LEABHARLANN CHOLÁISTE NA TRÍONÓIDE, BAILE ÁTHA CLIATH | TRINITY COLLEGE LIBRARY DUBLIN
Ollscoil Átha Cliath | The University of Dublin

Terms and Conditions of Use of Digitised Theses from Trinity College Library Dublin

Copyright statement

All material supplied by Trinity College Library is protected by copyright (under the Copyright and Related Rights Act, 2000 as amended) and other relevant Intellectual Property Rights. By accessing and using a Digitised Thesis from Trinity College Library you acknowledge that all Intellectual Property Rights in any Works supplied are the sole and exclusive property of the copyright and/or other IPR holder. Specific copyright holders may not be explicitly identified. Use of materials from other sources within a thesis should not be construed as a claim over them.

A non-exclusive, non-transferable licence is hereby granted to those using or reproducing, in whole or in part, the material for valid purposes, providing the copyright owners are acknowledged using the normal conventions. Where specific permission to use material is required, this is identified and such permission must be sought from the copyright holder or agency cited.

Liability statement

By using a Digitised Thesis, I accept that Trinity College Dublin bears no legal responsibility for the accuracy, legality or comprehensiveness of materials contained within the thesis, and that Trinity College Dublin accepts no liability for indirect, consequential, or incidental, damages or losses arising from use of the thesis for whatever reason. Information located in a thesis may be subject to specific use constraints, details of which may not be explicitly described. It is the responsibility of potential and actual users to be aware of such constraints and to abide by them. By making use of material from a digitised thesis, you accept these copyright and disclaimer provisions. Where it is brought to the attention of Trinity College Library that there may be a breach of copyright or other restraint, it is the policy to withdraw or take down access to a thesis while the issue is being resolved.

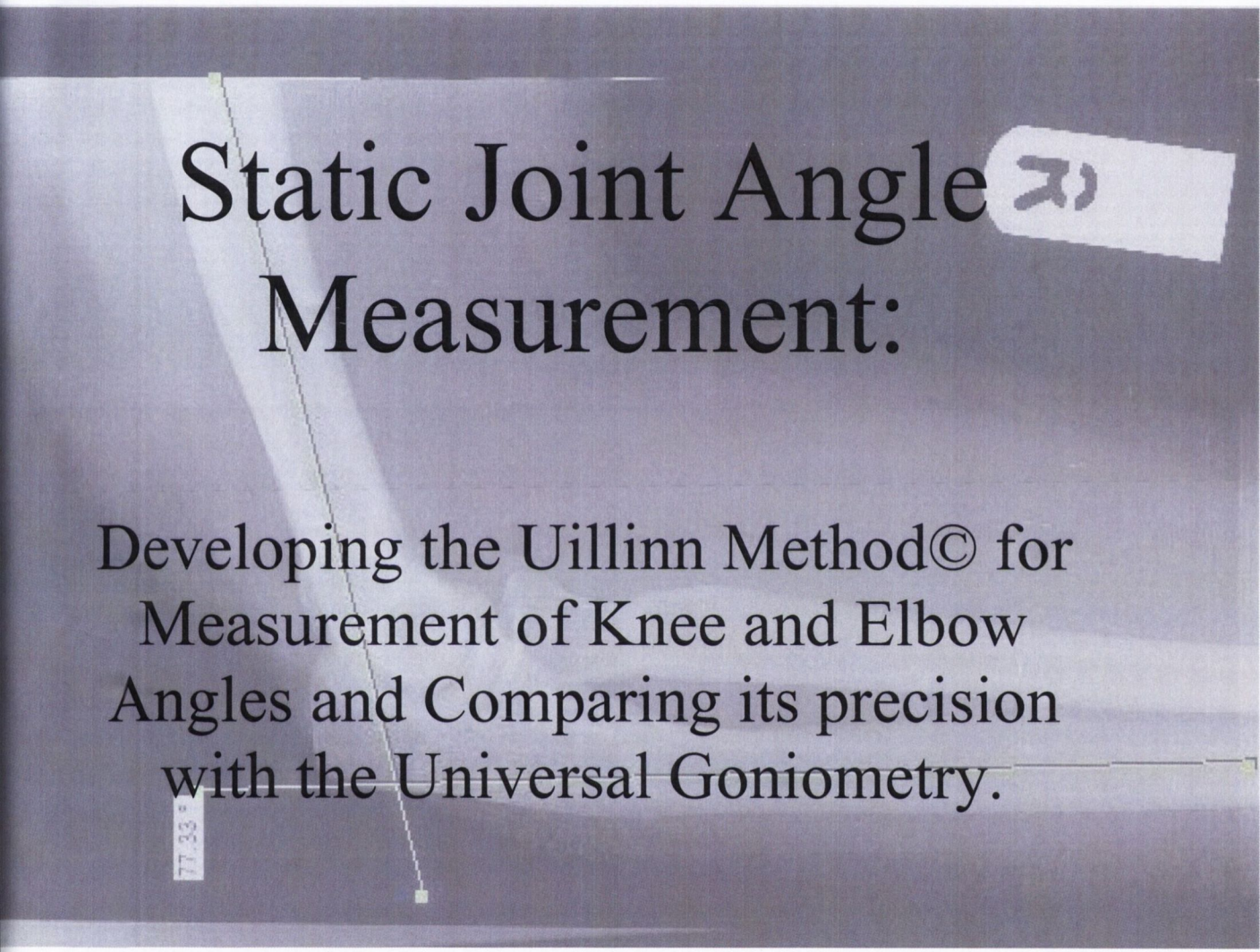
Access Agreement

By using a Digitised Thesis from Trinity College Library you are bound by the following Terms & Conditions. Please read them carefully.

I have read and I understand the following statement: All material supplied via a Digitised Thesis from Trinity College Library is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of a thesis is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form providing the copyright owners are acknowledged using the normal conventions. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

Static Joint Angle Measurement:

Developing the Uillinn Method© for Measurement of Knee and Elbow Angles and Comparing its precision with the Universal Goniometry.



PhD. Thesis
April 2005
Colin Dunlevy

TRINITY COLLEGE
19 MAY 2005
LIBRARY DUBLIN
THESIS
7583

Permission to Copy.

I grant permission to reproduce this thesis at the discretion of the depository library at Trinity College Dublin.

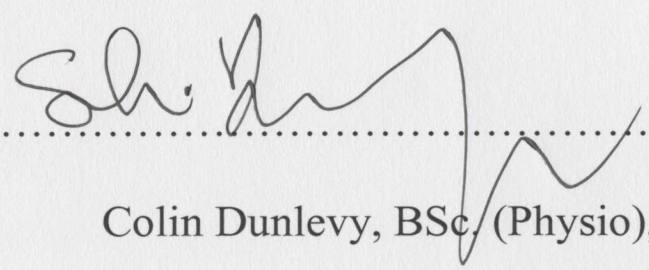
Declaration.

I declare that all work in the thesis is authentic and has not been previously published or submitted to TCD or any other institution.

The author devised, planned and carried out all experiments in this thesis. All intellectual property for the conception, design and construction of the Uillinn Method© should be attributed to the author. This system of virtual goniometry has not been previously submitted for acedemic consideration.

Paul Dunlevy is credited with writing the software code for the system, under the instruction of the author.

The radiographs that were taken during the validity trial were taken by the radiographic department at the Adelaide and Meath Hospital incorporating the National Children's Hospital (AMNCH) Dublin and are therefore their property.

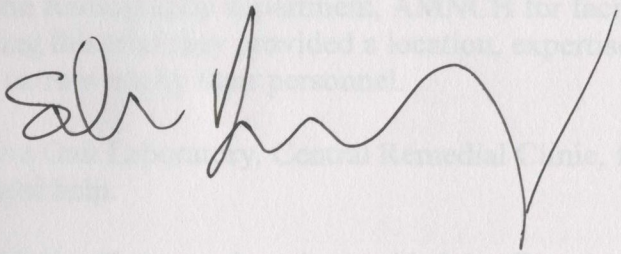


Colin Dunlevy, BSc. (Physio), MISCP.

DECLARATION

This thesis has not been submitted as an exercise for a degree at any other university. Except where stated, the work described therein was carried out by me alone.

I give permission for the Library to lend or copy this thesis upon request.

Signed: 

Acknowledgements.

I would like to thank and acknowledge the help from the following people, without whom this study would not have been possible;

- All the subjects that took part in the study.
- The management at Tallaght hospital (AMNCH), who continuously provide support and encouragement with research.
- To the physiotherapy department, AMNCH and particularly to Grace Cooke for both providing a location and equipment for the study, for facilitating study leave and providing most of the subjects for the reliability trial.
- To the Radiography department, AMNCH for facilitating the validity trial. During this trial they provided a location, expertise, equipment, afforded time and extra work by their personnel.
- To the Gait Laboratory, Central Remedial Clinic, for printing facilities and general help.
- To Marese Cooney who, along with John Gormley, were enduring in their efforts, provided encouragement and criticism during the study and without whom this study would not have been possible.
- To the examiners, Mr. Eoin O'Raithalaigh (MISCP), Ms. Emma Gallivan (MISCP) and Ms. Deborah Gillanders (MISCP) who gave up a substantial amount of time and were extremely tolerant during their involvement in the study. Particular thanks is imparted to Mr. Eoin O'Raithalaigh for proof-reading and constructive criticism.
- To my family and friends who were very supportive and understanding of the commitment needed to complete the work.
- To my brother, Paul Dunlevy who whose IT skills were invaluable and without whom the Uillinn Method© would not have been created.
- Finally and especially, to my wife, Clare who is always very patient and supportive on a personal level, but also gave a massive amount of her time and wisdom during the compilation study.

Table of Contents.

Permission to Copy. i
Declaration. ii
Acknowledgements. iii
Table of Contents. iv
List of Figures. x
Summary. xix
List of Abbreviations. xxi
Introduction. xxiii

Chapter 1 STATIC OSTEOKINEMATIC MEASUREMENT. 1

1.1 Joint Angle Measurement 2
1.1.1 Introduction to Osteokinematics Measurement. 2
1.1.2 The Rationale for Osteokinematic Measurement. 4
1.1.3 Methods of Joint Angle Evaluation. 5
1.1.4 The Universal Goniometer 5
1.15 Inclinometers 7
1.1.6 Electrogoniometer 7
1.1.7 Photography 7
1.1.8 Visual Estimation 8
1.2 Literature Review of the Joint Angle Measurement Techniques 9
1.2.1 Universal Goniometry. 9
1.2.2 Summary of Universal Gonometer. 27
1.2.3 Inclinometers. 28
1.2.4 Summary of Inclinometers. 29
1.2.5 Visual Estimation. 29
1.2.6 Summary of Visual Estimation 35
1.2.7 Electrogoniometers. 36
1.2.8 Summary of Electrogoniometry. 48
1.2.9 Photography. 49
1.2.10 Summary of Photographic based Measurement Systems 57
1.2.11 Radiography. 58
1.2.12 Summary of Radiographic based Measurements 60
1.3 Biomechanics Consideration at the Knee and Elbow. 62
1.3.1 Anatomy of the Knee 62
1.3.2 Normal Range of Movement 64
1.3.3 Anatomy of the Elbow 65
1.3.4 Normal Range of Movement 66
1.3.5 Considerations for Joint Measurement 67
1.4 Discussion 68
1.4.1 Measurement Error 68

1.4.2 Scientific Research 71
1.4.3 Study Populations 71
1.4.4 Statistical Analysis 73
1.4.5 Osteokinematic Measurement Tools 81
1.4.6 Challenges to Static Osteokinematic Measurement. 82
1.4.7 Specific Challenges to Photographic Measurement Systems. 84
1.4.8 Current Knowledge of Measurement Error 88
1.4.9 Implications of Measurement Error 92
1.4.10 Areas Requiring Further Research 93
1.5 Conclusion. 94

Chapter 2 METHODS: THE DEVELOPMENT AND APPLICATION OF THE UILLINN METHOD© 96

2.1 Introduction 97
2.1.1. Background to Developing the Uillinn Method©. 98
2.2 Software Development 100
2.2.1. The Uillinn Method© Version 1.x 101
2.2.2. Operational Procedures for the Uillinn Method© 102
2.2.3. The Rationale for Further Development 103
2.2.4. Developing the Uillinn Method© Version 2.x. 104
2.3 Applying the Uillinn Method© to the Human Model 109
2.3.1. The 3 Marker Approach (3Mks) 109
2.3.2 The 4 Marker Approach (4Mks) 111
2.3.3. The No Marker Approach (NMks) 112
2.3.4. The Parallel Approach (Parallel) 113
2.3.5. The Midline Approach (Mid-Point) 115
2.4. Discussion 118
2.5. Conclusion 121

Chapter 3 METHODS: RATIONALE FOR THE EXPERIMENTAL WORK 122

3.1 Introduction 123
3.1.1 Gold-Standard Measurements 124
3.1.2 Comparisons Between Methods 125
3.1.3 Deficits in Knowledge for Universal Goniometry. 125
3.1.4 Deficits in Knowledge for Photographic Measurement Methods. 127
3.1.5 Clinical Usage of Measurement Technology. 128
3.2 Specific Aims 130
3.2.1 Outline of the Study Design 130
3.2.2 Data Analysis 132
3.2.3 Ethics 134
3.3 Discussion 135
3.4 Conclusion. 137

Chapter 4 THE ACCURACY OF THE UILLINN METHOD© CALCULATIONS:A NON-CLINICAL TRIAL. 138

4.1 Introduction. 139
4.2 Specific Aims of the Study. 140
4.2.1 Justification of the Aims. 140
4.3 Methodology. 143
4.3.1 Calculation of the Control Data. 143
4.3.2 Data Analysis. 145
4.3.3 Equipment 145
4.4. The Experiments & Results. 147
4.4.1. Experiment 1: To determine the potential accuracy of the UM under optimal conditions. 147
4.4.2 Results: Experiment 1. 147
4.4.3 Experiment 2: To compare error for the UM when tripod mounted versus free-hand photography is used. 151
4.4.4 Results: Experiment 2. 151
4.4.5 Comparing Free-Hand and Tripod Mounted Photography 155
4.4.6. Experiment 3: To investigate the influence of the pillow effect, lens imperfections and parallax error on the UM. 157
4.4.7. Results: Experiment 3. 158
4.4.8. Experiment 4: To investigate the influence of camera or target rotation on the UM. 164
4.4.9. Results: Experiment 4. 166
4.4.7 Guidelines for Use of the Uillinn Method©. 171
4.5 Discussion. 173
4.5.1 Limitations of the Study. 175
4.5.3 Limitations of Data Analysis 177
4.5.4 Recommendations for Future Research 179
4.6 Conclusion. 183

Chapter 5. RELIABILITY OF THE UNIVERSAL GONIOMETER AND THE UILLINN METHOD©. 184

5.1. Introduction. 185
5.2 Specific Aims of the Study. 187
5.2.1 Justification of the Aims. 187
5.3 Methodology. 190
5.3.1 The Examiners 191
5.3.2 The Subjects 191
5.3.3 Equipment 192
5.3.4 Order of Measurements 196
5.3.5 Examination Procedures 196
5.3.6 Measurements by the Uillinn Method© 198
5.3.7 Data Analysis 199
5.4 Results. 202
5.4.1 Intra-Rater Reliability 202
5.4.2 Universal Goniometer at the Knee 202
5.4.3 Uillinn Method© (No Marker Approach) at the Knee. 203
5.4.4 Uillinn Method© (Parallel Approach) at the Knee. 203
5.4.5 Uillinn Method© (Mid-Point Approach) at the Knee. 204

5.4.6 Uillinn Method© (3-Marker Approach) at the Knees. 204
5.4.7 Uillinn Method© (4-Marker Approach) at the Knee. 205
5.4.8 Universal Goniometer at the Elbow 206
5.4.9 Uillinn Method© (No Marker Approach) at the Elbow. 206
5.4.10 Uillinn Method© (Parallel Approach) at the Elbow. 207
5.4.11 Uillinn Method© (Mid-Point Approach) at the Elbow. 207
5.4.12 Uillinn Method© (3-Marker Approach) at the Elbow. 208
5.4.13 Uillinn Method© (4-Marker Approach) at the Elbow. 208
5.4.14 Inter-Rater Reliability 209
5.4.15 Universal Goniometer for Knee Measurements 209
5.4.16 Uillinn Method (No Markers) for Knee Measurements 212
5.4.17 Uillinn Method (Parallel) for Knee Measurements 214
5.4.18 Uillinn Method (Mid-Point) for Knee Measurements 216
5.4.19 Uillinn Method (3-Markers) for Knee Measurements 218
5.4.20 Uillinn Method (4-Markers) for Knee Measurements 220
5.4.21 Universal Goniometer for Elbow Measurements 223
5.4.22 Uillinn Method (No Markers) for Elbow Measurements 225
5.4.23 Uillinn Method (Parallel) for Elbow Measurements 227
5.4.24 Uillinn Method (Mid-Point) for Elbow Measurements 229
5.4.25 Uillinn Method (3-Markers) for Elbow Measurements 231
5.4.26 Uillinn Method (4-Markers) for Elbow Measurements 233
5.4.27 Confidence in the Data 236
5.5. Discussion 239
5.5.1. The Most Reliable Method 247
5.5.2. Other Work 248
5.5.3. Reliability in Different Joint Ranges 251
5.5.4. Limitations of the Study 255
5.6. Conclusion 263

Chapter 6. CONCURRENT VALIDITY OF UNIVERSAL GONIOMETRY AND THE UILLINN METHOD©. 265

6.1. Introduction. 266
6.2. Specific Aims of the Study 268
6.2.1. Justification of the Aims. 268
6.3. Methodology. 271
6.3.1. Equipment. 272
6.3.2. Subjects. 273
6.3.3. Procedures. 274
6.3.4. Calculation of the X-ray data 279
6.3.5. Data Analysis. 281
6.4. Results. 282
6.4.1. Reliability of the PACS system 282
6.4.2. Validity Results for Measurement of the Knee 284
6.4.3. The Universal Goniometer (UG) at the Knee. 285
6.4.4. The Uillinn Method (No Markers Approach) at the Knee. 285
6.4.5. The Uillinn Method (Parallel Approach) at the Knee. 286
6.4.6. The Uillinn Method (Mid-Point Approach) at the Knee. 286
6.4.7. The Uillinn Method (3-Markers Approach) at the Knee. 287
6.4.8. The Uillinn Method (4-Markers Approach) at the Knee. 287
6.4.9. Validity Results for Measurement of the Elbow 290
6.4.10. The Universal Goniometer (UG) at the Elbow. 290

6.4.11. The Uillinn Method (No Markers Approach) at the Elbow. 290
6.4.12. The Uillinn Method (Parallel Approach) at the Elbow. 292
6.4.13. The Uillinn Method (Mid-Point Approach) at the Elbow. 292
6.4.14. The Uillinn Method (3-Markers Approach) at the Elbow. 293
6.4.15. The Uillinn Method (4-Markers Approach) at the Elbow. 293
6.4.16. Confidence of Results Data 295
6.5. Discussion 297
6.5.1. Examination Procedures 297
6.5.2. Value of the Research 298
6.5.3. Non-Random Error (Systematic Bias) 299
6.5.4. Random Error 300
6.5.5. The Criterion Data 300
6.5.6. Relationship of VMAs 303
6.5.7. Different Joint Ranges 309
6.5.8. Other Research 312
6.5.9. Skin Markers 314
6.5.10. Limitations 320
6.5.11. Further Work 321
6.5.12. Summary 321
6.6. Conclusion. 324

Chapter 7 A CLINICAL TRIAL OF THE UILLINN METHOD©. 325

7.1. Introduction. 326
7.2. Specific Aims of the Study 328
7.2.1. Justification of the Aims. 328
7.3. Methodology. 331
7.4. Results 336
7.4.1. Uillinn Method© Scoring Sheet (Practical Training) 336
7.4.2. Audit of Systems Usage. 338
7.4.3. Questionnaire Responses. 338
7.4.4. Training Issues 340
7.4.5. Operational Issues 340
7.4.6. Professional Issues 341
7.4.7. Design Issues 342
7.5. Discussion 344
7.6. Conclusion. 349

Chapter 8 DISCUSSION & CONCLUSION 350

8.1. Introduction 351
8.2. Over-view of the Literature 352
8.3. Development and Application of the Uillinn Method©. 355
8.3.1.The Rationale for Research 355
8.4. Overview of the Experiments 356
8.4.1. The Uillinn Method© Calculation System: A Calibration Question. 357

8.4.2. Reliability of the Universal Goniometer and the Uillinn Method©. 357
8.4.3. Concurrent Validity of Universal Goniometry and the Uillinn Method©. 359
8.4.4. Using the Uillinn Method© in the Clinic Environment 359
8.5. Discussion 361
8.5.1 The Best Quantitative Approach to Static Joint Angle Measurement. 361
8.5.2. Measurement Error 362
8.5.3. Value of the Research. 363
8.5.4. The Uillinn Method©: Implications for the Clinical Environment. 365
8.5.5. The Uillinn Method©: Implications for Research. 368
8.6. Conclusion 371

Appendices 372
Appendix A 373
Appendix B 374
Appendix C 376
Appendix D 380
Appendix E 383
Appendix E 389
Appendix E (cont.) 396
Appendix E (cont.) 403
Appendix F 410
Appendix G 428
Appendix H 433
Appendix I 435
Appendix J 437
Appendix K 438
Appendix L 441
Appendix M 445
Appendix N 449
Appendix O 453
Appendix P 454
Appendix P 462
Appendix Q 476

References 477

List of Figures.

Figure 1.1 The Universal Goniometer.
Figure 1.2 The Parallelogram Goniometer
Figure 1.3. The Flexible/ Strain Gauge Goniometer.
Figure 1.4. The Orthoranger Electrogoniometer.
Figure 1.5. The Genucom Electrogoniometer.
Figure 1.6. The NK Hand Assessment Laboratory (Short Arms).
Figure 1.7. The Pillow Effect.
Figure 2.1. The Craniovertebral angle.
Figure 2.2. The Uillinn Method©.
Figure 2.3. Version 2.1. of the Uillinn Method©.
Figure 2.4. Version 2.2. of the Uillinn Method© (Parallel Approach).
Figure 2.5. Version 2.3. of the Uillinn Method© (Parallel Approach).
Figure 2.6. Version 2.4. of the Uillinn Method© (Mid-point Approach).
Figure 2.7. The '3 Marker Approach' to Applying the Uillinn Method©.
Figure 2.8. The '3 Marker Approach' at the Elbow
Figure 2.9. The '4 Marker Approach' to the Uillinn Method©.
Figure 2.10. The '4 Marker Approach' at the Elbow.
Figure 2.11. The 'No Marker Approach' to the Uillinn Method©.
Figure 2.12. The 'No Marker Approach' at the Elbow.
Figure 2.13. The 'Parallel' Approach to the Uillinn Method©.
Figure 2.14. The 'Parallel Approach' at the Elbow.
Figure 2.15. The 'Midline Approach' at the Elbow.
Figure 4.1. The Uillinn Method© used to assess angle 'a' from one of the triangles
Figure 4.2. Illustration of the Cosine Rule.
Figure 4.3. The Smart Media Disk and Smart Media Disk Drive
Figure 4.4. A description of the lay out of each photograph.
Figure 4.5. A grid of 25 squares, each containing a triangle.
Figure 4.6. Illustration of Experimental Scheme into the effects of Rotation on the Uillinn Method©.
Figure 4.7. Controlling rotation of the whiteboard.
Figure 4.8. The Uillinn Method© of measurement: A photograph with the angle in 40° of rotation.
Figure 4.9. Potential Measurement Error of 3° of an Equilateral Triangle due to Inaccurate Placement of the Measurement
Figure 5.1. The 2 kg weight bags used to stabilise the foot (no markers, first rotation).
Figure 5.2. The 2 kg weight bags used to stabilise the arm (with markers, first rotation).
Figure 5.3. The 2 kg weight bags used to stabilise the foot with the wooden wedge.
Figure 5.4. The wooden wedge used to stabilise Knee hyperextension (with markers, 2nd rotation).
Figure 5.5. One of the Blind Universal Goniometers (UG).
Figure 5.6. The Photographic procedure at the Elbow.

Figure 5.7. The Uillinn Method Software. Analysis of the knee flexion angle (Parallel method).
Figure 5.8. The Uillinn Method Software. Analysis of the knee flexion angle (3-Marker method).
Figure 5.9. The Uillinn Method Software. Analysis of the knee flexion angle (4-Marker method).
Figure 5.10. Overlay of Photographs, Movement of Subject 34.
Figure 5.11. Overlay of Photographs, Movement of Subject 50.
Figure 5.12. Overlay of Photographs, Movement of Subject 18.
Figure 5.13. Overlay of Photographs, Movement of Subject 29.
Figure 6.1. Lower Limb Cast with Velcro Straps and Wooden fixing rod.
Figure 6.2. Medial View of the Knee with Adhesive Markers.
Figure 6.3 Four Lateral Limb Markers on the Greater trochanter, lateral epicondyle, head of the fibula and the lateral malleolus.
Figure 6.4 The PACS System Being Used for Angular Assessment of the Knee.
Figure 6.17 Example of concavities seen in the humerus during PACS assessments.
Figure 6.5. Overlay of the PACS angles calculation on the Uillinn Method (No Marker Approach) screen.
Figure 6.6. Overlay of the PACS angles calculation on the Uillinn Method (No Marker Approach) screen.
Figure 6.7. Overlay of the PACS angles calculation on the Uillinn Method (Parallel Approach) screen.
Figure 6.8. Overlay of the PACS angles calculation on the Uillinn Method (Parallel Approach) screen.
Figure 6.9. Overlay of the PACS angles calculation on the Uillinn Method (Mid-Point Approach) screen.
Figure 6.10. Overlay of the PACS angles calculation on the Uillinn Method (Mid-Point Approach) screen.
Figure 6.11. Overlay of the PACS angles calculation on the Uillinn Method (3-Marker Approach) screen.
Figure 6.12. Overlay of the PACS angles calculation on the Uillinn Method (3-Marker Approach) screen.
Figure 6.13. Overlay of the PACS angles calculation on the Uillinn Method (4-Marker Approach) screen.
Figure 6.14. Overlay of the PACS angles calculation on the Uillinn Method (4-Marker Approach) screen.
Figure 6.15 Individual Differences in Joint Axis as represented by the PACS measurement lines (yellow).
Figure 6.16 Example of X-ray Measurements made by Brosseau et al (2001), lines
drawn through the longitudinal bone axes (Black lines).
Figure 6.18. Typical Error to the X-ray data. All measurements.
Figure 7.1 Blinding the Uillinn Method©.

List of Tables.

Table 1.1. Data Provided by Nigel-Keenan et al (2004) for Reliability of measurements made with a Universal Goniometer.

Table 1.2. Correlation Coefficients for Upper Limb Joint Measurement Reliability.

Table 1.3. Correlation Coefficients for Lower Limb Measurement Reliability.

Table 1.4. Reliability of Upper Limb Measurement Methods (Reported without using Correlation Coefficients).

Table 1.5. Reliability of Lower Limb Measurement Methods (Reported without using Correlation Coefficients).

Table 1.6. Validity of Goniometric Measurement Methods.

Table 1.7. Correlation Table Between Two Cohorts with Large Bias

Table 1.8. Example of the Calculations Made with Means and Standard Deviations

Table 4.3. The Accuracy of the UM for Angular Measurements (Optimal Conditions) (n=60).

Table 4.4. The Control Data and the Uillinn Method© (Free-hand Photography, n=60).

Table 4.5. Comparison of Freehand Vs Tripod Mounted Photography.

Table 4.6. The Accuracy of the UM for Angular Measurements (n=120).

Table 4.7. The Pillow Effect, Lens Imperfections and Parallax Error on the UM; Off Centre on Photograph.

Table 4.8. The Pillow Effect, Lens Imperfections and Parallax Error on the UM; Periphery of Photograph.

Table 4.9. Comparison of Variables for Measurements in Different Areas of the Photograph.

Table 4.10. Control Data (CD) and the Uillinn Method© (UM) data with Rotation during Photography .

Table 4.11. Calculations of Agreement and Random Error for the UM due to Rotation.

Table 4.12 Typical Error due to Rotation (with Confidence Limits).

Table 5.1. Intra-rater Reliability of the Universal Goniometer (the Knee)

Table 5.2. Intra-rater Reliability of the No Marker Approach to the Uillinn Method© (the Knee).

Table 5.3. Intra-rater Reliability of the Parallel Approach to the Uillinn Method© (the Knee)

Table 5.4. Intra-rater Reliability of the Mid-Point Approach to the Uillinn Method© (the Knee)

Table 5.5. Intra-rater Reliability: Uillinn Method©, 3-Marker Approach (the Knee)

Table 5.6. Intra-rater Reliability of the 4-Marker Approach to the Uillinn Method© (the Knee)

Table 5.7. Intra-rater Reliability of the Universal Goniometer (the Elbow)

Table 5.8. Intra-rater Reliability of the No Marker Approach to the Uillinn Method© (the Elbow)

Table 5.9. Intra-rater Reliability of the Parallel Approach to the Uillinn Method© (the Elbow)
Table 5.10. Intra-rater Reliability of the Mid-Point Approach to the Uillinn Method© (the Elbow)
Table 5.11. Intra-rater Reliability of the 3-Marker Approach to the Uillinn Method© (the Elbow)
Table 5.12. Intra-rater Reliability of the 4-Marker Approach to the Uillinn Method© (the Elbow)
Table 5.13. Inter-rater Reliability of the Universal Goniometer (the Knee)
Table 5.14. Inter-rater Reliability of the Uillinn Method (No Marker) (the Knee)
Table 5.15. Inter-rater Reliability of the Uillinn Method (Parallel) (the Knee)
Table 5.16. Inter-rater Reliability of the Uillinn Method (Mid-Point) (the Knee)
Table 5.17. Inter-rater Reliability of the Uillinn Method (3-Markers) (the Knee)
Table 5.19. Inter-rater Reliability of the Uillinn Method (4-Markers) (the Knee)
Table 5.20. Inter-rater Reliability of the Universal Goniometer (the Elbow)
Table 5.21. Inter-rater Reliability of the Uillinn Method (No Marker) (the Elbow)
Table 5.22. Inter-rater Reliability of the Uillinn Method (Parallel) (the Elbow)
Table 5.23. Inter-rater Reliability of the Uillinn Method (Mid-Point) (the Elbow)
Table 5.24. Inter-rater Reliability of the Uillinn Method (3-Markers) (the Elbow)
Table 5.25. Inter-rater Reliability of the Uillinn Method (4-Markers) (the Elbow)
Table 5.26. Inter-rater Reliability of all Techniques. Typical error (TE) and ICC (3,1) between all examiners.
Table 5.27. Intra-rater Reliability: Standard Error of the Differences for the different measurement techniques.
Table 5.32. Inter-rater Reliability: Standard Error of the Differences for the different measurement techniques.
Table 5.27. Intra-rater Reliability of All Techniques (Repeatability Coefficient)
Table 5.28. Intra-rater Reliability of All Techniques (Typical Error).
Table 5.29. Inter-rater Reliability of All Techniques (Repeatability Coefficient)
Table 5.30. Inter-rater Reliability of All Techniques (Typical Error).
Table 5.32. Inter-rater Reliability: Standard Error of the Differences for the different measurement techniques.
Table 5.33 Typical Error for all Examiners. Both the Intra-rater and Inter-rater Reliability Data.
Table 5.34 Average Typical Error Between all Examiners Different Ranges of the Knee.
Table 5.35 Average Typical Error Between all Examiners Different Ranges of the Elbow.
Table 6.1. Intra-rater Reliability of the PACS system of Joint angle measurements
Table 6.2. Inter-rater Reliability of the PACS system of Joint angle measurements.
Table 6.3. Data Analysis for each of the Experimental Measures, Comparisons with X-ray (PACS) Measurements of the Knee.
Table 6.4. Data Analysis for each of the Experimental Measures, Comparisons with X-ray (PACS) Measurements of the Elbow.
Table 6.5. Standard Error and Confidence Limits for the Knee Data Calculations.

Table 6.6. Standard Error and Confidence Limits for the Elbow Data Calculations
Table 6.7. Comparison of the Validity of each tool for Knee measurements in Outer and Mid Range.
Table 6.8. Comparison of the Validity of each tool for Elbow measurements in Mid Range and Inner Range
Table 7.1 Results from the Practical Section of the Training Session. All repeated measurements fall within the Repeatability Coefficient (RC) of 3.3°.
Table 7.2 Results from the Practical Section of the Training Session. All repeated measurements fall within the Limits of Agreement (LOA) of –7.9°, 10°.
Table 7.3 Results from the Practical Section of the Training Session. All repeated measurements fall within the Repeatability Coefficient (RC) of 4.5°.
Table 7.4 Results from the Practical Section of the Training Session. All repeated measurements fall within the Limits of Agreement (LOA) of –10°, 7.3°.
Table 7.5 Total Usage of the Uillinn Method© of the 3-Month trial.
Table 7.6. Training for the Uillinn Method©.

List of Graphs.

Graph 1.1 An Illustration of Unrelated Data Maintaining Means and Standard Deviations (from Table 1.8).

Graph 4.2 Difference Between Uillinn Method and Control Data (Optimal Conditions/ with Tripod).

Graph 4.3 The Uillinn Method (Optimal Conditions/ with Tripod) Vs Control Data.

Graph 4.4 Bland & Altmans 95% Limits of Agreement Scatter Plot. The Uillinn Method (Optimal Conditions/ with Tripod) Vs Control Data.

Graph 4.5 Difference Between Uillinn Method and Control Data (Freehand Photography/ without Tripod).

Graph 4.6 The Uillinn Method (Freehand/ without Tripod) Vs Control Data.

Graph 4.7 Bland & Altmans 95% Limits of Agreement Scatter Plot. The Uillinn Method (Free Hand Photography) Vs The Control Data.

Graph 4.8 Bland & Altmans 95% Limits of Agreement Scatter Plot. The Uillinn Method (Tripod & Freehand) and the Control Data.

Graph 4.9 Uillinn Method Vs Control Data, Off-Centre in Photograph.

Graph 4.10 Uillinn Method Vs Control Data, at Periphery of Photograph.

Graph 4.11 Bland & Altmans 95% Limits of Agreement Data Plot. The Uillinn Method Vs The Control Data (Angles off Centre in Photograph).

Graph 4.12 Bland & Altmans 95% Limits of Agreement Data Plot. The Uillinn Method Vs The Control Data (Angles at Periphery of Photograph).

Graph 4.13. Uniform Increase in Observed Error as Measured Angles Move away from the Centre of the Photograph.

Graph 4.14. Random Error with Increasing Rotation between the Angle and the Camera.

Graph 4.15. Bland & Altmans 95% Limits of Agreement Scatter Plot. The Uillinn Method Vs the Control Data (5° Rotation During Photograph).

Graph 4.16. Bland & Altmans 95% Limits of Agreement Scatter Plot. The Uillinn Method Vs the Control Data (10° Rotation During Photograph).

Graph 4.17. Bland & Altmans 95% Limits of Agreement Scatter Plot. The Uillinn Method Vs the Control Data (20° Rotation During Photograph).

Graph 4.18. Bland & Altmans 95% Limits of Agreement Scatter Plot. The Uillinn Method Vs the Control Data (30° Rotation During Photograph).

Graph 4.19. Bland & Altmans 95% Limits of Agreement Scatter Plot. The Uillinn Method Vs the Control Data (40° Rotation During Photograph).

Graph 5.1. Inter-rater Reliability of the Knee Angle Measurements. Universal Goniometer. Examiner 1 Vs Examiner 2.

Graph 5.2. Inter-rater Reliability of the Knee Angle Measurements. Universal Goniometer. Examiner 1 Vs Examiner 3.

Graph 5.3. Inter-rater Reliability of the Knee Angle Measurements. Universal Goniometer. Examiner 2 Vs Examiner 3.

Graph 5.4. Inter-rater Reliability of the Knee Angle Measurements. Uillinn Method (No Markers). Examiner 1 Vs Examiner 2.
Graph 5.5. Inter-rater Reliability of the Knee Angle Measurements. Uillinn Method (No Markers). Examiner 2 Vs Examiner 3.
Graph 5.6. Inter-rater Reliability of the Knee Angle Measurements. Uillinn Method (No Markers). Examiner 1 Vs Examiner 3.
Graph 5.7. Inter-rater Reliability of the Knee Angle Measurements. Uillinn Method (Parallel). Examiner 1 Vs Examiner 2.
Graph 5.8. Inter-rater Reliability of the Knee Angle Measurements. Uillinn Method (Parallel). Examiner 2 Vs Examiner 3.
Graph 5.9. Inter-rater Reliability of the Knee Angle Measurements. Uillinn Method (Parallel). Examiner 1 Vs Examiner 3.
Graph 5.10. Inter-rater Reliability of the Knee Angle Measurements. Uillinn Method (Mid-point). Examiner 1 Vs Examiner 2.
Graph 5.11. Inter-rater Reliability of the Knee Angle Measurements. Uillinn Method (Mid-point). Examiner 2 Vs Examiner 3.
Graph 5.12. Inter-rater Reliability of the Knee Angle Measurements. Uillinn Method (Mid-point). Examiner 1 Vs Examiner 3.
Graph 5.13. Inter-rater Reliability of the Knee Angle Measurements. Uillinn Method (3 Markers). Examiner 1 Vs Examiner 2.
Graph 5.14. Inter-rater Reliability of the Knee Angle Measurements. Uillinn Method (3 Markers). Examiner 1 Vs Examiner 3.
Graph 5.15. Inter-rater Reliability of the Knee Angle Measurements. Uillinn Method (3 Markers). Examiner 1 Vs Examiner 3.
Graph 5.16. Inter-rater Reliability of the Knee Angle Measurements. Uillinn Method (4 Markers). Examiner 1 Vs Examiner 2.
Graph 5.17. Inter-rater Reliability of the Knee Angle Measurements. Uillinn Method (4 Markers). Examiner 2 Vs Examiner 3.
Graph 5.18. Inter-rater Reliability of the Knee Angle Measurements. Uillinn Method (4 Markers). Examiner 1 Vs Examiner 3.
Graph 5.19. Inter-rater Reliability of the Elbow Angle Measurements. Universal Goniometer. Examiner 1 Vs Examiner 2.
Graph 5.20. Inter-rater Reliability of the Elbow Angle Measurements. Universal Goniometer. Examiner 2 Vs Examiner 3.
Graph 5.21. Inter-rater Reliability of the Elbow Angle Measurements. Universal Goniometer. Examiner 1 Vs Examiner 3.
Graph 5.22. Inter-rater Reliability of the Elbow Angle Measurements. Uillinn Method (No Markers). Examiner 1 Vs Examiner 2.
Graph 5.23. Inter-rater Reliability of the Elbow Angle Measurements. Uillinn Method (No Markers). Examiner 2 Vs Examiner 3.
Graph 5.24. Inter-rater Reliability of the Elbow Angle Measurements. Uillinn Method (No Markers). Examiner 1 Vs Examiner 3.
Graph 5.25. Inter-rater Reliability of the Elbow Angle Measurements. Uillinn Method (Parallel). Examiner 1 Vs Examiner 2.
Graph 5.26. Inter-rater Reliability of the Elbow Angle Measurements. Uillinn Method (Parallel). Examiner 2 Vs Examiner 3.
Graph 5.27. Inter-rater Reliability of the Elbow Angle Measurements. Uillinn Method (Parallel). Examiner 1 Vs Examiner 3.

Graph 5.28. Inter-rater Reliability of the Elbow Angle Measurements. Uillinn Method (Mid-Point). Examiner 1 Vs Examiner 2.
Graph 5.29. Inter-rater Reliability of the Elbow Angle Measurements. Uillinn Method (Mid-Point). Examiner 2 Vs Examiner 3.
Graph 5.30. Inter-rater Reliability of the Elbow Angle Measurements. Uillinn Method (Mid-Point). Examiner 1 Vs Examiner 3.
Graph 5.31. Inter-rater Reliability of the Elbow Angle Measurements. Uillinn Method (3 Markers). Examiner 1 Vs Examiner 2.
Graph 5.32. Inter-rater Reliability of the Elbow Angle Measurements. Uillinn Method (3 Markers). Examiner 2 Vs Examiner 3.
Graph 5.33. Inter-rater Reliability of the Elbow Angle Measurements. Uillinn Method (3 Markers). Examiner 1 Vs Examiner 3.
Graph 5.34. Inter-rater Reliability of the Elbow Angle Measurements. Uillinn Method (4 Markers). Examiner 1 Vs Examiner 2.
Graph 5.35. Inter-rater Reliability of the Elbow Angle Measurements. Uillinn Method (4 Markers). Examiner 2 Vs Examiner 3.
Graph 5.36. Inter-rater Reliability of the Elbow Angle Measurements. Uillinn Method (4 Markers). Examiner 1 Vs Examiner 3.

Graph 5.37. Intra-rater Reliability: Standard Error within the Data for the Difference between the Repeated Assessments.
Graph 5.38. Intra-rater Reliability: Standard Error within the Data for the Difference between the Repeated Assessments.
Graph 5.39. Inter-rater Reliability: Standard Error within the Data.
Graph 5.40. Inter-rater Reliability: Standard Error within the Data.
Graph 5.41. Intra-rater Reliability: Typical Error for Different Measurement Techniques.
Graph 5.42. Intra-rater Reliability: Typical Error for Different Measurement Techniques.
Graph 5.43. Intra-Rater Reliability at the Knee – Worst Observed Cases
Graph 5.44. Intra-Rater Reliability at the Elbow – Worst Observed Cases
Graph 5.45. Systematic Bias Between Examiners for the Different Measurement Techniques.
Graph 5.46. Inter-rater Reliability: Typical Error for Different Measurement Techniques, Between Examiners.
Graph 5.47. Inter-rater Reliability: Typical Error for Different Measurement Techniques Between Examiners.
Graph 5.48. Inter-rater Reliability: Typical Error for Different Measurement Techniques, Between Examiners.
Graph 5.49. Intra-rater Reliability: Average Typical Error with the Highest Observed Error for all Measurement Techniques
Graph 5.50. Inter-rater Reliability: Typical Error with Highest Observed Error for Paired Measurements.
Graphs 6.1. Limits of Agreement between PACS data (X-ray) and Universal Goniometer for Knee measurements.
Graphs 6.2. Limits of Agreement between PACS data (X-ray) and the No Markers Approach to the Uillinn Method for Knee measurements.
Graphs 6.3. Limits of Agreement between PACS data (X-ray) and the Parallel Approach to the Uillinn Method for Knee measurements.

Graphs 6.4. Limits of Agreement between PACS data (X-ray) and the Mid-Point Approach to the Uillinn Method for Knee measurements
Graphs 6.5. Limits of Agreement between PACS data (X-ray) and the 3-Marker Approach to the Uillinn Method for Knee measurements.
Graphs 6.6. Limits of Agreement between PACS data (X-ray) and the 4-Marker Approach to the Uillinn Method for Knee measurements.
Graphs 6.7. Limits of Agreement between PACS data (X-ray) and the Universal Goniometer for Elbow measurements.
Graphs 6.8. Limits of Agreement between PACS data (X-ray) and the No Marker Approach to the Uillinn Method for Elbow measurements.
Graphs 6.9. Limits of Agreement between PACS data (X-ray) and the Parallel Approach to the Uillinn Method for Elbow measurements.
Graphs 6.10. Limits of Agreement between PACS data (X-ray) and the Mid-Point Approach to the Uillinn Method for Elbow measurements.
Graphs 6.11. Limits of Agreement between PACS data (X-ray) and the 3-Marker Approach to the Uillinn Method for Elbow measurements.
Graphs 6.12. Limits of Agreement between PACS data (X-ray) and the 4-Marker Approach to the Uillinn Method for Elbow measurements.
Graphs 6.13. Standard Error in the Data; Differences between PACS and the other measurement techniques.
Graphs 6.14. Standard Error in the Data; Differences between PACS and the other measurement techniques.
Graph 6.15 Typical Error (with t-tests shown) of Agreement with PACS in different ranges for Knee measurements.
Graph 6.16 Typical Error (with t-tests shown) of Agreement with PACS in different ranges for Elbow Measurements.
Graph 7.1 Total System usage by each therapist.

Summary.

PURPOSE: There remains a deficit of knowledge regarding the expected error associated with the universal goniometer (UG). The Uillinn Method© (UM) was devised as an alternative approach to joint angle measurement. This system uses personal computer (PC) software to form a virtual goniometer to assess joint angles, taken from digital photographs. There was a need to establish the best procedure to apply the UM to measure human movement.

RELEVANCE: Clinicians use joint range of motion (ROM) data to establish diagnoses and make treatment decisions. It is essential that ROM assessment tools are accurate and trustworthy. Therefore, validity and reliability investigations must be conducted to inform the clinicians about the limitations of available measurement tools.

SUBJECTS: The reliability of the UM and the UG was established on a population of healthy knees (n=51) and elbows (n=45). The validity investigation used a patient population (knees, n=42; elbows, n=50).

METHODS AND MATERIALS: Two digital cameras (Polariod 2.1 megapixels; Olympus 1.3 megapixels) were used interchangeably during the study in conjunction with a laptop computer. An initial non-clinical trial established the accuracy of the UMs calculations and provided evidence for best photographic procedures. The Reliability trial used three examiners to perform repeated blind examinations of the elbow and knee. The Validity trial made comparisons between joint angles calculated by X-rays against those assessed by the UG and each of 5 different approaches to the UM. These examinations were also blind. A final qualitative study reported of the experiences of 11 physiotherapists who had used the new UM system clinically over a 3-month period.

ANALYSES: Results were expressed in terms of random and non-random error. The conclusions of the study were based on tests of the 95% limits of agreement (LOA), repeatability coefficient (RC) and Typical Error (TE). Significance testing and correlation coefficients were also used to a limited degree. Means and standard deviations were also reported

RESULTS: The non-clinical trial revealed that in order to achieve optimal results photographic procedures were important. The calculation system would typically be accurate to 0.3°. For 95% of measurements it would be accurate to 0.8°.

The reliability trial showed that for the UG repeated measurements by the same examiner would agree by better than 7° (TE≈ 2.5°) at the knee and 9° (TE≈ 2.5°) at the elbow (for 95% of occasions). The UM (Mid-Point approach) would agree by under 3.5° (TE≈ 1.5°) (knee) and 6° (TE≈ 2.3°) (elbow). Between examiners knee measurements using the UG agreed by 95% LOA of -11°, 14° (TE≈ 4.5°) and by –5.5°, 7° (TE≈ 2.2°) for the UM. At the elbow there was a systematic bias of approximately 6° between examiners for the UG with a 95% LOA of -7°, 14° (TE≈ 7.2°). The UM (Mid-Point) had no systematic bias between examiners at the elbow and a 95% LOA of -4.5°, 4.5° (TE≈ 2.1°).

The validity trial showed that at the knee the UG had a 95% LOA with X-ray data of –7.5°, 7.9° (TE≈ 2.75°), while the UM (Mid-Point) had a systematic bias of 2.75° and a 95% LOA of -10°, 4.5° (TE≈ 2.6°). At the elbow the UG had a bias of 2.2° and a 95% LOA of -14°, 18.5° (TE≈ 6°), the UM had a large bias of 12°, but a low 95% LOA of -1°, 25° (TE≈ 4.6°).

The system received a modest amount of use from the therapists over a 3-month period. Most users reported that it took an additional 2-minutes to conduct the assessments and that this amount of time was too long and presented a barrier to it clinical use.

CONLUSIONS: Conducting human joint angle measurements is a challenge. This study found that elbow measurements hold more error than the knee measurements. The UM is a useful method of assessing joint angles and it improves upon the validity and reliability of the conventional UG. Using adhesive markers on anatomical landmarks do not improve either the validity or reliability of the UM.

List of Abbreviations.

2-D	=	Two dimensional
3-D	=	Three dimensional
AAOS	=	American Academy of Orthopaedic Surgeons
ACL	=	Anterior cruciate ligament
ADF	=	Ankle dorsi-flexion
ANOVA	=	Analysis of Variance.
APAS	=	Ariel Performance Analysis System™
APF	=	Ankle plantar flexion
CD	=	Control Data
cm	=	Centimetre
CT	=	Computer topography
CV	=	Craniovertebral
EG	=	Electrogoniometer
EOR	=	End of Range.
EMS	=	Error Mean Square (from ANOVA table)
EPR	=	Electronic patient record
FFG	=	Fluid Filled Goniometer
Fig	=	Figure
IBG	=	Internet Based Goniometer
ICC	=	Intraclass Correlation Coefficient
IT	=	Information Technology
JPEG	=	Joint Photographic Experts Group
kVp	=	kilo Volts
LED	=	light emitting diodes
LOA	=	Limits of Agreement
m	=	Metre
MAVS	=	Motion Analysis™ Video System
mA/s	=	milliamps per second
MCL	=	Medial collateral ligament
mm	=	Millimetre

MRI	=	Magnetic Resonance Imaging
OA	=	Osteoarthritis
PACS	=	Picture Archiving and Communication System
PADS	=	The Postural Analysis Digitising System
PC	=	Personal Computer
PG	=	Parallelogram Goniometer
PPMCC	=	Pearsons Product Moment Correlation Coefficient
r	=	Pearsons Product Moment Correlation Coefficient
ROM	=	Range of Movement
SB	=	Systematic Bias
SEM	=	Standard Error of Measurement
Std Dev (SD)	=	Standard Deviation
SUN	=	International standards for symbols, units and nomenclature
TMJ	=	Temperomandibular joint
UM	=	Uillinn Method©
UG	=	Universal Goniometer
VE	=	Visual Estimation

Introduction.

Health professionals who deal with the musculoskeletal system regularly have call to assess the precise quantity of movement available at a joint. This is done to consider limitations of movement at a joint which in turn may help the diagnostic process, may provide information about the effectiveness of a treatment programme or provide evidence regarding the overall function of the patient (Arnheim and Prentice, 1993; The American Academy of Orthopaedic Surgeons (AAOS), 1994; Macnicol, 1995; Norkin and White, 1995; Szulc et al, 2000).

When conducting joint range of movement (ROM) examinations, the clinician must be sure that their measurements are as precise as possible (Rothstien, 2001), to this end there are many measurement tools available which aim to improve the accuracy of assessments.

When the plethora of published research that investigated the performance of joint angle measurement tools was examined, it was clear that our professional knowledge regarding this fundamental task was incomplete. This deficit presented an urgent need to improve our understanding of the issues associated with clinical joint angle measurement. Further research was required to clearly describe the error associated with commonly used tools. In addition, it was felt there was a need to develop alternative approaches to these tools.

To address this problem a type of joint angle measurement tool called the 'Uillinn Method©' (UM) (version 2.4) was created. This device was based on a method of quantifying human postural angles (Uillinn Method©, version 1.2) described by Dunlevy et al (2000). The technique uses the relatively new technology of digital photography and a software application which runs on a standard personal computer (PC). The software forms a virtual goniometer which can measure joint angles displayed in digital photographs.

There was a need to investigate the ability of the UM to conduct measurements of joint angles and to assess its performance against the commonly used universal goniometer (UG).

Therefore, this thesis is concerned with the development of a new method of static joint angle measurement. It also deals comprehensively with all the issues surrounding the testing and use of clinical joint angle measurement tools.

The joints chosen as the subject matter for the experimental work were the knee and the elbow. The reason that these joints were selected was because they both have hugely complicated biomechanical properties which present a challenge in measurement. In addition they are both commonly injured joints, which places a priority on the need to understand the error associated with their measurement. The study focused on sagittal angles, i.e. flexion and extension, as it is in this plane that the majority of movement occurs for both joints.

Chapter 1

STATIC OSTEOKINEMATIC MEASUREMENT.

1.1 Joint Angle Measurement.

1.1.1 Introduction to Osteokinematics Measurement.

The study of joint motion recognises two fundamental components. Those movements that occur inside the joint capsule, the 'arthrokinematics' and those movements that occur outside the joint capsule the 'osteokinematics'.

Arthrokinematics may be thought of as pertaining specifically to intricate joint dynamics regardless of the gross movement. It considers the dynamics between the adjacent bone surfaces inside the capsule and typically describes the relative elements of gliding, spinning and rolling motions.

While osteokinematics suggests a description of the movement that takes place outside the joint capsule and relate to the assessment of the orientation of the bone shafts. These may be thought of as dealing with the movement of the body segments and are related to gross function. Osteokinematic measurement can further be divided into the assessment of dynamic movement and assessing static joint angles. Dynamic assessments are commonly associated with 3 dimensional (3-D) motion analysis systems. This type of measurement is outside the scope of this investigation.

Static osteokinematic measurement is associated with the assessment of a stationary end-of-range (EOR) joint angle, it is this type of measurement that is central to this current work.

Assessments of static joint angles are often used to quantify the available range of movement (ROM) between two body segments. These are calculated by measuring the arc of available movement of the bone shafts about the joint axis. ROM measurements are usually angular in nature. Movement of the joint is typically described in terms of direction and magnitude and whether it is active or passive, e.g. 100° active shoulder abduction, 90° passive knee flexion, etc. The displacement of the joint is referenced to the zero neutral position, which is an anatomical description of body posture where joints are in 0° of displacement (AAOS, 1994). For health professionals working in the fields of physical medicine and rehabilitation osteokinematic measurements are a fundamental part of their daily work (Soames, 2003).

Goniometry is the term given to the method or tools used to measure joint angles. It is derived from two Greek words, *gonia*, meaning angle and *metron*, meaning to measure (Norkin and White, 1995). There are a number of goniometric measurement methods and tools available to clinicians that are designed to quantify osteokinematic joint motion, of these, the universal goniometer (UG) is probably the most commonly used physical tool in the clinical setting (Coxford et al, 1998; Burton et al, 1999). However, there are numerous other tools available including novel innovative measurement techniques that offer an alternative to the UG. Some of these tools claim to have advantages over the UG. Whatever type of osteokinematic assessment tool is used, it is vital to the clinician that measurement results can be trusted. Rothstein (2001) underlines the danger of blurring the divisions between measurement knowledge and measurement opinion, implying that the need for trustworthy accurate measurement is paramount.

A paper by Sims and Arnell (1993) states that the concepts of validity and reliability are central to building confidence towards the measurement process. They define measurement validity as relating to the extent to which an instrument measures what it is intended to measure and reliability as concerning the extent to which the measurement instrument yields the same results on repeated uses, either by the same examiner (intra-rater reliability) or different examiners (inter-rater reliability) (Sims and Arnell, 1993).

The research process can yield information regarding validity and reliability of a measurement tool. It is critical that research has been carried out on all measurement methods and tools used by clinicians and that the methodology of the research has followed standard scientific principles. Data on the validity and reliability of osteokinematic measurement procedures should be easily available and presented in a clinically relevant manner. If there is to be confidence in health professionals' clinical assessments and subsequent diagnoses, clinicians must know the limitations and associated errors of their tools. The onus is on the clinicians to be aware of the findings of scientific research that has been conducted on their clinical equipment (Rothstein, 2001).

1.1.2 The Rationale for Osteokinematic Measurement.

Repeated goniometric measurement of a joint or body segment undergoing therapy is essential to monitor the success of a particular treatment (Arnheim and Prentice, 1993; Szulc et al 2000). Maitland (1991) and Macnicol (1995) maintain that repeated joint angle assessment is necessary to monitor treatment and assess recovery from injury. They continue that information gathered during an assessment can be used as a baseline to compare with future assessments. Szulc et al (2000) hold that goniometric measurement is vital when assessing joint function and locomotion.

The guidelines towards clinical joint measurement published by the American Academy of Orthopaedic Surgeons (AAOS) (1994) state that ROM deficits may indicate joint pathology and functional impairment. Therefore, it is important to assess the exact amount of available movement and related movement deficit at a joint (Greene and Heckman, 1994). This publication also asserts that quantifying and recording joint ROM is imperative for other reasons, including helping to identify clinical patterns of movement and dysfunction, providing an index for the progression and severity of a disorder and providing valuable information about the efficacy of a particular treatment regime. They go on to state that the data gathered is objective information that is important when assessing and calculating disability ratings. A disability rating scale or the primary goniometric data may be useful to statutory or insurance agencies that must make financial decisions regarding a personal injury or disability claim.

To summarise, the benefits of accurate joint angle assessments include the ability to:

- Establish a baseline of available movement.
- Help to determine the presence or absence of joint dysfunction.
- Help to establish a diagnosis.
- Evaluate progress in relation to treatment goals.
- Establish the need to modify a treatment.
- Help to motivate the patient.
- Allow research into the effects of a specific treatment regime.
- Optimise fabrication of orthoses and prosthetic equipment.
- Establish a capsular or non-capsular pattern of movement.

- Provide information to assist making decisions regarding surgical management.

1.1.3 Methods of Joint Angle Evaluation.

There are numerous ways of calculating joint angles. However, all of the different methods may be classified under two headings, 'invasive imaging' and 'non-invasive' or external methods. External methods of measurement are those procedures that assess the joint externally at the surrounding tissue, i.e. they employ any type of goniometric device or procedure that does not attempt to penetrate the skin. Invasive imaging methods of joint angle measurement include X-ray and other diagnostic imaging techniques where the aim of the procedure is to image the actual bones themselves ignoring the over-lying soft tissue.

However, systems that have the ability to conduct such assessments do not lend themselves well to routine clinical measurement. X-ray, computer topography (CT) or magnetic resonance imaging (MRI) are highly specialised expensive machines that have biohazardous issues (Fuji et al 1997, Shakhatreh, 2001).

External procedures are the most commonly used methods of assessing joint angles. This is primarily because they are quicker, cheaper and generally more practical than invasive imaging methods, but not because they are thought of as being more accurate. The major disadvantage of external methods of joint measurement is that they can only estimate the position of the bones and therefore the data gathered for the precise orientation of the bones in space are approximated. Commonly used goniometric tools include the universal goniometer (UG), inclinometer, electronic goniometers (EG), visual estimation (VE) and photographic measurement.

1.1.4 The Universal Goniometer

The universal goniometer (UG) remains the most commonly found measurement tool in a clinical setting (Coxford et al, 1998; Burton et al, 1999). This device comprises of 2 measuring arms situated about a calibrated fulcrum (see Figure1.1).

The operational procedure for this tool involves placing the axis of the device close to the rotational axis of the joint. The arms of the goniometer are then aligned with proximal and distal bony landmarks. Joint position or available ROM can then be read from the incremental analogue scale on the goniometer. Norkin and White (1995, 2003) outline the common practice for placement of this device over each of the measurable joints in the body.

Other variations of this tool include different sizes, different materials and design variations. One variation on the design is the parallelogram goniometer (PG) (see Figure 1.2). This latter device is a tool almost identical to a plastic UG apart from a parallelogram device built into the moveable arm that allows it to shift up and down. This feature apparently enables the device to measure a joint angle without using a fixed point to represent the joints centre of rotation.

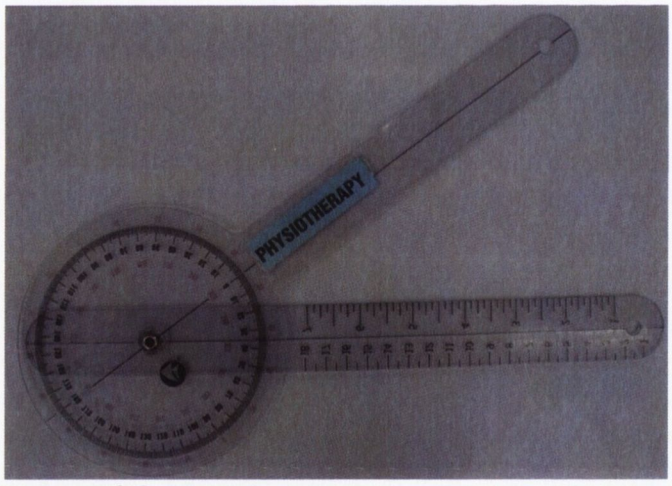


Figure 1.1 The Universal Goniometer.

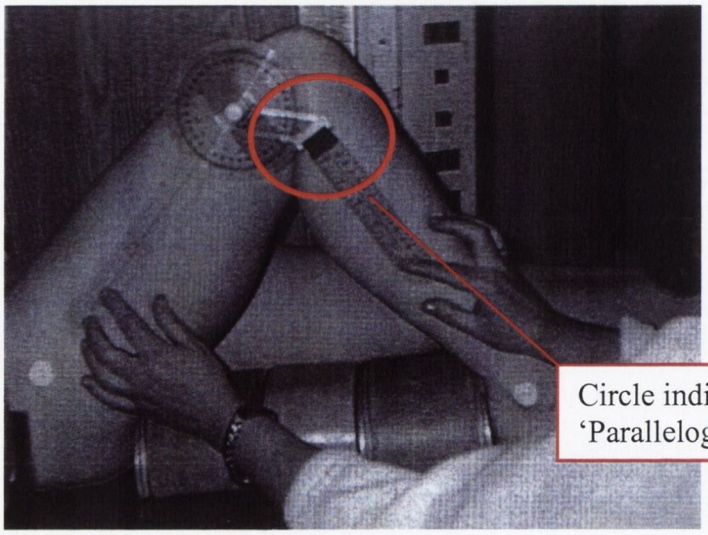


Figure 1.2 The Parallelogram Goniometer.

1.15 Inclinometers

Inclinometers are a type of goniometer that rely on external forces such as global magnetism, hydrodynamics or the direct influence of gravity to ascertain an angular measurement.

Two common types are the pendulum goniometer and the fluid filled goniometer (FFG). The FFG is a tool that uses the properties of liquid and air much like the common spirit level. This type of device will require either passive or active movement from the zero neutral position through to the end-of-range position.

1.1.6 Electrogoniometer

The term electrogoniometer (EG) covers a number of devices all of which base their calculations on electrical equipment. These devices are based on a number of different and diverse designs. These devices are dealt with more comprehensively further in this section (see 1.2.9).

1.1.7 Photography

Remondino (2001) defines the term photogrammetry as describing a technique of conducting measurements made from photographs. This was first used in published work in 1867 when the art and science of photography itself was still new. Over the last 80 years the principal application of photogrammetry has been the compilation of maps from aerial photographs. In addition, there has been a continuing development of the applications of photogrammetric close-range techniques to many other fields including engineering, architecture, archaeology, medicine, industrial quality control and robotics (Remondino, 2001). Photographic and video images have been used to assess joint movement and gait analysis for many years. The concept is that when a body segment is captured on film during dynamic movement or in a desired static position, the processed image can be analysed independent of the subject (Cameron et al, 1993; Gajdosik et al, 1994; Marks and Karkouti, 1996; Marks, 1996; Marks and Quinney, 1997; Karkouti and Marks (1997) and Marks, 1999). This type of measurement may be called osteokinematic photogrammetry.

1.1.8 Visual Estimation

The visual estimation (VE) of joint ROM is used by clinicians in the absence of more objective tools (Coxford et al 1998). This technique relies on the subjective interpretation of the assessor to evaluate a joint angle by sight only. This is a very common technique as it is very quickly done without the requirement of equipment (Watkins et al, 1991; Youdas et al, 1993; Norkin and White, 1995, Burton et al, 1999).

In conclusion, it is necessary for the clinician to measure available joint movement if a physical examination is to be comprehensive (Soames, 2003). Assessment of joint ROM provides objective information about the joint. The data gathered from joint angle measurement can be used as a stand-alone treatment outcome or to infer information relating to joint function (Greene and Heckman, 1994). The type of data that is sought from goniometric measurement is *ratio* data, this type of data is the most objective and accountable, includes a natural zero and has empirical meaning (LoBiondo-Wood and Haber, 1998). A valid and reliable joint angle measurement tool will yield this powerful ratio type data, in which the clinician can have confidence. This type of authoritative data is essential to scientific and health care research (Hicks, 2004; Atkinson and Nevill, 2000).

1.2 Literature Review of the Joint Angle Measurement Techniques

In order to better understand the issues associated with the practice of joint angle measurement there was a need review the published literature on this topic. Many studies were identified that have investigated the efficacy of the different common goniometric tools such as the universal goniometer (UG), inclinometer, electronic goniometers (EG), visual estimation (VE) and photographic measurement. The literature regarding these types of measurement is discussed in this section.

The subsequent experimental work was concerned with sagittal plane angles of the elbow and knee only, however, the literature review considered a wide range of bodily joints. This was done in order to evaluate variations in error between different measures, to understand the limitations of inferals between joints and to provide a more comprehensive understanding of the challenges facing static joint angle measurement.

1.2.1 Universal Goniometry.

The UG appears to have been the most frequently investigated goniometric method to date. Rothstein et al (1983) undertook a reliability study that investigated intra-rater, inter-rater and inter-device reliability of different sized UG measurements. The 3 devices they investigated were all UG's, a large metal UG, a large plastic UG and a small plastic UG. Their study used a total of 12 physical therapists to randomly assess a population of 24 patients using the 3 different UG's. The examiners were randomly paired together and measured 12 different joints using each of the different goniometers. The Pearson product moment correlation coefficient (PPMCC), which correlates 2 cohorts and returns a coefficient (r), was used to calculate a reliability score. The results for the intra-rater reliability of flexion and extension measurements for both the knee and elbow was high at between r=0.91 to 0.99. Inter-rater reliability was also high at r=0.88 to 0.97, except for knee extension which was r=0.63 to 0.70. The 3 different goniometers used compared well with each other when assessed with the intraclass correlation coefficient (ICC), which is a statistical tool that allows correlations of 2 or more cohorts. The results for this trial yielded ICCs of between 0.92 and 0.99. The study

states that the reliability of measuring knee extension fared poorly. Nonetheless, they conclude that intra-rater and inter-rater reliability will be maintained irrespective of UG type. This study did not use confidence intervals or mean scores to illustrate the magnitude of error between the data which limits its clinical interpretation.

Grohmann (1983) undertook a study to determine which of two different methods of UG placement at the elbow was the more reliable. The two methods identified were the 'Lateral method' and the 'over-the-joint method'. The former method was similar to the method described by Norkin and White (1995), except it did not specify an anatomical landmark as the joint axis, but states that if the measurement arms are placed correctly the UG fulcrum will be at the approximate joint centre. The second method, the over-the-joint method requires the UG to be placed along the dorsal aspect of the limb segments. The examiners were told that during this method care must be taken to place the arms of the UG evenly against the skin, leaving the axis of movement well away from the joint.

The experimental procedure used 40 physiotherapists as examiners who used a blinded UG to assess 1 healthy elbow. Two different joint positions were used, which resulted in 4 different measures (over-the-joint method at ≈125°, over-the-joint method at ≈45°, lateral method at ≈125° and lateral method at ≈45°). A plywood board with a series of perpendicular pegs inset in to it, were used as a stabilising device for the 2 upper limb positions. The measurements were carried out over 4 consecutive days.

The data were analysed using a 3 way-ANOVA, F-ratios, mean square values as estimates of variance and means and standard deviations (SD). The results revealed no significant inter-rater differences for either relative flexion or extension between each method. Mean difference between the over-the-joint and lateral methods were +1° and –1.2° respectively (at approx. ≈125°). Standard deviations of the differences between each method are not given. The authors conclude that both methods of UG alignment are reliable, but propose that during the more standard lateral method, maintaining correct UG alignment may be difficult and that over-the-joint method may be easier.

Another reliability study by Mayerson and Milano (1984) repeated measurements of a total 22 joint positions using only 1 subject and 2 examiners. The joints measured were described as the major joints of the upper and lower limbs. They found that both inter-rater and intra-rater measurements shared a similar 95% confidence interval (CI) that ranged from between –2.3° to + 4.3°. The PPMCC for intra-rater reliability was r=0.98 and inter-rater was r=0.97, which were both high. They conclude that it is reasonable to expect an average fluctuation of about 4° when measuring joint angles with a UG across all major joints through all ranges of movement. They suggest that this error can be attributed to instrument error, which they describe as either erroneous placement or misreading the tool. Other sources of error that were identified were external (environmental) and subjective (emotional) factors, which may add to the magnitude of error. A very small patient population of 1 and that fact that they do not specify which joints they measured or how the joints were positioned, limits the importance of this study. However, because they used confidence intervals to analyse their data, it is possible to appreciate the actual magnitude of difference that they found between the repeated measurements.

Walker et al (1984) undertook a study that aimed to add to the knowledge of normal ROMs for the white population. This study used healthy subjects which were assessed for 24 extremity movements. Each measurement was repeated 5 times by each of the 4 examiners. They used a standard UG, which was not blinded.

The results found that there was a significant difference in the average ROM between men and women in the age group 60 to 84 years (n=60). As part of the study, intra-rater and inter-rater reliability was assessed using PPMCC, means and SD. The results show that intra-rater reliability averaged a high correlation coefficient, r=0.81 (range 0.78 to 0.99). The mean error between repeated measurements was 5° ± 1°. Only the 2 most agreeable (out of the 4) examiners were assessed for inter-rater reliability, this analysis was carried out using descriptive analysis. The mean and SD of error between these two examiners were reported as 6° ± 5° in total. For 3 of the 24 motions (shoulder medial rotation, radioulnar pronation and ankle dorsiflexion) the average difference exceeded 10°.

Interpreting results from this study must be taken with caution especially with regard to the inter-rater reliability scores. In that part of the study the

methodology seemed to show bias by only comparing the 2 examiners who had the most similar results. In addition, limited data analysis was conducted. However, this study did not claim to be primarily a reliability trial, but an investigation to broaden the knowledge base of normal ROMs of caucasians.

Gogia et al (1987) conducted a reliability and validity study of UG measurement at the knee joint. It claims to be the first study to attempt to validate UG by correlating it against a gold-standard radiographic measurement. They used 30 volunteers with no history of knee injury. The subjects were positioned in side-lying with their uppermost leg supported on a flat wooden board. Goniometric measurements of a random angle of the knee were carried out by 2 examiners, which were followed immediately by an X-ray. The data for all measurements were assessed using means and standard deviations as well as statistical analysis using both PPMCC and the ICC. Inter-rater reliability was high at r=0.99 and ICC=0.99. Validity was also very high r=0.97 and ICC=0.98. However, the mean and standard deviation (SD) used in the study appear only to describe the average test angle used in the study and not the mean and SD of the differences between the data sets. This is required for in-depth analysis of the data (Hopkins, 2000a). Nonetheless, the difference between the mean X-ray data and the data for the examiners was 3° for examiner 1, and 1.7° examiner 2. This figure of 3° is the average difference and does not give an account of the SD or the largest differences between the data sets. The somewhat limited data analysis is the only limiting aspect of an otherwise useful study, which demonstrates a SB of approximately 3° for measurements taken with X-ray and UG.

Enwemeka (1986) is the only study reviewed that investigated validity of UG measurement alone. This study compared UG measurements of the knee with measurements taken from X-ray. Ten volunteers each had their legs assessed by X-ray and UG in 6 different positions, 0°, 15°, 30°, 45°, 60° and 90° flexion. This yielded a total of 60 X-rays and 60 UG measurements. Both datasets were then compared.

The data was statistically analysed using PPMCC and the student t-test was used to test the null hypothesis that, joint angles determined by goniometry did not

differ from measurements obtained by bone angle measurement using X-ray at 0°, 15°, 30°, 45°, 60° and 90° knee flexion. The PPMCC was high (r=0.99) for the entire data. The null hypothesis was accepted for measurements at 30°, 45°, 60° and 90° flexion, implying similarity of X-ray and UG measurements. However, the null hypothesis was rejected for measurements between 0° to 15° of knee flexion, implying significant differences between X-ray and UG measurements in this outer range knee position. The mean difference between the 2 data sets was 4.59° (± 4.76 °) in this outer range, which the author deemed to be too high. For the ranges between 30° to 90° of knee flexion, the average differences were 0.52° to 3.81° (± 5.58° to ±7.2°), which the study claims is an acceptable magnitude of difference.

However, the acceptance of significant differences between X-ray and UG in the outer range is based on mean differences only and does not appear to take account of the standard deviations (SD) seen in this range. The SDs seen in the outer range (0° to 15°) are actually the lowest across all the data, implying most consistent agreement with X-ray and the lowest random error. Having the highest mean difference in this range merely implies a slightly larger systematic error between the 2 cohorts. In fact, the highest SD of the differences was ±7.2° and was recorded at 90° knee flexion (mean difference 2.32°). A SD of this magnitude would translate to a 95% limits of agreement (LOA) spread of approximately 14° (at 90°), compared to 9.3° (at 15°).

The author attributes the proposed significant difference in the outer range of knee joint movement to external rotation of the tibia that may occur as the knee moves into extension. This rotation is known as the 'screw-home mechanism', which describes the rotational movement between the tibia and femur in the knees outer ROM. From 30° to 90° the author states that goniometric measurement does not differ significantly from measurements obtained using X-ray, and therefore that UG is a valid tool in this range. The study was limited by the sample size which gave it a low statistical power (degrees of freedom (df) = 9) for each angular band.

Riddle et al (1987) set out to examine both intra-rater and inter-rater reliability of passive shoulder joint ROM using both a small and large UG. The 7 passive ROMs that were assessed were; shoulder flexion, extension, abduction into elevation, horizontal abduction, horizontal adduction, medial rotation and lateral

rotation. The 50 subjects that were used in the study were recruited from a patient population. A total of 16 physical therapists were randomly paired to measure the different joint movements. All therapists used both the large and small UGs to complete the measurements during the trial. When all assessments were complete there were a total of 100 measurements for each joint motion. The interclass correlation coefficient (ICC) was used to evaluate the intra-rater reliability, inter-rater reliability and the inter-device reliability of the large and small UGs for each joint motion. Intra-rater reliability was high for all measurements for both devices (large UG ICC=0.90 to 0.99; small UG ICC=0.93 to 0.98). The inter-rater reliability for both size UGs followed the same trends across different joints and were generally variable, ranging from poor results for shoulder extension (large ICC=0.27, small ICC=0.26) to good results for lateral rotation (large ICC=0.88, small ICC=0.90). The inter-device reliability was high for all measurements made (ICC=0.87 to 0.98).

Based on their correlation coefficient results the authors conclude that there is a high level of reliability between measurements made by the same examiner, but care should be taken when interpreting results of measurements taken by different examiners for shoulder medial rotation, extension, horizontal abduction and horizontal adduction. They are in agreement with other researchers (Rothstein et al, 1983) in their finding of good reliability between different sized goniometers. This study did not conduct any type of descriptive data analysis and does not indicate the actual magnitude of differences between the repeated measurements made, which limits the clinical interpretation of the findings.

Bovens et al (1990) correctly propose that reliability investigations should consider 3 factors; variations for the individual examiner, variations between examiners and the biological variations of the patient, but their study does not explore these issues specifically. Their study investigated repeated goniometric measurement at several joints on a total of 48 subjects. Measurements were made by 3 examiners, all of whom were qualified physicians said to be experienced in joint angle measurement. Nine different movements were assessed on both the left and right hand sides of each subject. The 9 movements measured were; external rotation of the glenohumeral joint, pronation and supination of the radioulnar joint, flexion and extension of the wrist, eversion and inversion of the foot, dorsiflexion and

plantar flexion of the ankle. Only 8 of the subjects were included in the intrarater reliability trial. Each of the 3 examiners assessed each of these 8 subjects a total of 8 times over a 3 month period. Inter-rater reliability was investigated by dividing the total population of 48 into 5 subject groups of either 9 or 10 people per group. Each examiner assessed each patient within 3 weeks of the other examiners.

Descriptive analysis of the data was performed using mean, standard deviation and coefficient of variance calculations, while statistical analysis was conducted using the formula $R= ó^2T/ ó^2T + ó^2e$. This formula is described as a reliability coefficient (R), where, T is the true underlining value, X=T+e is the classical linear model for an observed score and $ó^2$ is the error variance. The reason the authors chose this formula and the statistical standard to which it belongs, is not discussed. This calculation method is not justified in the text and adds confusion to the interpretation and comparisons of the results.

The results show that intra-rater reliability varied for each examiner and for each joint measured. Intra-rater reliability was shown to be high for external rotation of the shoulder, supination of the superior radio-ulnar joint, ankle inversion and dorsiflexion. Intra-rater reliability was fair for ankle eversion and plantar flexion, while the result for wrist extension was poor. There were large disparities in intra-rater reliability between examiners for radio-ulnar pronation which ranged from fair to poor and wrist flexion which ranged from excellent to poor.

The results for inter-rater reliability again showed large variations between examiners for the different joints. Results were illustrated for each of the 5 subject groups and ranged from poor to fair for external rotation of the glenohumeral joint, radioulnar pronation, wrist extension, foot eversion, ankle plantar flexion and dorsi-flexion. Results improved slightly for radioulnar supination, wrist flexion and foot inversion, ranging between poor and excellent.

The standard approach to reliability research is to investigate either intra-rater (same examiner), inter-rater (different examiners) or inter-device (different tools) reliability. However, in this study Bovens and colleagues (1990) confuse their analysis as they appear to label the statistical analysis of the data as *Reliability* and the descriptive analysis of the reliability as *Variability*. Also the rationale for the 5 sub-groupings is not clear and appears only to weaken statistical power and further confuse the results.

In agreement with other studies, the authors conclude that intra-rater reliability is stronger than inter-rater reliability. They surmise that during the 2½ years of the trial over all reliability improved in the final stages implying that experience improves reliability. They hold that error associated with repeated UG measurement may be as much as 10°. However, due to the unusual nature of the definitions of reliability, a complicated methodology and the irregular statistical analysis employed, the results are difficult to interpret or to relate to other studies.

Brosseau et al (1997) investigated the reliability and validity of the UG and the parallelogram goniometer (PG). In their study both UG and PG were used by 2 examiners to assess ROM on 60 healthy knee joints. The procedure was repeated twice in both relative flexion and relative extension at the knee and X-rays were taken in each position. Intra-rater reliability was good for both the UG (extension ICC=0.86-0.94, flexion ICC=0.95-0.97) and PG (extension ICC=0.85-0.87, flexion ICC=0.91-0.96). Inter-rater reliability was good for both devices for knee flexion (UG ICC=0.91-1.0, PG ICC=0.82-0.88), but not as good as for extension (UG ICC=0.62-0.71, PG ICC=0.43-0.52).

The Pearsons product moment correlation coefficient (PPMCC) was used to assess the validity of both devices by correlating the results against measurements taken from X-rays. Again validity for knee extension was found to be poor for both the UG (r=0.33-0.48) and the PG (r=0.33-0.41), but much better for flexion at the knee (UG r=0.73-0.78, PG r=0.73-0.77). This study seems to illustrate that the UG is slightly more reliable tool than PG. It also agrees with Enwemekas' (1986) assertion of poor validity for the UG when measuring knee extension, as well as showing poor inter-rater reliability in this range, which was in agreement with Rothstein et al (1983).

Brosseau et al (2001) repeated the above study on a group of 60 subjects with residual musculoskeletal impairments of the knee. Again 2 examiners assessed each subject in relative flexion and in relative extension, twice, with both a UG and a PG. They also visually estimated the joint angle prior to the goniometric measurement. Each subject underwent radiographs in all test positions and angular data was calculated for each X-ray. There were good results for intra-rater reliability in knee flexion and extension, for both the UG (flexion ICC=0.99, extension

ICC=0.97-0.98) and PG (flexion ICC=0.99, extension ICC=0.95-0.95). Inter-rater reliability in both flexion and extension for both methods were good (UG: flexion ICC=0.97-0.98, extension ICC=0.89-0.92; PG: flexion ICC=0.95-0.97, extension ICC=0.85-0.89).

As in their previous study PPMCC was used to correlate validity from the X-ray data. Knee flexion correlated well with the X-ray data for both methods (UG r=0.97-0.98, PG r=0.97-0.98). However, knee extension again faired poorly for both devices (UG r=0.39-0.44, PG r=042-0.51). They conclude that both the UG and PG are reliable instruments, but for best results the same examiner should not use different measurement tools when re-assessing a joint angle. These authors agree with Brosseau et al (1997) who illustrated that both devices appear to have poorer validity when assessing knee extension.

A study by Sprigle et al (2003) was undertaken to determine the reliability and validity of a new pelvic goniometer designed to assess pelvis tilt and hip flexion in sitting. This tool is specifically designed to provide information on the sitting position of wheelchair users. The device has two components: a band that is placed over the iliac spines and a thigh bar to be placed along the femur. The design is apparently a hybrid between a standard goniometer and an inclinometer. Both validity and reliability of the device was investigated, in addition the reliability of the standard UG was tested. Data analysis used means and standard deviations, PPMCC, ANOVA tables, Bland and Altmans 95% limits of agreement (LOA) and Repeatability Coefficient (RC). Validity was established by assessing 10 healthy male subjects using both the pelvic goniometer and lateral radiographs. Each of the subjects were assessed in 3 self-selected positions of erect posture, anterior pelvic tilt and posterior pelvic tilt. In each position both radiographs and pelvic goniometry were conducted. One experienced physical therapist assessed all of the angles.

Results displayed high validity for the pelvic goniometer for both pelvic tilt (r=0.93) and hip flexion (r=0.81). The 95% LOA suggest that the average measurements with that goniometer will be -4.9° less than radiographic studies but have a LOA of -18°, 8°. These can be interpreted as equating to a variability of 13° over a 5° SB. Surprisingly these differences were not found to be significant.

In a separate procedure, both intra- and inter-rater reliability was tested for both the pelvic goniometer and the UG. Eight physical therapists were paired and asked to assess hip and pelvic positions for a sitting posture in a total of 20 healthy mixed gender subjects.

The RC (Bland and Altman, 1999) for intra-rater reliability when measuring hip flexion with the pelvic goniometer was 5.6° (mean difference = 3.2°), while the UG had a RC of 6.3° (mean difference = 3.1°). Inter-rater reliability suggested that for the pelvic goniometer the average difference between examiners would be 2.3° and will fall within the LOA range –3.9°, 8.4°. The UG also had a SB of 2.3° but a larger range of expected agreement of -10°, 14.6°. The authors feel that the result of this study show that the pelvic goniometer is better suited to assess sitting posture at the pelvis than conventional UG.

Nigel-Keenan et al (2004) underline the importance of the validity and reliability of osteokinematic measurement data when making decisions regarding the type of orthopaedic surgery that may be indicated for a patient. They state that inability to acquire closely related data is a major concern for motion analysis systems that rely heavily on certain clinical data. These authors set out to investigate the reliability of the UG for 4 different clinical measures of the lower limb. Examiner agreement was investigated for, the Thomas test, i.e. hip extension, knee extension, the popliteal angle (knee extension with the hip at 90° flexion) and measures of ankle dorsiflexion. The different joints were passively moved to their end of range and in that position they were assessed using a standard UG. Other aims of this study were to define and describe the techniques of physical examination of lower limb joint ROM and to investigate inconsistencies between static and dynamic clinical assessments.

The study used a patient population of 20 children and young adults (mean age 12.5 years, range 5-25) all of whom had a diagnosis of cerebral palsy (CP) (13 diplegic, 3 asymmetric diplegia, 4 hemiplegia) and 2 clinical examiners (1 physiotherapist and 1 orthopaedic surgeon). Each examiner assessed each subject twice; once prior to, and once post a routine 3 dimensional gait analysis. During the assessments the UG was not blinded, but the examiners were blind to other testers

measurement data. A third party who assisted during the all of the measurement procedures ensured a standardised assessment technique was followed.

The data analyses included a correlation coefficient identified only as the concordance correlation coefficient (CCC), the 95% Bland and Altman limits of agreement test (LOA) and general descriptive data analysis. The intra-rater scores were assessed by comparing the 1st and 2nd measurements of each examiner while the inter-rater scores were assessed by comparing the averages of each of the examiners 2 measurements. Measurements for the left and right lower limb were analysed separately.

The intra-rater reliability results showed a CCC ranging between 0.67 (dorsiflexion) to 0.96 (hip extension). While CCC scores for inter-rater reliability ranged from 0.34 (hip extension) to 0.87 (dorsiflexion). Ninty-five percent LOA scores indicated that expected agreement for measurements made by the same examiner varied from as low as 2.15° for hip extension to 12.3° for the popliteal angle. The 95% LOA scores for inter-rater reliability predict that agreement between examiners would range from 5.85° gastrocnemius length (ankle dorsiflexion) to 11.85° soleus length (ankle dorsiflexion) (see Table 1.1, 1.3, 1.5).

	Intra-Rater Reliability*		Inter-Rater Reliability	
	Left Lower Limb	*Right Lower Limb*	*Left Lower Limb*	*Right Lower Limb*
Clinical Measure	95% Limits of Agreement	95% Limits of Agreement	95% Limits of Agreement	95% Limits of Agreement
Hip Extension	-5.5° , 5.5°	-7.8° , 8.3°	-13.9° , 4.4°	-12.9° , 5.2°
Popliteal Angle	-9.0° , 10°	-11.7° , 13.2°	-13.6° , 7.8°	-13.6° , 4.9°
Knee Extension	-7.1° , 4.1°	-6.5° , 7°	-12° , 5°	-9.1° , 5.9°
Soleus Length	-9.1° , 12.1°	-9.2° , 11.7°	-10.7° , 13°	-10.5° , 8.5°
Gastroc. Length	-9.0° , 6.5°	-10.5° , 9.5°	-7.2° , 4.5°	-12.1° , 5°

* = Examiner 'A' only

Numeric Data in Degrees (°)

Table 1.1. Data Provided by Nigel-Keenan et al (2004) for Reliability of measurements made with a Universal Goniometer.

The authors established a limit of agreement of ±10° as being clinical acceptability prior to undertaking the study. The results found that this level of agreement was exceeded by repeated measurements of the popliteal angle and gastrocnemius length (ankle dorsiflexion) by the same examiner (intra-rater). For measurements made by different examiners (inter-rater) the ±10° limits was exceeded by measurements of the popliteal angle and soleus length (ankle

dorsiflexion). In addition to the limits of expected measurement error described by this type of data analysis, the 95% LOA test also provided information on the SB seen during the measurement procedures. This SB between examiners was approximately 3.5° for most measures while maximum bias was found to be for assessments of hip extension which was 4.8°.

The study can be criticised on a number of issues which include the non-blinding of the UG, a small number of examiners (n=2) which limits the inferences of the data, the poor description of the correlation coefficient (CCC) and the fact that they did not appear to have achieved their last aim, which was to investigate inconsistencies between dynamic and static gait analysis. In addition to these limitations, it is unclear why the authors choose to analyse the left and right lower limbs separately, particularly as only 4 of their subjects were hemiplegic. This latter methodology appears only to have confused and weakened the inferences of the data by lowering the degree of freedom. However, this study provides useful data regarding the reliability of clinical measurements of neurological patients.

The appropriate placement of the UG fulcrum over the joint axis has been questioned in the literature (Enwemeka, 1986; Nicol, 1989). To address this issue, an interesting study by Szulc et al (2000) attempted to investigate the distance between the bony landmark at which the axis of the goniometer is placed and the actual axis of the knee joint. The standard goniometer placement was used, as proposed by Norkin and White (1995), including the placement of the UG fulcrum at the lateral epicondyle of the knee. The distance from this point to the centre of the knee joint in different amounts of knee flexion was measured. The study also calculated the impact that the discrepancy between the 2 axes, i.e. the lateral epicondyle and the true rotational axis, may have on angular measurements. The study used 3 exposed knee joint preparations taken from cadavers. An electrogoniometer (EG) was attached directly to the bone shafts and used to assess angular position throughout the trial. The lateral epicondyle of the femur was marked with a nail and the joint was moved from 10° to 100° flexion through 10° increments. In each position, the knee was photographed and the developed image was assessed using a common protractor. It appears that the true joint centre was identified by assessments of the photographs. The distances from the lateral

epicondyle to joint centre was measured directly from the joint preparation. The author reported that the dynamic axis of the joint can be up to 4.81cm away from the lateral epicondyle and that this discrepancy increases as the knee is moved from extension into flexion. The author claims that this disparity in the apex of movement may relate to in excess of 15° difference in angular measurement, as calculated by correlations from the photographic and biaxial EG data.

The methodology of the study used measurements taken from a biaxial EG and photographs taken of the exposed cadaveric bones. From these photographs the position that the UG would be placed was supposed, i.e. no actual UG measurements were made during the study, rather the UG data was gained by its predicted position in the photographs. In addition, the author does not describe the methodology used to take the photographs. Dunlevy et al (2000) found that the photographic procedure is vital to ensure accurate measurement results are taken from the photographs. The fact that the measurements were taken from cadaver preparations minus soft tissue may also limit inferences to actual clinical measurement, as it is over soft tissue that both EG and standard UG measurements are made in normal practice.

Szulc et al (2000) found that in knee extension the chances of measurement error due to inappropriate positioning of the UG over the joint axis lessened, this finding is contrary to other clinical studies (Rothstein et al, 1983; Enwemeka, 1986; Clapper and Wolf, 1988; Watkins et al, 1991; Marks and Karkouti, 1996; Karkouti and Marks, 1997; Brosseau et al, 1997, 2001), these studies found that measurements of knee extension were less accurate than knee flexion measurements. The study illustrates the difficulties associated with using the lateral epicondyle as a fixed reference point for the centre of sagittal plane knee rotation, when in fact the knee axis is dynamic.

Researchers	Population	Joint(s) Assessed	Movement	Instrument	Intra-rater	Inter-rater
Hayes et al	Patient	Shoulder	Shoulder - Flexion	UG	ICC=0.53	ICC=0.69
(2001)	Intra-rater: n=9		Shoulder - Abduction		ICC=0.58	ICC=0.69
	Examiner=1	(External Rotation	Shoulder - Flexion	Photo	ICC=0.56	ICC=0.73
		not reported)	Shoulder - Abduction		ICC=0.61	ICC=0.73
	Inter-rater: n=8		Shoulder - Flexion	VE	ICC=0.59	ICC=0.70
	Examiner=4		Shoulder - Abduction		ICC=0.60	ICC=0.66
Armstrong et al	Patient	Upper Limb	Elbow - Flexion	UG	ICC=0.55-0.98	ICC=0.58-0.62
(1998)	n=38	(Supination/Pronation	Elbow - Extension	UG	ICC=0.45-0.98	ICC=0.58-0.87
	Examiner=5	not reported)	Elbow - Flexion	NK Goni	ICC=0.96 to 0.99	ICC=0.94 to 0.95
			Elbow - Extension	NK Goni	ICC=0.96 to 0.98	ICC=0.89 to 0.90
Leen T'Jonck et al	Healthy n=8	Shoulder	Shoulder	UG	R=0.79 to 0.89	R=0.88
(1997)	Patient n=15	(Only Active Abduction	Active Abduction	EDI 320	R=0.87 to 0.94	R=0.82
	Examiners=1 & 2	reported here)		(Electro)		
Goodwin et al	Healthy	Upper Limb	Elbow - Flexion	UG	r=0.79 (0.61-0.92)	n/a
(1992)	n=23				ICC=0.77 (0.56-0.91)	
	Examiner=3			FFG	r=0.72 (0.53-0.85)	n/a
					ICC=0.71 (0.50-0.84)	
				EG	r=0.36 (0.00-0.43)	n/a
					ICC=0.27 (0.00-0.43)	
Greene and Wolf	Healthy n=20	Upper Limb (only	Elbow - Flexion	UG	ICC=0.94	n/a
(1989)	Examiner=1	Elbow Reported Here)	Elbow - Extension	UG	ICC=0.95	n/a
Petherick et al	Healthy n=30	Upper Limb	Elbow - Flexion	UG	n/a	ICC=0.53
(1988)	Examiner=2			FFG	n/a	ICC=0.92
Riddle at al	Patient	Shoulder	Flexion	UG Ψ	ICC=0.98	ICC=0.89
(1987)	n=100 (intra)	End of Range	Extension	UG Ψ	ICC=0.94	ICC=0.27
	n=50 (inter)	Movements	Abduction	UG Ψ	ICC=0.98	ICC=0.87
	Examiner=16	(only Selection	Reported Here)			
Rothstein et al	Patient	Elbow	Flexion	UG	r=0.98; ICC=0.97	r=0.97; ICC=0.96
(1983)	n=24 (intra)		Extension	UG	r=0.99; ICC=0.96	r=0.95; ICC=0.94
	n=12 (inter)	Knee	Flexion	UG	r=0.99; ICC=0.99	r=0.91; ICC=0.92
	Examiner=12		Extension	UG	r=0.96; ICC=0.91	r=0.63; ICC=0.64
Mayerson &	Healthy	Major Joints	22 Joint positions	UG	r=0.99-0.98	r=0.97
Milano (1984)	n=1, Examiner=2	(unspecified)	(unspecified)			

Ψ Large and Small Universal Goniometer : UG= Universal Goniometer : n/a= Not Applicable : NK Goni = NK Goniometer : FFG= Fluid-filled Goniometer : PG= Parallelogram Goniometer : R= Reliability Coefficient : VE= Visual Estimation: EG= Electrogoniometer

Table 1.2. Correlation Coefficients for Upper Limb Joint Measurement Reliability.

Researchers	Population	Joint(s) Assessed	Movement	Instrument	Intra-rater	Inter-rater
Nigel-Keenan	Patient	Hip Extension	Thomas Test	UG	CCC=0.75 to 0.96	CCC=0.34 to 0.42
et al (2004)	n=20	Knee	Popliteal Angle	UG	CCC=0.86 to 0.92	CCC=0.84 to 0.86
	Examiner=2	Knee	Knee Extension	UG	CCC=0.82 to 0.91	CCC=0.67 to 0.75
		Ankle Dorsiflexion	Soleus Length	UG	CCC=0.72 to 0.81	CCC=0.64 to 0.71
			Gastroc. Length	UG	CCC=0.69 to 0.80	CCC=0.59 to 0.87
Russell et al	Healthy	Knee	Between	IBG (Photo)	ICC=1.0	ICC= 1.0 to 0.96
(2003)	n=1; Examiner=2		3° to 97°			
Brosseau et al	Patient	Knee	Flexion	UG	ICC=0.99	ICC=0.97 to 0.98
(2001)	n=60		(mean 118 ±20)	PG	ICC=0.99	ICC=0.97 to 0.96
	Examiner= 2		Extension (mean 8 ±6)	UG	ICC=0.97 to 0.98	ICC=0.92 to 0.89
				PG	ICC=0.95	ICC=0.89 to 0.85
Holm et al	Patient. n=27	Hip	Total Movement	UG &	ICC=0.85	n/a
(2000)	Examiner=4		6 ROMs	VE		
Karkouti and	Healthy	Knee	Extension (Full)	Photo	ICC=0.84	n/a
Marks (1997)	n=32		Flexion (approx. 30°)	Photo	ICC=0.87	n/a
	Examiner=1	Ankle	Neutral	Photo	ICC=0.81	n/a
			Dorsiflexion	Photo	ICC=0.71	n/a
Brosseau et al	Healthy	Knee	Large Flexion	UG	ICC=0.95-0.97	ICC=0.91-1.0
(1997)	n=60		Small Flexion	UG	ICC=0.86-0.94	ICC=0.62-0.71
	Examiner= 2		Large Flexion	PG	ICC=0.91-0.96	ICC=0.82-0.88
			Small Flexion	PG	ICC=0.85-0.87	ICC=0.43-0.52
Youdas et al	Patient	Ankle	Dorsiflexion	UG	ICC=0.82 (0.64-0.92)	ICC=0.28
(1993)	n=45			VE	n/a	ICC=0.34
	Examiner=10		Plantar Flexion	UG	ICC=0.86 (0.46-0.96)	ICC=0.25
				VE	n/a	ICC=0.48
Watkins et al	Patient	Knee	Flexion	UG	ICC=0.99	ICC=0.90
(1991)	n=43	approx 0° to 155°		VE	n/a	ICC=0.83
	Examiner=14		Extension	UG	ICC=0.98	ICC=0.86
				VE	n/a	ICC=82
Bovens et al	Healthy	Upper Limb & Foot	Ankle - Dorsiflexion	UG	R=0.82 to 0.87	R=0.76
(1990)	n=48	(only Selection	Ankle - Plantar Flexion	UG	R=0.64 to 0.75	R=0.63
	Examiner=3	Reported Here)	Wrist - Flexion	UG	R=0.27 to 0.87	R=0.59
			Wrist - Extension	UG	R=0.09 to 0.40	R= 0.09
Clapper and Wolf	Healthy	Lower Limb	Knee - Flexion	UG	ICC=0.95	n/a
(1988)	n=20	(only Ankle & Knee	Knee - Extension	UG	ICC=0.85	n/a
	Examiner=1	Reported Here)	Ankle - Flexion	UG	ICC=0.96	n/a
			Ankle - Extension	UG	ICC=0.92	n/a
Rheault et al	Healthy n=20	Knee	Flexion	UG	n/a	r=0.87
(1988)	Examiner=2			FG	n/a	r=0.83
Gogia et al	Healthy n=30	Knee	0 to 120	UG	n/a	r=0.98; ICC=0.99
(1987)	Examiner=2					
Walker et al	Healthy n=4	Upper Limb	24 ROMs (Averaged)	UG	r=0.81 (0.78 - 0.99)	n/a
(1984)	Examiner=2	& Lower Limb				
Boone and Azen	Healthy n=12	only Elbow & Knee	Knee Flexion	UG	ICC=0.87; ICC=0.94	ICC=0.50; ICC=0.88
(1979)	Examiner=4	Reported Here	Elbow - Flexion			

UG= Universal Goniometer : n/a= Not Applicable : NK Goni = NK Goniometer : FFG= Fluid-filled Goniometer : r=PPMCC : PG= Parallelogram Goniometer: R= Reliability Coefficient : CCC= Concordance Correlation Coefficient : ICC= Intraclass Correlation Coefficient

Table 1.3. Correlation Coefficients for Lower Limb Measurement Reliability.

Researchers	Population	Joint(s) Assessed	ROM (Degrees)	Instru-ment	Intra-Rater	Inter-Rater
Hayes et al (2001)	Patient Intra-rater: n=9 Examiner=1 Inter-rater: n=8 Examiner=4	Shoulder (External Rotation not reported)	Shoulder - Flexion Shoulder - Abduction Shoulder - Flexion Shoulder - Abduction Shoulder - Flexion Shoulder - Abduction	UG Photo VE 	SEM= 17°; 95%CI=34° SEM= 23° ; 95%CI=46° SEM= 19° ; 95%CI=38° SEM= 22° ; 95%CI=44° SEM= 13° ; 95%CI=26° SEM= 21° ; 95%CI=42°	SEM= 25° ; 95%CI=50° SEM= 21° ; 95%CI=42° SEM= 23° ; 95%CI=46° SEM= 23° ; 95%CI=46° SEM= 19° ; 95%CI=38° SEM= 19° ; 95%CI=38°
Burton et al (1999)	Healthy n=1 Examiner=40	Hand Metacarpo-phalangeal	Flexion (approx. 40°) Extension (approx. 8°)	UG VE UG VE	n/a n/a n/a n/a	±1.1° RC 4.4° ±15.7° RC 44.5° ±1.7° RC 5.9° ±18° RC 51°
Armstrong et al (1998)	Patient n=38 Examiner=5	Upper Limb (Supination/Pronation not reported)	Elbow - Flexion Elbow - Extension Elbow - Flexion Elbow - Extension	UG UG NK Goni NK Goni	95% CI = 5.9° 95% CI = 6.6° 95% CI = 5.2° 95% CI = 6.3°	95% CI = 9.5° 95% CI = 10.2° 95% CI = 4.1° 95% CI = 6.1°
Leen T'Jonck et al (1997)	Healthy n=8 Patient n=15 Examiners=1 & 2	Shoulder (Only Active Abduction reported here)	Shoulder Active Abduction	UG EDI 320 (Electro)	SEM=2.6°: SD ± 10.1° SEM=2.8°: SD ± 11°	n/a n/a
Bovens et al (1990)	Healthy n=48, Examiner=3	Upper Limb & Foot	9 Joint Movements	UG	Mean SD ± 4.8°(2.5° to 8.1°) Expected Variation ±10°	
Williams & Callaghan (1990)	Healthy n=1 Examiner=22	Shoulder	Flexion	UG VE	n/a n/a	105°±5.91° to 102.27°±11.48° 103.63°±8.5° to 106.36°±11.85°
Greene and Wolf (1989)	Healthy n=20, Examiners=1	Upper Limb - All Movements (only Elbow Reported Here)	Elbow - Flexion Elbow - Extension	UG	95% CI = 3.0° 95% CI = 1.9°	n/a n/a
Fish & Wingate (1985)	Healthy n=1, Examiners=46	Elbow	Flexion 135° Exention 50°	UG UG	n/a n/a	SD 3.8° (2SD=7.6°) SD 4.2° (2SD=8.4°)
Walker et al (1984)	Healthy n=4, Examiner=2	Upper Limb & Lower Limb	24 ROMs (Averaged)	UG	Mean Error & SD 5°±1°	Mean Error & SD 6°±5°
Mayerson & Milano (1984)	Healthy n=1, Examiner=2	Major Joints (unspecified)	22 Joint positions (unspecified)	UG	Mean Error -0.9° 95% CI -2.3° to 4.1°	Mean Error 1.7° 95% CI -0.9° to 4.3°

Mean Error (without Standard Deviation => Caution) * Calculated from Data in Published Article
SD (±) = Standard Deviation CI= 95% Confidence Intervals CV= Coefficient of Variance
RC=Repeatability Coefficient LOA = Limits of Agreement. UG=Universal Goniometer
VE= Visual Estimation PG= Parallelogram Goniometer NK Goni = NK Goniometer

Table 1.4. Reliability of Upper Limb Measurement Methods (Reported without using Correlation Coefficients).

Researchers	Population	Joint(s) Assessed	ROM (Degrees)	Instru-ment	Intra-Rater	Inter-Rater
Nigel-Keenan et al (2004)	Patient n=20 Examiner=2	Hip Extension	Thomas Test	UG	95% LOA^= -7.8°, 8.3°	95% LOA^= -13.9°, 4.4°
		Knee	Popliteal Angle	UG	95%LOA^= -12.3°, 12.3°	95%LOA^= -13.6°, 7.8°
		Knee	Extension	UG	95% LOA^= -6.5°, 5.0°	95% LOA^= -12°, 5.0°
		Ankle Dorsiflexion	Soleus Length	UG	95% LOA^= -9.1°, 12.1°	95% LOA^= -10.7°, 13°
			Gastroc. Length	UG	95% LOA^= -10.5°, 9.5°	95% LOA^= -12.1°, 5.1°
Russell et al (2003)	Healthy n=1; Examiner=2	Knee	Between 3° to 97°	IBG (Photo)	95% LOA = -1.0° to 1.6° (Average)	95% LOA = -1.0° to 1.4° (Average)
Sprigle et al (2003)	Healthy n=20; Examiner=8	Hip and Pelvis	Flexion 3 Sitting Position	UG Pelvic Gonio	RC=6.32° RC=5.57°	95% LOA=-10°, 14.6° 95% LOA=-3.9°, 8.4°
Brosseau et al (2001)	Patient n=60 Examiner= 2	Knee	Flexion (mean 118 ±20)	UG PG VE	Mean Err* 0.33°- 0.08° Mean Err* 0.64°- 0.33° N/a	Mean Error* 2.4° Mean Error* 4.44° Mean Error* 4.6°
			Extension (mean 8±6)	UG PG VE	Mean Err* 0.5° to 0.2° MeanErr.* 0. 5°to0.17° n/a	Mean Error* 0.12° Mean Error* 0.03° Mean Error* 0.37°
Holm et al (2000)	Patient n=27 Examiner=4	Hip	Total Movement 6 ROMs	UG & VE	CV= 7% 95% CI= 5.5° to 26.1° RC= 9.5° to 3.9°	n/a n/a n/a
			Flexion	UG	* 0.01° to 0.05°	* 0.05°
Brosseau et al (1997)	Healthy n=60 Examiner= 2	Knee	(approx. 103°)	PG	Mean Err * 0.13° to 0.16°	Mean Err * 0.59°
			Flexion	UG	Mean Err * 0.25° to 0.28°	Mean Err * 0.07°
			(approx. 16°)	PG	Mean Err * 0.16° to 0.08°	Mean Err * 0.21°
Karkouti and Marks (1997)	Healthy n=32 Examiner=1	Knee	Extension (Full)	Photo	SEM=2.09°	n/a
			Flexion (approx. 30°)	Photo	SEM=1.65°	n/a
		Ankle	Plantar Grade	Photo	SEM=1.92°	n/a
			Dorsiflexion	Photo	SEM=1.63°	n/a
Youdas et al (1993)	Patient n=45, Examiner=10	Ankle	Dorsiflexion	UG	SD Median 3.27°	n/a
			Plantar Flexion	UG	SD Median 6.04°	n/a
Watkins et al (1991)	Patient n=43 Examiner=14	Knee approx 0° to 155°	Knee - Flexion	UG	Mean Error 0° to 1°	Mean Error 3°
				VE	n/a	Mean Error 4°
			Knee - Extension	UG	Mean Error 0° to 1°	Mean Error 1°
				VE	n/a	Mean Error 2°
Croxford et al (1989)	Healthy n=1 Examiner=12	Ankle	Dorsiflexion	UG	n/a	95% CI=5.5° CV= 34.8%
				VE	n/a	95% CI=11.1° CV= 69.4 - 53%
Rheault et al (1988)	Healthy n=20, Examiner=2	Knee	Flexion	UG	n/a	Mean Error 2.85°
				FFG	n/a	Mean Error 2.47°
Gogia et al (1987)	Healthy n=30, Examiner=2	Knee	0 to 120	UG	Mean Error* 1.64°	n/a

Mean Error (without Standard Deviation => Caution) * Calculated from Data in Published Article

SD (±) = Standard Deviation CI= 95% Confidence Intervals CV= Coefficient of Variance

RC=Repeatability Coefficient LOA = Limits of Agreement. UG=Universal Goniometer

VE= Visual Estimation PG= Parallelogram Goniometer NK Goni = NK Goniometer

(L)=Left Hands Side IBG=Internet Based Goniometer ^ = Worst Published Data

Table 1.5. Reliability of Lower Limb Measurement Methods (Reported without using Correlation Coefficients).

Reasearcher	Population	Joint	Position	Instru-ment	Gold Stanard	Correlation Coefficient	Mean Error	Stanard Deviation
Sprigle et al (2003)	Healthy n=20 Examiner=8	Hip and Pelvis	Flexion 3 Sitting Position	Pelvic Goniom.	Vs X-ray	r=0.93	95% LOA 18° , +8°	
Brosseau et al (2001)	Patient n=60	Knee	Flexion (mean 118 ±20)	UG	Vs X-ray	r=0.97 to 0.98	*1.59°	n/a
				PG	Vs X-ray	r=0.97 to 0.98	*1.02°	n/a
				VE	Vs X-ray	r=0.91 to 0.94	*1.36°	n/a
			Extension (mean 8 ±6)	UG	Vs X-ray	r=0.39 to 0.44	*3.82°	n/a
				PG	Vs X-ray	r=0.42 to 0.51	*2.75°	n/a
				VE	Vs X-ray	r=0.24 to 0.36	*2.64°	n/a
Brosseau et al (1997)	Healthy n=60	Lower Limb	Knee - Large Flexion	UG	Vs X-ray	r=0.73-0.78	*2.4°	n/a
			Knee - Small Flexion	UG	Vs X-ray	r=0.33 to 0.48	*6.85°	n/a
			Knee - Large Flexion	PG	Vs X-ray	r=0.73-0.77	*4.21°	n/a
			Knee - Small Flexion	PG	Vs X-ray	r=0.33-0.41	*8.1°	n/a
Gogia et al (1987)	Healthy n=30	Knee	0 to 120	UG	Vs X-ray	r=0.97-0.98 ICC=0.98-0.99	2.37° (1.64°-3.07°)	n/a
Enwemeka (1986)	Healthy n=10	Knee	Flexion	UG	Vs X-ray	r=0.99	2.32°	±7.2°
			Extension				4.59°	±4.76°
			From 0° to 90°				*2.52°	*±6.15
Szulc et al (2000)	Cadaver n=3 (measures=30)	Knee	Extension to Flexion 10° to 100°	UG Φ	Vs EG	n/a	* 6.8° Φ	* 2.3° Φ
				Photo	Vs EG	n/a	* 4.3°	* 1.5°
Ishii et al (1995)	Patient n=50 (measures= 100)	Knee	Tibiofemoral Angle (Coronal Plane)	X-ray (Short &	Vs X-ray Long Plate)	r=0.82 - 0.97	0.5°	± 1.2°
Rawes et al (1996)	Healthy n= 1 (measures= 20)	Wrist	Combined Movements	EG	Vs Fabricated Structure	n/a	0.2°	±0.6°
Ojima et al (1991)	Healthy & Patient n=13	Wrist	Combined Movements	EG	Vs Fabricated Structure	n/a	2.1° (Max Error 7°)	

* Calculated from Data in Published Article Φ Theoretical Calculation of Goniometric Position
UG= Universal Goniometer : n/a= Not Applicable : EG = Electrogoniometer :
PG= Parallelogram Goniometer: r=Pearsons Product Moment Correlation coefficient:
ICC= Intraclass Correlation Coefficient

Table 1.6. Validity of Goniometric Measurement Methods.

1.2.2 Summary of Universal Gonometer.

The fact that the UG is a cheap, user-friendly device is not disputed in the literature. Coxford et al (1998) and Burton et al (1999) agree that the UG is probably the most widely used measurement tool in clinical practice. Different variations of this tool which include different sizes, UGs fabricated from different materials and slight variations in design, such as the parallelogram goniometer (PG) (Brosseau et al, 1997) and the pelvic goniometer (Sprigle et al, 2003) have been investigated. From these investigations there appears to be some evidence that the UG performs slightly better than the PG, however, not as well as the pelvic goniometer (for hip joint measurements).

Norkin and White (1995) proposed standard operational procedures for the UG including specific guidelines for each of the measurable joints in the body. The literature has investigated several different approaches to applying the UG, including those proposed by Norkin and White (1995), but has not shown that one approach is clearly superior to any other approach (Grohmann, 1983; Szulc et al, 2000). However, the literature has shown that when a consistent procedure is used reliability will be improved (Rothstein et al, 1983). There is also some agreement as to the increased difficulties associated with measuring knee extension using a UG (Rothstein et al, 1983; Enwemeka, 1986; Brosseau et al, 1997, 2001), although there is no clear reason for this phenomenon. It has been suggested that the inability of the UG to cope with the joints' dynamic joint centre may be to an influence (Enwemeka, 1986; Norkin and Levangie, 1992; Szulc et al, 2000).

Some studies have only reported fair or weak correlation coefficients for the UG for either intra-rater (Armstrong et al, 1998; Hayes et al, 2001) or inter-rater reliability (Boone and Azen, 1979; Riddle et al, 1987; Petherick et al, 1988; Bovens et al, 1990; Youdas et al, 1993; Armstrong et al, 1998; Hayes et al, 2001), however the majority seem to agree that when reliability and validity are assessed using correlation coefficients results will be high (see Table 1.2 to 1.5).

In the main studies have used inadequate data analysis methods which has limited the information about the expected performance of the UG. There remains controversy as to the expected error associated with the UG. This lack of conclusive knowledge regarding validity and reliability of the UG presents problems comparing it against the performance of new or existing measurement tools against it.

Further studies that have investigated the performance of the UG in addition to other measurement tools will be discussed in the following sections.

1.2.3 Inclinometers.

Rheault et al (1988) investigated the validity and reliability of the fluid filled goniometer (FFG). The study used 2 examiners, who each repeated measurements of knee flexion on a population of 20 healthy volunteers with both a FFG and a standard universal goniometer (UG). Validity was investigated by correlating data from the FFG against data from the UG, using the UG as a gold-standard. Inter-rater reliability was assessed by correlating the datasets between the 2 examiners. The data was analysed using the PPMCC, which showed a high inter-rater reliability for both the UG ($r=0.87$) and FFG ($r=0.83$). Claimed validity was also high when the data was assessed using the PPMCC ($r=0.82$). However, when all the data was analysed using a paired t-test, statistically significant differences were seen ($p<0.05$) for the UG and FFG measurements. The authors conclude that both UG and FFG are both valid and reliable, but warn that these measurements should not be used interchangeably. The disagreement in the results between the statistical tools of the PPMCC and the t-tests underlines the importance of appropriate statistical analysis and methodology.

In a similar study, Petherick et al (1988) undertook a trial to establish concurrent validity for UG and a fluid-filled goniometer (FFG), and to establish inter-rater reliability for both devices. Only full elbow flexion was measured, using a population of 30 healthy subjects. Two physiotherapists were employed to conduct the examinations, which were repeated 3 times. The examiners were instructed to follow the same measurement procedures as those described by Norkin and White (1995). During the assessments each subject was positioned in supine with their shoulder stabilised and a towel placed under the distal humerus. When each examiner was happy with their UG placement, the blinded instrument was read by a different examiner.

Data analysis was conducted using PPMCC, ICC (2,3) ('2,3' indicates the specific ICC formula used) and the paired t-test. The latter test was used in an attempt to establish concurrent validity. The results of this inter-device analysis showed significant difference between devices ($p<0.05$), despite a relatively high

PPMCC result for measurements between devices (r=0.83). This conflict in results between the PPMCC and the t-test is similar to that seen by Rheault et al (1988) above. Inter-rater reliability of elbow flexion for both devices was assessed using the ICC (2,3). This data suggests a poorer result for UG (ICC=0.53) than for FFG (ICC=0.92). No intra-rater reliability investigation was attempted.

These investigators only assessed one joint and joint position which may limit the implications of the study. The authors concede that further validity and clinically relevant trials are required for both FFG and UG.

Both of the above studies (Rheault et al, 1988; Petherick et al, 1988) claimed to have assessed concurrent validity of both the UG and the FFG, however, this claim seems inappropriate as no gold-standard measurement method was used as a reference against which to gauge validity (Norkin and White, 2003). As has already been discussed there is much controversy towards the expected error associated with the UG, which would certainly seem to undermine its roll as the gold-standard for these studies.

1.2.4 Summary of Inclinometers.

Norkin and White (1995) agree with Rheault et al (1988) and Petherick et al (1988) when they report that the FFG may be a useful tool, but it should not be used interchangeably with a UG or any other type of goniometer. The AAOS (1994) criticise the FFG commenting that it is generally more expensive, relatively bulky and may be easily broken. Nonetheless they hold that FFGs are useful in measurement in some areas such as the spine and hindfoot, however, this claim does not appear to have been tested in the literature.

1.2.5 Visual Estimation.

Visual estimation (VE) is the practise of assessing joint angles simply by looking at the joint. This is a subjective method of assessment and has been investigated by several authors. Watkins et al (1991) state that therapists will often use VE in lieu of other goniometric methods, which is probably due to the speed of these assessments.

In their study they investigated the reliability of VE and UG of knee ROM in a clinical setting. The population comprised of 43 patients, with a total of 50 knees being assessed. Fourteen experienced physical therapists were then randomly paired, so that 2 therapists measured each knee. This presented data on a total of 100 knee measurements. Each therapist assessed each knee 3 times, firstly by VE then re-assessed twice using a blinded UG.

Statistical analysis was carried out using a different ICC formula than that which was used in similar previous studies. In fact one of the justifications for the study which was put forth was to use the ICC (1,1) rather than the ICC formula that was used by Rothstein et al (1983). Although neither of the published articles (Rothstein et al (1983) or Watkins et al (1991)) offers the original formula used by Rothstein et al (1983). Watkins et al (1991) states that the importance of the choice of ICC (1,1) over other ICC formulas is that the other formulas underestimate measurement error.

Intra-rater reliability was excellent for UG measurements for both flexion (ICC=0.99) and extension (ICC=0.98). Inter-rater reliability was also high for UG for flexion (ICC=0.90), while knee extension was slightly lower but remained high (ICC=0.86). Only the inter-rater reliability of VE was investigated for practical reasons. Significant examiner bias would have been introduced on repeated VE assessment due to the inevitable recall by the examiners post initial assessment. Statistical analysis revealed that while VE was good (flexion ICC=0.83, extension ICC=0.82), the correlations were less than measurements made with a UG (flexion ICC=0.90, extension ICC=0.86).

Descriptive analysis of the data reveal that the average intra-rater error associated with UG measurement was on average only 1°, while inter-rater measurements had a mean difference of approximately 3°(no standard deviations for repeated tests are reported). Visual estimation compared with UG measurements showed a mean difference of approximately 4°. The authors conclude that VE adds more error to the measurement process, and that UG is a more reliable way of assessing knee joint ROM. However, due to the lack of means and standard deviations (SD) for all data inferences towards expected agreement are not possible.

Williams and Callaghan (1990) investigated VE and UG. The aim of this study was to compare measurements made with VE and 3 different types of UG (long arm, short arm and digital read-out), and to compare measurements made between practised (experienced) and non-practised physiotherapists (limited experience). Two groups of 11 physiotherapists were used as examiners for the study. The first group regularly conducted joint angle assessment, while the second group did not regularly assess joint angles as part of the career speciality. One healthy male subject was used as the subject. This person stood with his heels, back and head against a wall, with his elbow extended, he had his wrist strapped to an over-head cage, leaving his shoulder held in passive flexion at about 100°. Each examiner firstly assessed the shoulder flexion angle visually, then used each of the 3 UGs to measure the same joint angle. Data was analysed using a 2-way ANOVA, means and SD.

No significant differences were found between any of the measurements for any of the groups. The largest mean difference between any of the measurement cohorts were between VE and the short arm UG for the practised group (4°), while SD for each of the methods ranged between ±11.85° (VE) to ±4.03° (digital read-out UG).

The authors concede that the study design is rudimentary, only assessing 1 joint angle from 1 subject. The study ignored intra-rater reliability, did not subject its data to adequate analysis and did not blind the measurement tools. This latter issue must be seen as a major source of methodological error as examiner bias would seem to have been inevitable. In addition, the joint position chosen for the study was very close to the horizontal, which may have helped the examiners gain a reference against which to make the VE. In fact, the authors state that in a limited separate study, when student physiotherapists were asked how they would approach VE they stated that they would try to visualise a right angle over a joint and from that reference line they judged the approximate position of the joint. This suggests that a joint positioned close to a right angle may be relatively easier to assess using VE.

The authors conclude that VE is probably as reliable as standard UG measurements, a statement that seem contrary to other researchers and was not rigorously tested by their methodology. Their results also did not support this

conclusion as the SD of the differences which relate to random error and agreement were almost 3 times as high for VE over UG. However, one apparently relevant point put forth by the authors is that VE is not just about quantifying angles, but that the practice is important to the qualitative aspect of the assessment.

Youdas et al (1993) investigated the reliability of UG and VE in measuring ankle dorsi-flexion (ADF) and ankle plantar flexion (APF). They recruited 38 patients with orthopaedic disorders for their study, assessing a total of 45 ankles. A random pairing technique was employed to couple 10 physical therapists for the purpose of assessments. This method is similar to that used in previous studies (Rothstein et al, 1983, Watkins et al, 1991). The interclass correlation coefficient (ICC) (1,K) was used to assess the data.

Intra-rater reliability for UG varied for ADF (ICC=0.64 to ICC=0.92) and APF (ICC=0.47 to ICC=0.96). Inter-rater reliability for UG was low for both ADF (ICC=0.28) and APF (ICC=0.25). The inter-rater reliability for VE seemed to be slightly better for both ADF (ICC=0.34) and APF (ICC=0.48), but still remained low. The authors hold that examiners should use a goniometer when making repeated measurements of the ankle joint and that considerable inconsistency exists when two or more examiners make repeated goniometric and visual measurements of ankle motion on the same subject. This study showed poorer reliability for UG and VE than those displayed by Watkins et al (1991), however, different joints were assessed in both of these studies. Youdas et al (1993) conclude that therapists using VE may erroneously conclude that a patient's ankle ROM has changed because of treatment when the change could be attributed to a lack of VE measurement reliability. These authors did not discuss the complicated arthrokinematic nature of ankle joint or account for how the complex joint dynamics of ankle movement may influence the measurement process.

A study by Coxford et al (1998) was designed to test the expected reliability of UG and VE in a clinical setting as opposed to the tightly controlled setting of seen in some of the research literature. Twelve experienced physiotherapists were used to carry out measurements of active ankle dorsiflexion on 1 healthy subject. A small masked half circle goniometer was used for the investigation. Each examiner carried out both VE and UG measurements once. The subject rested for 3 minutes between

each measurement to prevent fatigue. Each examiner was asked assess the subject in the position that they would normally use as opposed to guidelines or instruction give by the researcher, i.e. the subjects measurement position was not tightly controlled by the studies methodology.

Data analysis involved the F ratio test, t-test, Wilcoxon rank order test (non-parametric), coefficient of variance (CV) and 95% confidence interval (CI), as well as measures of central tendency and dispersion (descriptive). The results show an average difference between UG and VE as 4.2°. The standard deviation (SD) for VE between examiners was ±8.75, while UG had a lower SD of ±4.38, which indicates better agreement between testers for the UG. This is seen clearly when the range of measurements are inspected (UG=5°-20°; VE=5°-30°), data which shows an increase in VE data range of almost 50%. This increase is reflected in the CV data, with UG returned a CV of 34.8% and VE showed a CV of 69.4%. The 95% CI calculations suggested that UG should expect variations of between 5.5°, while VE should expect variations between examiners of between 11.1°.

This study confused results by terming the inter-method reliability investigation as concurrent validity. From the results of this investigation and the fact that no significant difference was seen between the two methods, the authors feel that both approaches could be used interchangeably, despite the actual clinically relevant differences seen in the results. It would seem that the claim of interchangeable devices is poorly supported by their results, as is the usefullnes of VE.

Burton et al, (1999) conducted a study that compared reliability of VE and UG for assessment of the metacarpophalangeal joint (MCP). This investigation used 40 different examiners, 20 physiotherapists and 20 occupational therapists. One healthy subject was used in the study. The subject had their hand splinted in 2 different positions, one relatively flexed and the other was relatively extended. A small goniometer (Roylan, manufactured by Smith and Nephew, Hull, UK) was used during the measurements, as opposed to a standard large joint UG. The scale of the small UG was blinded. Each examiner firstly assessed with VE, then used the small UG.

Mean measurement for the first position for VE was 42.1°±15.7°, while UG had a much lower SD (44.5°±1.1°). The second position showed similar poor results for VE (mean 9.4°±18°) and similarly strong results for UG (mean 7.8°±1.7°). These results show massive differences between the ranges and SD between the UG and VE for inter-rater reliability and mean that consistency for UG measurements are much better than VE. This finding was reflected statistically when the data was analysed using the RC. This analysis reveals inter-rater reliability for VE having a RC=44.5° to 51°, while the UG was again much lower with an RC=4.4° to 5.9°.

The investigation revealed that there was no differences in the reliability of the measurements between the different occupations, years of experience or clinical speciality. The authors propose that VE carries an end digit bias, that is the examiner is more likely to round to the nearest 5°. The authors conclude that it is unwise for anyone to use VE as an objective means of joint angle assessment.

Holm et al (2000) aimed to evaluate inter-rater and intra-rater reliability and agreement between both VE and UG. The study population were 25 patients diagnosed with hip *osteoarthrosis*. Six examiners were used to conduct the assessments, 5 physiotherapists and 1 orthopaedic surgeon. The examiners were divided into 4 groups, some paired and some single. The fourth team, which comprised of the orthopaedic surgeon only, did not use a UG, but assessed purely using VE. A steel half circle UG with a scale marked in 5° increments was used for all the goniometric measurements. This device was not blinded. Six hip movements were assessed; hip flexion, extension, abduction, adduction, external rotation and internal rotation. Each team assessed all measurements on two occasions at 1-week intervals.

Extensive data analysis included means, SD, a paired t-test, ANOVA tables, ICC, PPMCC, RC, coefficient of variance (CV) and 95% confidence intervals (CI).

The results of the study showed significant differences for measurement of hip abduction between the first and second measurement occasions. There were also differences between measurements made with 1 examiner and those made with a team of 2 examiners, and between the visual estimator and one of the other groups. The repeatability coefficient (RC) showed variations in measurements of between 9.5° for hip flexion (ICC=0.76) and 3.9° for hip extension (ICC=0.83). When all

measurements for all the teams were totalled the resulting RC was 21.5°, however, this figure includes both VE and UG. Overall VE correlated well with UG yielding an ICC=0.80.

The authors state that hip abduction has the worst repeatability for measurement at the hip, while measurement of hip flexion showed best scores. Measurements carried out by one person are different to those carried out by two people, a finding the authors propose is due to the difficulty in stabilising the pelvis when measurements are conducted single-handedly.

Despite using an extensive range of data analyses these researchers did not provide enough information regarding the reliability of VE compared to the UG. Insufficient analysis was conducted for inter-rater reliability and there was poor isolation of data between VE and UG.

Brosseau et al (2001), as discussed previously (1.1.4) investigated reliability and validity of parallelogram goniometry (PG), universal goniometry (UG) and visual estimation (VE). Their study assessed a population of 60 patients with residual musculoskeletal impairments of the knee. Two examiners visually estimated knee joint angles in relative flexion and in relative extension. The subjects were also X-rayed in each position. Validity was statistically tested using PPMCC, which revealed a high correlation between VE and X-ray measurement. This implied a high level of validity for VE assessment for knee flexion (r=0.91 to r=0.94), but was not sustained for measurement of knee extension (r=0.24 to r=0.36). When the VE data was subjected to PPMCC analysis results ranged from r=0.24 to r=0.94, while similar analysis for PG and UG ranged from r=0.89 to r=0.99. These results illustrate that measurement by either UG or PG are significantly more valid than VE measurements. The authors conclude that the UG or PG are better methods of joint angle measurements than VE.

1.2.6 Summary of Visual Estimation

Many authors agree that VE is probably the most commonly employed approach towards the assessment joint angles (Watkins et al, 1991, Youdas et al, 1993, Norkin and White, 1995). Some authors feel that it is an acceptable approach to the science of joint angle measurement (Williams and Callaghan, 1990; Coxford

et al, 1998) while others feel that it is a poor method of objectifying joint angles (Watkins et al, 1991; Youdas et al, 1993; Burton et al, 1999; Brosseau et al, 2001). These latter authors appear to have based their findings on stronger research methodologies and therefore it appears there is more evidence to support limited validity and reliability for angular measurements made using VE alone. Croft et al (1994) give reasons of systematic variation in examination technique and random variation in visual interpretations for the flawed reproducibility of VE of joint angles.

1.2.7 Electrogoniometers.

Electrogoniometers (EG) were first introduced by Karpovich and Karpovich in 1959 (cited by Nicol, 1989). EG's work by converting angular joint motion into digital information using one of a number of different electronic principles. In 1989, Nicol stated that the most basic of these devices is the simple electric potentiometer, a device that is placed at the joint so that its 2 goniometric arms are adjacent to the body segment, proximal and distal to the joint. Motion at the joint causes movement of the measurement arms and so a change in resistance at the potentiometer. This change in resistance can be calibrated to indicate a specific joint angle. Ojima et al (1991) claim that the potentiometer EG is difficult to align to the centre of the joint and is too cumbersome to allow smooth normal movement.

The 2 other common types of electrogoniometry are the digital inclinometer and the electronic strain-gauge. These devices are more advanced than the simple electric potentiometer. The digital inclinometer is based on the conventional pendulum type goniometer and is gravity dependant (see Figure 1.3). The electronic strain-gauge is sometimes known as the 'Penny and Giles EG' or the 'Flexible EG'. The electrical properties of this device result in a change in electrical resistance as the device is deformed. Different amounts of bend show different amounts of electrical resistance, therefore a specific change in bend angle of the goniometer can be calibrated to display specific angles (see Figure 1.2). Both the digital inclinometer and the original electronic strain-gauge only allow single plane movement examination. However, more than one device can be attached to the skin at the same time, thus allowing simultaneous multi-plane movement analysis. Nicol (1989) claims that the electronic strain-gauge or flexible EG allows natural

movement analysis because it can be comfortably worn next to the skin underneath clothes. Analysis of unrestricted multidirectional movement can be very useful as it enables in-depth functional movement analysis. Systems like this have been developed and used to gain valuable information regarding gait and other functional activity analysis (Nicol, 1989). Therefore, there is a divergence in the application of EGs for either ROM assessments or functional movement analysis.

The Orthoranger electrogoniometer is an inclinometer type EG. This device has as an encased, gravity dependant electronic pendulum, which is attached to the skin at the relevant body segment (see Figure 1.3). As the limb moves, the pendulum slides across a series of electrical resisters that are calibrated to output angular data. This device was investigated by a number of authors (Clapper and Wolf, 1988 Greene and Wolf, 1989; Stude, 1994).

Clapper and Wolf (1988) investigated the reliability of the Orthoranger and UG for measurements of hip flexion, extension, abduction, adduction, lateral and medial rotation, knee flexion and extension, and ankle plantar flexion and dorsiflexion were also assessed. A total of 5 measurement sessions were repeated with 3 measurements at each joint per session.

They found that repeated measurements made with the UG and EG for knee flexion were better than those for knee extension (UG = flexion ICC=0.95; knee extension ICC=0.85: EG = flexion ICC=0.91; extension ICC=0.80). This finding is in agreement with other studies which employed differing methods of osteokinematic measurement (Rothstein et al, 1983; Enwemeka, 1986; Watkins et al, 1991; Marks and Karkouti, 1996; Karkouti and Marks, 1997; Brosseau et al, 1997, 2001). While both intra-rater reliability results are high, UG faired better over all. Results for all the joints tested showed that repeated UG measurements (ICC=0.80 to ICC=0.96) faired better than repeated Orthoranger EG measurements (ICC=0.72 to ICC=0.93).

In conclusion, the authors state that when compared with a standard UG, the added costs of purchase and increased time needed to use for the Orthoranger EG, appear to make this an unnecessary tool which has questionable accuracy.

Greene and Wolf (1989) further studied the 'within session' reliability of the Orthoranger EG and UG. This study assessed shoulder flexion, extension, abduction, adduction, internal and external rotation; elbow flexion and extension;

wrist flexion, extension, pronation, supination, and ulnar and radial deviation, in 20 healthy volunteers. Similar to the previous study, they assessed each subject 3 times during each session and repeated these measurement procedures over 3 different sessions.

Intra-rater reliability was high for the UG in all of the upper limb ROM measurements (ICC=0.91 to ICC=0.98). It was also high for the Orthoranger (ICC=0.84 to ICC=0.97). The intra-rater reliability of the Orthoranger EG did not score better than the UG on any of the individual measurements. The PPMCC results comparing the 2 data sets varied (r=0.11 to r=0.92) which suggests a limited relationship between the Orthoranger EG and the UG. These results compare well with the earlier Clapper and Wolf (1988) study, although the exact ICC formula is not published. The authors conclude that the Orthoranger EG does not provide any advantages over UG for joint measurement science.

Stude (1994) studied the more advanced Orthoranger II EG. This device had some technical and design changes to its predecessor the Orthoranger EG. The study investigated the intra-rater and inter-rater reliability of this digital inclinometer at the lumbar spine. Twenty-eight asymptomatic people were recruited for the study. Two examiners measured each subject twice, for the lumbar flexion, extension and right and left lateral flexion. The ICC revealed a low level of reliability for this device, except when measuring lumbar flexion. The rest of the movements assessed for reliability did not meet the studys recommended minimum value of R=0.75. The authors felt that the device failed to meet the minimum statistical criteria due to the intrinsic limitations of the instrument design. The study concludes that the Orthoranger II digital inclinometer does not show adequate intra-rater or inter-rater reliability for measuring lumbar spine ROM.

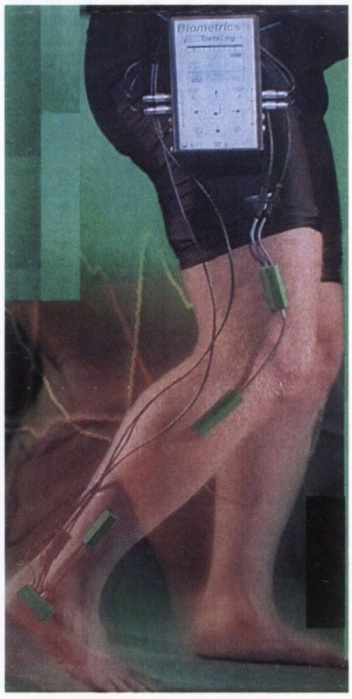


Figure 1.3. The Flexible/ Strain Gauge Goniometer.

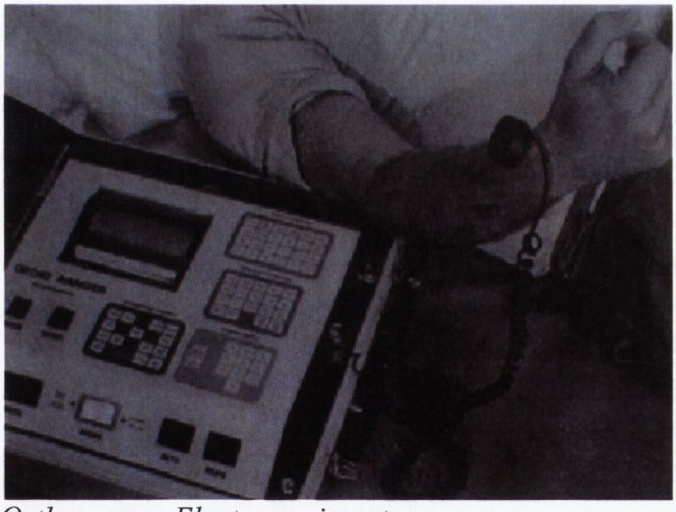


Figure 1.4. The Orthoranger Electrogoniometer.

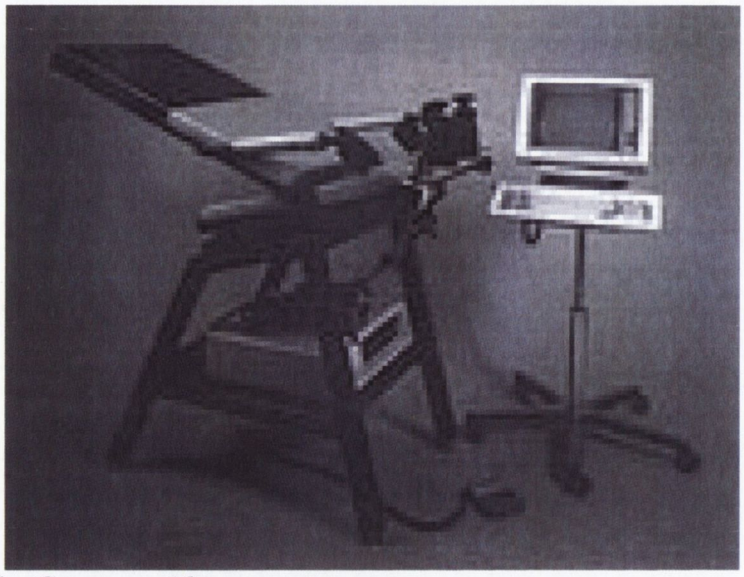


Figure 1.5. The Genucom Electrogoniometer.



Figure 1.6. The NK Hand Assessment Laboratory (Short Arms).

Another EG, the Cybex EDI 320 electronic digital inclinometer (EDI), wich is also based on the inclinometer model was investigated by Leen T'Jonck et al (1997). This investigation examined the reliability of the UG and EDI. Both of these devices were used to assess ranges of active and passive shoulder movement. The shoulder movements measured were active and passive elevation, abduction, passive horizontal adduction, passive internal and external rotation. The population they used for the investigation was a mixed population of both patient (15) and healthy (8) subjects. The patient population were in-patients who were assessed prior to shoulder surgery, these patients (23-66 years) were used to test the intra-rater reliability of 1 examiner using both of the devices. The healthy population were 8

subjects (age range 20-24 year) all of whom were right handed and were engaged in athletics favouring the dominant extremity. This population was used to investigate the inter-rater reliability of the devices between 2 examiners. Both upper limbs were assessed once with the UG and then twice with the EDI, after a 2 hour rest the measurements were repeated again.

The UG was placed at the shoulder in a standard procedure as described by Norkin and White (1995), while the EDI systems use skin markers at anatomical landmarks and placed the device away from the axis of rotation on the dependent part of the limb. Data analysis was performed with PPMCC, means, SD and the standard error of measurement (SEM). Significant differences were investigated between opposite arms using the student t-test.

The intra-rater reliability results for the 2 methods showed that active elevation was generally better for the EDI than for the UG (UG range r=0.45 to 0.94; EDI range r=0.71 to r=0.97). The inter-rater reliability results varied for both devices for different measurements (UG range from r=0.39 to r=0.88 and EDI range from r=0.59 to r=0.87). Inter-device reliability ranged from r=0.21 to r=0.97 (mean r=0.76, SD=0.2). SEM in measurement ranged from 1.5° to 3.2° for UG and 1.4° to 2.7° for the EDI. These findings imply that the EDI performed slightly better than the UG.

The study may be criticised on a number of issues, firstly the methodology did not control patient/subject positions over repeated tests, but allowed them to self-select end of range movements. This implies that the results carry an unknown quantity of error that originates from the test subjects, so called biological error. This is especially important with the patient population who were imminently going for surgery and were probably in some amount of pain. It is fair to suppose that the presence of local shoulder pain in this population would have an influence on their ability to replicate movements. The measurement instruments do not appear to have been blinded during the examinations and finally, both the intra-rater and inter-rater test population were small, as too were the number of examiners involved in the study.

Unlike the EDI or the Orthoranger, the genucom EG system is based on the simple electric potentiometer. Granberry et al (1990) compared measurements made with the genucom (see Figure 1.5) against radiographic measurements for detecting

movement during the anterior draw test at the knee. This test is designed to assess the integrity of the anterior cruciate ligament (ACL) by placing stress on the ligament when the knee is flexed (Petty and Moore, 1998). The study used 8 lower extremities from cadavers on which repeated tests were performed to assess the magnitude of anterior-posterior movement, both before and after sacrifice of the ACL. Measurements of the magnitude of laxity were calculated by both the Genucom and X-ray measurements. The data for the intact ACL for X-ray measurements showed that anterior-posterior translations of the tibia with respect to the femur was 2.7 +/- 0.4 millimetres, while the genucom EG data for the same movement was found to be 3.5 +/- 8.2 millimetres. Comparison of these 2 measurements showed that the genucom EG tended to overestimate the amount of tibial translation. When the ACL was sacrificed and the tests repeated, the extent of overestimation increased slightly further. The results show that the difference between measurements taken with the genucom EG and X-rays varied greatly, with a 95% confidence interval of +/- 5.5 millimetres. The authors conclude that measurements made with the genucom EG, when used to assess translation at the knee joint during the anterior draw test were poor. However, it should be noted that the anterior draw test is not the purpose for which the genucom EG was designed. In a commentary paper Melby and Askew (1991) state that the genucom EG is a reliable tool arguing that it is probably accurate to up to 1°, however they have not published any data to support this statement. They also say that the tool could not be used in a clinical setting because of the length of set-up time that is required.

The flexible/ strain gauge goniometer, as described above, is a completely different type of goniometer. One of the first physiotherapy journal papers that reported using this tool was a study by Rowe et al (1989). This study employed a flexible/ strain gauge EG (designed by the Bioengineering Unit, University of Strathclyde, UK) to gathered normative gait pattern data for a population who had undergone total hip replacement. They did not attempt to investigate the performance of the EG itself.

A later study undertaken by Paquet et al (1991) was designed to determine the validity and reliability of this type of EG for the measurement of sagittal dorsolumbar movements (T8-S1). Validity was tested by measuring lumbar flexion

in 10 normal subjects with the EG and a two-inclinometer method that the authors had previously validated against X-ray, although this data does not appear to be published. The total ROM was divided into 5° intervals, and the angle obtained with both methods was recorded at each of these intervals. The testing procedure was repeated after removal and reattachment of the EG. Analysis of data in relation to corresponding inclinometric angles gave a slope of 1.03 and an R= 0.97. The authors assert that this indicated a high concurrent validity between the two methods. Intra-rater reliability of the measurement procedures were also high for repeated measurements using the EG (ICC=0.98). The authors conclude that under standardised conditions the flexible EG can provide reliable measurements of sagittal dorsolumbar movements that are accurate when compared to those measurements obtained with the two-inclinometer method.

Ojima et al (1991) aimed to investigate the validity and reliability of the bi-axial flexible goniometer [Penny and Giles, Blackwood Ltd., Gwent, UK] for movement at the wrist. This device has the ability to measure 2 directions of joint movement at a time. Validity was assessed by combined movements of a flexible board that had the EG blocks attached to it. The movements were progressed in increasing increments of 10°, as measured by a common protractor. Cineradiography was used to assess the movement of the carpal bones during motion, but not to gain angular data of the movements.

Validity was assessed by comparing the results between the EG and the protractor, which showed an average difference of 2.1° (standard deviations are not reported) and a maximum error of 7°, for bi-axial movement. From this data the authors conclude that the flexible EG is both a valid and reliable way to assess wrist joint ROM. However, the use of the common protractor to assess joint ROM is unorthodox and unsubstantiated, while the published article gives no information regarding the methodology or results for the reliability section of the trial. On the evidence provided in this published article, it is difficult to accept the authors' findings of reliability and validity for the device.

Goodwin et al (1992) compared 3 different types of goniometry; (fluid-filled goniometry (FFG), universal goniometry (UG) and the flexible EG. The aims of this

study were to establish the effect on measurements of using different examiners, different measurement devices and to establish which of the different goniometric methods are most consistent from one occasion to another.

Three physiotherapist examiners were used in the study. These examiners assessed right elbow flexion on 23 healthy female student volunteers. Measurements were made using each goniometer and repeated on two separate occasions. Anatomical reference points were marked by different examiners before full elbow flexion was assessed.

Data analysis consisted of PPMCC, ICC, means and SDs. The results showed high correlations between testers, between repeated measurements made over two separate occasions and between FFG and UG (ICC=0.90). However the correlation between EG and UG was only moderate (ICC=0.51) and was low between FFG and EG (ICC=0.33). Mean correlation coefficients (both PPMCC and ICC) for intra-rater reliability for each of the 3 examiners for each devices were as follows; UG r=0.79 (0.61-0.92), ICC=0.77 (0.56-0.91); FFG r=0.72 (0.53-0.85), ICC=0.71 (0.50-0.84) and EG r=0.36 (0.00-0.43), ICC=0.27 (0.00-0.43). These showed the UG to have the best consistency for repeated measurements by the same examiner.

Inter-rater reliability showed significant differences between the examiners for both FFG and UG (paired t-test). The authors suggest that using the EG is preferable when examinations are conducted between different clinicians, but that in order to achieve reliable measurement results none of the devices should not be used interchangeably.

There were some procedural differences between EG and the other 2 measurement methods, particularly the fact that the EG procedure positioned the subjects forearms midway between supination and pronation, while the other procedures positioned the forearm in full supination. Only one position of full elbow flexion was assessed which may limit inferences to a clinical situation.

A study by Rawes et al (1996) also attempted to investigate the intra-rater reliability and clinical validity of the bi-axial EG for movement at the wrist. The movements assessed were wrist flexion-extension, radioulnar deviation and circumduction. So called clinical validity was said to be investigated by assessing 20 known fixed angles of a wrist in a plaster of Paris slab. There is no indication as to

how the authors calculated the data for these *known* angles. Intra-rater reliability was assessed by twice repeating 5 consecutive measurements on 10 healthy wrists. Data was analysed using means and standard deviations (SD).

Results for measurement validation of the bi-axial EG displayed an average error of 0.2° (±0.6). The intra-rater reliability of the device showed an average error of 3.8°. The authors felt that these errors were not significant although they did not attempt to substantiate this belief with statistical analysis. They conclude that this device has shown high accuracy and reliability. The results of this study are weakened by a lack of information regarding the validation methodology and limited data analysis.

Rome and Cowieson (1996) investigated the intra-rater and inter-rater reliability of three goniometers, the universal (UG), the fluid filled (FFG), and the strain-gauge electrogoniometer (EG) for measurements of ankle dorsiflexion. The strain gauge EG that they used was made by Biometrics Ltd. [Gwent, UK]. The study assessed 10 healthy volunteers using 5 different examiners. A standardised ankle position was used to measure full range of active dorsiflexion (ADF). Intra-rater reliability was assessed using 1 examiner over 2 successive occasions. Inter-rater reliability was assessed among the 5 examiners. A one-factor analysis of variance was used to examine intra-rater reliability, which demonstrated no significant difference between each of the devices on repeated measurements. A multi-factorial analysis of variance demonstrated significant differences among examiners and again among devices ($P < 0.001$). The study concludes that while each device displayed high intra-rater reliability, inter-rater reliability was only moderate. Inter-device reliability demonstrated significant differences among the three devices ($p < 0.1$). The study suggests that each measurement tool cannot be used interchangeably and repeated measurements should not be made by different examiners.

Rowe et al (2001) investigated sources of error and the validity of the Biometrics biaxial strain-gauge EG. Sources of error were established by conducting a series of experiments in a non-clinical environment. These experiments used a large protractor to calibrate a series of large amplitude movements of the EG's

measurement arms. During the trial repeated measurements were made in different environments, for example the trial was repeated using different EG (same model) and with exposure to differing artificial electromagnetic noise. The study also investigated the significance of the introduction of kinematic cross talk, i.e. co-planar movement that naturally occur at a joint. The results show that there is an average error of approximately 1° without any of the above listed stresses. Small errors of an additional 1° or 2° may be introduced when using different EGs or as a result of measurement latency. The effects of co-planar movement may further increase measurement error by approximately 2°, however at the limits of sagittal and coronal movement measurement error may be as high as 10°.

The clinical validity trial used concurrent analysis of 5 healthy subjects assessed using the EG and the Vicon VX gait laboratory system. The data from this investigation was analysed using means, standard deviations, 95% confidence intervals and percentage scores. The results show some variations in detecting the exact timing of joint movement between the 2 systems. The average difference between the methods was 1.5° (± 2.8°).

The authors point out that the validation process was limited by the fact the both the Vicon system and the EG use adhesive skin markers to anchor the points on the limb from where the measurement are taken. This practice may hold significant systematic error and not reflect true bone orientation (Sati et al, 1996). However, the authors recommend the Biometrics strain-gauge EG as a precise and accurate measurement tool for assessing dynamic knee movements. They feel that the errors of approximately 2° or 3° (which is approximately 1 standard deviation) associated with the EG are acceptable for this type of measurements. This study reports on the need for strict operational and set up procedures for the EG, which are necessary to achieve the results that they obtained. The need for these strict procedures further illustrates the challenging nature of attempting to gain accurate measurements of human movement.

A different design of EG is the NK Hand Assessment Laboratory joint motion module system (NK goniometer) [NK Biotechnical Engineering Company, Minneapolis, Minn] (see Figure 1.6). This type of goniometer is similar to the UG in that it has two rigid measurement arms that are manually aligned about limb

segments which enables calculation of a static joint angle. The system works by calculating angles based on electrical resistance between the position of the measurement arms. The main practical difference between the UG and the NK seems to be that the latter device provides a digital read-out of the calculated angles. Armstrong et al (1998) compared the intra-rater, inter-rater and inter-device reliability of both the NK system and a standard UG for measurements of elbow flexion and extension.

The study used a total of 38 one-time patients for its measurement subjects. Five examiners (4 physiotherapists and 1 orthopaedic surgeon) were employed to perform the assessments. The subjects were positioned in a standardised manner, and asked to move their elbow through its full range of flexion and extension. Each subject was measured twice by each examiner within a 45 minutes session. Both end of range flexion and extension was assessed using each of the measurement tools. The NK goniometer was blinded, but the UG was not. The placement of the device used anatomical landmarks as described by Norkin and White (1995). The subjects were brought back for a subsequent test session on another day and the procedures were repeated again.

The ICC was used to illustrate the levels of correlation between the cohorts. In addition means of measurement differences and 95% confidence intervals (CI) were used to quantify the levels of expected error associated with the measurement tools. Scatterplot graphs were also drawn to illustrate the spread of difference between cohorts.

The results of the study showed that intra-rater reliability for the UG had an average 95% CI of 5.9° (max. 11.7°; ICC=0.55 to 0.98) for flexion and 6.6° (max. 14.3°; ICC=0.45 to 0.98) for extension. The NK goniometer displayed an average 95% CI of between repeated tests of 5.2° (max. 8.1°; ICC=0.96 to 0.99) for flexion and 6.3° (max. 11.3°; ICC=0.96 to 0.98) for extension. Inter-rater reliability revealed an upper 95% CI figure for UG of 9.5° for flexion (ICC=0.58 to 0.62) and 10.2° (ICC=0.58 to 0.87) for extension, whereas CI results for the NK goniometer were 4.1° (ICC=0.94 to 0.95) for flexion and 6.1° (ICC=0.89 to 0.90) for extension.

From these results the authors feel that differences of approximately 6° to 7° for UG measurements at the elbow are needed to be confident that actual changes have occurred. This conclusion is based on the average 95% CI seen between repeated measurements performed by the same examiner using the same instrument.

Their results indicate that the NK goniometer may have less expected error associated with repeated measurements than for UG.

There appears to be a methodological problem with this study in that joint positions were not controlled during the measurements. The authors claim to have set out to assess the level of examiner and instrument error associated with the 3 different measurement tools. However, each of the subjects were asked to repeatedly find their own end of range positions for elbow flexion and extension. It is likely that these self selected positions held some variations over repeated testing which would have undoubtedly added an amount of uncontrollable error to the investigations. The authors admit that differences of temperature, fatigue and motivation were not controlled and that this may have been a limiting factor in the investigation.

While this study generally appeared to use appropriate statistical analysis, standard deviations between the data sets were not reported. If this type of analysis were reported it may have made comprehensive comparisons with other studies easier. The authors state that the figures reported for average error between the data sets indicates absolute error, however, it may be more correct to classify this figure as indicating systematic error rather than absolute error. They also suggest that standardised procedures are vital for best results during clinical measurement, but that the research environment may not be observed in routine clinical practice, implying that research findings from other authors may not be truly reflective of expected working error.

1.2.8 Summary of Electrogoniometry.

The EG has seen many developments and changes since its conception in the 1950's. There are now many different designs of electrogoniometry (EG) available to the clinician, so much so that it is no longer appropriate to categorise all of these devices together when considering issues of validity and reliability as each different design performs very different to the next. Designs of EGs based on similar fundamentals to the inclinometer have shown differing levels of performance. The digital inclinometer, EDI has shown improved performances over the UG (Leen T'Jonck et al, 1997), while the Orthoranger (I & II) digital inclinometer has shown

itself to be of limited usefulness (Clapper and Wolf, 1988; Greene and Wolf, 1989; Stude, 1994).

Probably the most accessible and versatile of EGs is the new generation of flexible EGs (see Figure 1.3). This device is capable of assessing dynamic and static joint angles. However, the published research on this tool generally seem to have had limited methodologies. There is also some disagreement between the studies as to the expected performance of this type of EG. Goodwin et al (1992) found that the intra-rater reliability of the flexible EG was poor compared to the UG and the FFG (EG, ICC=0.27; UG, ICC=0.77; FFG, ICC=0.71). Rome and Cowieson (1996) disagree and found no significant differences between results for both intra-rater and inter-rater reliability of the EG. Ojima et al (1991) and Rowe et al (2001) also found promising results from their investigations into the flexible EG. It would seem that, as with other types of goniometry furthur studies are required to assess the performance of this device.

A very positive aspect of EG in general appears to be the fact that these devices provide the examiner with a digital read-out of the calculated angle. This issue may help to minimise error which is attributable to the examiner misreading the device (Mayerson and Milano (1984). Some evidence for improved reliability with digital read-out is provided by Armstrong et al (1998) and Williams and Callaghan (1990).

1.2.9 Photography.

Fish and Wingate (1985) performed a study which aimed to determine sources of error of UG measurement at the elbow, to establish 95% confidence interval (CI) levels for the reliability of elbow measurements and to compare UG with photographic measurements at the elbow. Three sources of error were identified by these authors, which were; identification of bony landmarks, alignment of the UG arms and external forces acting on the limb during passive ROM measurement.

The study aimed to test the potential sources of error by using a total of 46 examiners who conducted measurements on 1 healthy male subject under different procedural stresses. The subject had anatomical landmarks in place over the acromion process, the lateral epicondyle of the humerus and the capitate depression

at the wrist, during these initial assessments. The examiners were divided into 2 groups of 23. The first group assessed the elbow stabilised at approximately 50°, while the second group assessed the subjects elbow when stabilised at approximately 135°. A thermoplastic splint was used to stabilise the elbow on both occasions. Two types of UG were employed a standard large plastic UG and a large steel UG. Each examiner conducted the measurement twice with each UG. This initial assessment procedure was called the ALIGN protocol.

The procedure was then repeated with the anatomical landmarkings removed. Thus the examiners ability to align the UGs without the anatomical reference points were tested. This was called the ASSIGN protocol. The last protocol called the PROM, involved removing the splint from the elbow and the examiners were now asked to work in pairs. One examiner was asked to align each of the UGs while the other examiner passively flexed the elbow to end of range. This was repeated twice with both goniometers.

Photography was also used during the assessments. A single-lens reflex (SLR) camera was mounted on a tripod and aligned to the approximate centre of the joint (lateral epicondyle), positioned 2.5m perpendicular to the elbow. The authors felt that this position minimised visual perspective (parallax) error. Joint angles were calculated from these photographs by projecting the images onto a digitiser which enabled a computer programme to calculate angles to a claimed accuracy of 0.01°.

Data analysis used ANOVA tables, Newman-Keuls post-hoc test and F-ratios. In addition 95% confidence intervals (CI) were calculated as well as means and standard deviations (SD). The results show significant inter-rater differences between the ALIGN and ASSIGN procedures, i.e. when the anatomical landmarks were taken away. There were also significant differences between the plastic goniometer, the steel goniometer and the photographic measurements when the elbow was at 50°. The SD of the differences in goniometric scores at 50° elbow flexion was 4.2° and at 135° flexion was 3.8°. Photography showed significantly less measurement variations than either the steel goniometer or the plastic goniometer.

The authors state that the alignment of the UG can give rise to a significant degree of error leading to variations between examiners of ± 2.1°. When the examiners had to identify their own landmarks measurement error increased by

±0.7° to ±2.8°. They conclude that photography is the preferable measurement option when trying to detect small changes in joint movement. However, the need to replace anatomical labels maybe be a major source of error.

Hayes et al (2001) assessed both intra-rater and inter-rater reliability of 3 methods of joint angle measurements (visual estimation (VE), still photography (photos), universal goniometer (UG)) for 4 shoulder movements (flexion, abduction, external rotation and hand behind back (measured to the vertebral level)). The population for these trials were drawn from patients with diagnosed shoulder dysfunction (intra-rater: n=9, examiner=1; inter-rater: n=8, examiner=4). The intra-rater reliability examiner was an orthopaedic surgeon who assessed each of the 9 patients on 3 different occasions within a 48-hour period. The inter-rater reliability examiners were an orthopaedic surgeon, a sports physician and 2 qualified physiotherapists. These examiners were blinded to each other's measurements but not to their own during their measurements.

Visual estimation assessed passive shoulder movement, while the UG and photos were used to assess active movement. The photographic method used a standard Polaroid camera. Each examiner pre marked 8 reference points on each subject prior to taking the photographs (the T1 spinous process, T7 spinous process, postero-lateral acromion process of the scapula, lateral epicondyle of the humerus, a point 6cm inferior to the postero-lateral aspect of the acromion process of the scapula, the end of 12^{th} rib, the olecranon process of the ulna, and ulna styloid process). The examiner used a small UG placed directly over the photograph to calculate the joint angle.

Data analysis included interclass correlation coefficients (ICC) (2,1) and confidence intervals to evaluate both intra-rater and inter-rater reliability, in addition to calculations of standard error of measurement (SEM) and 95% confidence intervals (±2 SEM) for each range of motion test.

Inter-rater reliability for VE ranged from ICC=0.57 (ext. rotation) to 0.70 (flexion), UG achieved a range of ICC=0.64 (ext. rotation) to 0.69 (flexion & abduction), while the reliability of photography ranged from ICC=0.62 (ext. rotation) to 0.73 (flexion & abduction). Intra-rater reliability of VE ranged from ICC=0.59 (flexion) to 0.67 (ext. rotation), for UG the range lay between ICC=0.53

(flexion) to 0.65 (ext. rotation), while for photography the ICC ranged between 0.56 (flexion) to 0.61 (abduction).

The SEM for the inter-rater reliability trial ranged from 14°-19° for VE, 14°-25° for UG, and 15°-23° for photography. While, the SEM for the intra-rater reliability ranged from 11°-21° for VE, 14°-23° for UG and 13°-22° for photography. Confidence intervals (CI) for the inter-rater trial were VE = ± 28°-38°; UG = ± 28°-50°; photos = ± 30°- 46° and CI for the intra-rater trial were VE = ± 22°-42°; UG = ± 28°-46°; photos = ± 26°-44°.

The authors claim that in general, comparable reliability and measurement error was demonstrated for VE, UG, and photography for 3 measurements of shoulder motion. However, there would seem to be some major flaws in their study design. The patients seem to have been positioned differently between the photographic and other measurements, some methods used active movement (UG, photo), while the other (VE) used passive movements, which may have had an effect on results (Riddle et al, 1987). It would seem that there would have inevitably been some bias for the VE measurements for the intra-rater reliability test especially as repeated measurements were made over such a short period of time. No attempt was made to ensure each subject achieved the same joint position for subsequent measurements. In, fact the researchers state that the joint positions were selected to be 'at the pain limit to range of motion'. This range may have been subject to change and seems to have added uncontrollable error to this reliability trial. Patients may not have been able to achieve equitable ROMs over repeated tests due to the variable nature of pain as influenced by analgesia, temperature, fatigue, motivation or emotional status (Sterigou, 2004). It is possible that the passive ROM test for VE may have been more repeatable than the actively moved joints secondary to the pain levels.

From this trial it is impossible to determine how much error was attributed to actual differences in shoulder position presented by the patient over repeated tests, or how much was attributable to error arising from the measurement tools themselves. Nonetheless, the authors feel there is considerable error associated with these 3 measurement tools.

Braun and Amundson (1989) undertook a study that included human postural assessments in its methodology. The study used angular data gathered from lateral

photographs of its subjects. To overcome the problem of diminished scale and to avoid using a protractor to work out postural angles Postural Analysis Digitising System (PADS) was used. This system uses a slide projector to cast photographs onto a pressure sensitive digitising pad. When pressure is applied to a point on the digitising pad, a computer which is connected to the pad, can interpret the spatial relationship of that point to any other point on the pad. A computer program can calculate the angles between any number of points triggered during this procedure. The authors conducted a small validity trial by projecting a protractor onto the pad and assessing the system at 10° increments of the protractor. The authors state that there was a high level of agreement between the scale on the protractor and the PADS system, although they provided scant information regarding the specific data for this small test. Similar systems were used by Braun (1991), Raine et al (1994, 1997) and Refshauge (1994). However, no further work was done by these authors to investigate the clinical validity or reliability of this measurement system.

To overcome the problem of reference point identification on the photograph, superficial adhesive markers have been placed over relevant anatomical landmarks, as reported by Fish and Windgate (1985). This has been done to assist in identifying the landmarks to be used for measurement reference points. However, Marks and Karkouti (1996) question the efficacy of this practice. In their study they repeatedly photographed 35 healthy volunteers in single leg standing with their legs extended, and at 30° flexion. There was a total of 3 test sessions for each subject. When the photographs were developed, lines were drawn directly onto the photographs through the adhesive markers. The angles of intersection of these lines were then calculated manually using a common protractor. The data was correlated for repeated measurements using the ICC. The results showed that when a measurement was repeated during the same test session the intra-rater reliability was high (extension and 30° flexion; ICC=0.87), however, when the measurement was repeated 4 weeks later the intra-rater reliability dropped substantially (extension ICC=0.53, 30° flexion ICC=0.68). The authors conclude that while reflective markers are beneficial when measurements are repeated within the test session, reliability will dramatically diminish as time progresses, this error may arise from the examiners ability to reselect the same reference point.

A later study by Karkouti and Marks (1997) investigated the reliability of photographic knee and ankle measurements. Photographic assessments were repeated over 3 sessions on 32 healthy volunteers. The examinations were done at an initial session, 7 days later and finally 30 days after the initial session. Two different identifying procedures were used to place adhesive markers at the head of the 5^{th} metatarsal, the lateral malleolus, the lateral epicondyle and the greater trochanter of the femur. Angular calculations and data analysis were similar to Marks and Karkouti (1996). Similarly, the results showed that as time progressed, the intra-rater reliability of photographic measurements using adhesive markers diminished. This was true for knee flexion (Day 1, ICC=0.87; Day 7, ICC=0.70; Day 30, ICC=0.68) and knee extension (Day 1, ICC=0.83; Day 7, ICC=0.75; Day 30, ICC=0.53). The reliability of their method of subject knee placement was not investigated for its repeatability over time, which may have lead to systematic errors and may have been a contributing factor to their poor results. The authors conclude that their results were favourable enough to warrant further investigation for the clinical use of photography and anatomical landmarking, especially if an instantaneous method of photographic analysis was available.

The use of adhesive skin markers was also examined in a study by Sati et al (1996). This study attempted to quantify the magnitude of skin movement in relation to bone, which occurs at the knee. The authors used a video fluoroscopy technique, which is a dynamic radiographic system, to compare the relationship of the 2 tissues. The study assessed 3 healthy males during 3 repeated knee flexion and extension movements (approximately 0° to 65°). Each subject had 5 radio-opaque adhesive markers placed on the lateral aspect of their knees. The direction and magnitude of each marker with regard to the under-lying bone was assessed. The investigation suggested that there may be up to 2 cm of movement between the skin and the bony tissue at the lateral epicondyle during knee flexion. This finding has negative implications for the validity of using adhesive skin markers to assess either static or dynamic human movement. It is not clear from the study whether skin movement is uniform across different areas of the limb, it may be that the lateral epicondyle is not reflective of magnitude of skin movement that occurs over muscle segments or other aspects of the limb. It is also unclear if other factors effect skin movement, i.e.

obesity, gender, age, etc. Nonetheless, the authors feel that the dynamic relationship between skin and bone is by far the greatest obstacle for non-invasive movement analysis. It is clear that further work in this area is required.

An Australian study by Russell et al (2003) set out to investigate the reliability and validity of an internet based measurement method of assessing joint angles. This system termed the internet based goniometer (IBG) uses video conferencing technology to gain real time images, incorporated with audio feed, and was designed to improve contact between physiotherapists and patients who live in remote parts of that continent. The IBG calculates its measurement from still digital photographs captured from the real time visual feed and has similarities to the UM. The physiotherapist instructs the patient to point to the bony point on their legs in order to highlight the measurement reference points outlined by Norkin and White (1995).

The methodology employed to assess reliability and validity used 2 examiners and 1 healthy female subject. The subjects' sagittal knee angles were assessed in 9 different positions (range ~3° to ~97°) with their leg supported on an adjustable wooden frame to stabilise it during the procedures. The first examiner assessed the knee with a standard UG and then twice with the IBG. This procedure was repeated twenty times with the knee in each of the 9 positions, i.e. a total of 180 measurements were made with the UG while there were 360 measurements made with the IBG by the first examiner. Both the IBG and the UG were blinded during the procedures and another individual read and recorded the measurements. The second assessor did not acquire the digital images themselves but used the photographs taken by the first examiner. This second examiner calculated a total of 180 measurements with the IBG only.

Data analysis included ICC (2,1), ANOVA and Bland and Altmans 95% LOA. Validity, as measured by comparisons between the UG and the IBG, was strong with an ICC=0.99 and an average 95% LOA of –1.9° to +1.8°. This 95% LOA data implies there was no systematic bias between the UG and the IBG. Intra-rater reliability for the IBG was measured by dividing the 360 measurements made by the first examiner into two matched groups each of 180. These two groups were then compared and a perfect correlation of ICC=1 was revealed, the 95% LOA

between the paired groups was –1.0° to 1.6°. Inter-rater reliability was assessed by comparing measurements made by the first examiner and the second examiner. The ICC for this question ranged between 1 and 0.96 while the average 95% LOA was –1.0° to 1.4°.

This study portrays an innovative approach to the use of the internet and digital photography as a method of assessing joint angles. Unfortunately its methodology is flawed in several ways which has a considerable impact on the studies claims of validity and reliability for this type of measurement system. Despite multiple measurements being made in an attempt to increase dataset numbers, a very low subject size of 1 implies that it is difficult to predict how the system would cope when measuring limbs of different sizes from different ages and genders that are part of the general population. The first examiner repeated 40 measurements of the knee using the IBG and 20 measurements with the UG in each of the 9 different knee positions. This amount of measurements of the same joint in the same position, made consecutively, must have led to familiarity and recognition of specific features on the limb during the repeated examinations and therefore lead to some degree of examiner bias. Further bias may have been introduced by the design of the wooden frame used to support the subject's lower limb during the trial. In the published article the photographic example of the system clearly illustrates that the frame had the potential to guide the examiners during the placement of the measurement arms in some fashion at least. Additionally the frame appears to impede the natural contours of the posterior thigh segment, a feature that would not be seen in clinical practice.

The authors claimed that the knee was assessed through the full range of motion (ROM), however, measurements were restricted to a range of approximately 3° to 97° flexion, therefore it is not clear how the system would perform with the knee joint in hyperextension or flexion over 100°. The fact that the examiners did not read and record their own measurements also limits inference to clinical conditions. This aspect of measurement has been implicated by Armstrong et al (1998) as having an effect on measurement error and therefore seems to be justified as an inclusion for a strong methodology. Inter-rater reliability was conducted on the software stage only, i.e. the second examiner did not actually take the photographs

themselves or instruct the subject to point to the bony landmarks of their legs. This implies a limited test of inter-rater reliability.

The claim of validity testing by comparisons against the UG has been made by other authors (Rheault, 1988; Petherick et al, 1988; Coxford et al 1998; Dunlevy et al, 2000) however, as previously mentioned and highlighted by this review, the UG does not appear to posses the characteristics of a strong gold standard measurement device. Finally, the subject used during the measurements was a young, healthy female with no history of knee injury, these features may limit claims of validity towards a patient population.

The standard operational procedures used to operate this system may prove to be controversial. In order to identify the measurement reference point the system requires that the patient, in a remote location, must identify their own anatomical landmarks and highlight them to the examiner. This feature may prove to be impractical as different individuals may not posses ability to be accurate recognise these points on their own legs.

1.2.10 Summary of Photographic based Measurement Systems

Photography holds both advantages and disadvantages when compared with other methods of assessing human movement. It is a hands-off method that does not interfere with normal movement, it can also capture both dynamic and static joint motion angles making it quite versatile. Once a photograph is taken angles can be assessed repeatedly and by different examiners away from the subject. The photograph may provide strong hardcopy evidence of the movement ability of a subject at the time of photography, which may be beneficial when performing repeated measurements on a subject over time.

Only a limited amount of work seems to have been done on the validity and reliability of photographic methods of joint angle assessment, which represents a sizeable gap in our professional knowledge base regarding the error associated with this type of measurement. Of the studies that were reviewed here, Fish and Wingate (1985) showed photography holds significantly less variations in measurement than the UG, however, Hayes et al (2001) disagreed and showed similar levels of the standard error of measurement (SEM) for VE, UG and photos. In fact this study suggested that VE performed slightly better over all. This finding is contrary to

Burton et al (1999) and Coxford et al (1998) who found VE to be less reliable than UG.

To date the practice of measuring joint angles with photography has invariably employed adhesive markers or other methods of highlighting the position of specific anatomical landmarks upon which the calculations are based. However, the validity and reliability of placing these landmarks may be limited and attach a significant source of error to the measurements (Marks and Karkouti, 1996; Sati et al, 1996; Karkouti and Marks, 1997). This issue seems to present a need to investigate the link between adhesive landmarking and photographic joint measurements. The development of digital photography may be another reason to consider further research into this area especially as this technology has presented novel innovations to clinical measurement science, such as the Uillinn Method© (UG) and the internet based goniometer (IBG).

In 1997 Karkouti and Marks felt that if photography could be made faster, cheaper and more practical, it may offer a new dimension to clinical and research methods of static measurement. The advent of digital photography appears to offer these features and therefore could hold new clinical possibilities for osteokinematic photogrammetry. The price of digital photography has dropped dramatically since it became publicly available, and will probably drop even further in the future (Balis, 1997). This new and exciting technology has emerged from its infancy and will probably render conventional chemical based photography largely obsolete (Balis, 1997), in fact recently digital camera sales figures have already bettered conventional sales figures (Gleeson, 2003).

1.2.11 Radiography.

Invasive joint ROM assessments, such as X-rays, in theory provide the highest levels of accuracy when measuring a joint angle because they actually map the position of the under-lying bones at the joint. This fact implies that they should provide highly accurate information about the joint structure and the relative positions of the bone shafts.

Radiography has many similarities to photography. The final construct for both techniques is a 2-D image. The emission point of the X-ray machine acts inversely to the lens of a camera, resulting in both systems exposing in one

direction. The resemblance between the 2 methods implies that X-ray osteokinematic assessment should have all the advantages of photographic methods, the only practical difference being the penetrative properties of the X-rays. Nonetheless, this is a crucial factor and the major reason why osteokinematic measurement made with X-ray should provide more trustworthy joint angle calculations. It is for this reason that many trials have used radiological joint measurements as the absolute gold-standard against which to test the validity of other methods of joint measurement (Enwemeka, 1986, Gogia et al, 1987, Granberry et al, 1990, Paquet et al, 1991, Ishii et al, 1995, Brosseau et al, 1997, 2001; Sprigle et al, 2003).

However, as X-ray technology shares many similarities to photography it also shares similar shortcomings, which may primarily be the negative effects of rotation and parallax error.

A study by Gray et al (1991) investigated the validity of radiographic measurements of temperomandibular joint (TMJ) angles. The subject matter used in the study was a resected TMJ taken from a cadaver. The experiment continuously changed the angle of incident of the X-ray beam and the orientation of the mounted TMJ specimen during repeated radiographs. The results showed that statistically significant errors were introduced to the measurement procedure when the specimen and radiological beam were rotated 5° away from each other. They concluded that when evaluating the position of the mandibular condyle in relation to the articular fossa from an X-ray, measurements are invalid if the rotation of the subject with regard to the X-ray beam is unknown.

Lonner et al (1996) conducted a study that investigated the inter-rater reliability of 7 different orthopaedic surgeons for measurement of the tibiofemoral angle from radiographs. A synthetic model of a femur and tibia was used to create a mock-resurfaced total knee arthroplasty. The femoral component was placed in 7° genu valgum while the tibial component was placed in 2° varum with a 5° posterior slope. The overall anatomic alignment was set in 5° valgum. A series of radiographs were taken on 35.5cm x 43cm (14 x 17 inch) X-ray plates, in full extension and 10° flexion, with the limb rotated through 5° increments, from 20° external rotation to 25° internal rotation. The results show that there were no significant differences

between the 7 examiners for most measurements. However, increased limb rotation had a highly statistically significant effect on measured values of the anatomic alignment. The study concludes that when orthopaedic surgeons use radiographs to assess limb alignment in the post total knee arthroplasty population, they must be aware that objective angular assessments may vary somewhat. The authors hold that this finding makes knee alignment indices imperfect.

A study by Ishii et al (1995) compared the tibiofemoral angle of the same knee taken on 2 different size radiographic plates (long 36cm x 130cm and short 36cm x 43 cm). Their population was 100 knees that had previously undergone arthroplasty. When they compared the measurements taken on both radiographs for the same tibiofemoral angle there was a mean difference of 0.5° (± 1.2°). They estimated that there is a significant chance of measurements made by different sized radiographic plates varying by about of 2.5°. However, they conclude that when the tibiofemoral angle is measured from X-ray changes of approximately 2° are not clinically significant, due to random error in the measurements.

Jackson and Waugh (1984) demonstrate that a coronal view radiograph of the knee can show a 4° varum of the tibiofemoral joint when the patient is non-weight bearing. However, genu varum can increase by 14° when the patient is standing. The authors use this as an example to underline the fallibility of X-ray measurements and the need for attention to the exact procedures employed when considering data.

1.2.12 Summary of Radiographic based Measurements

These above findings illustrate the need for caution when considering the role of X-ray as an unquestionable gold-standard for osteokinematic measurement (Jackson and Waugh, 1984, Gray et al, 1991, Ishii et al, 1995; Lonner et al, 1996; Lantz et al, 1999; Greene et al, 2001). Indeed Lantz et al (1999) report that no procedures or technologies, including radiography, have been tested thoroughly enough to be titled an undisputed gold-standard. Therefore, just like any other measurement procedures, radiographic procedures must use consistent and appropriate techniques in order to gain accurate osteokinematic measurement data. When optimal procedures are followed X-rays probably remains the most accessible

approach to validating joint angles and is the best available gold-standard against which researchers can test other joint angle measurements tools.

1.3 Biomechanics Consideration at the Knee and Elbow.

The knee and elbow joints share unique similarities, both having the longest mechanical levers in the body, with movements predominantly restricted to the sagittal plane, additionally both joints have a high incidence of injury (Nordin and Frankel, 1989; Soames, 1995),

There are many issues that must be considered in order to understand and avoid potential sources of error when measuring angles at these joints. Ganjikia et al (2000) caution that the complexity of the knees articulations makes its clinical evaluation extremely difficult. The close proximity of the radio-ulna joint to the elbow may present a significant potential for error due to 'kinematic crosstalk' (Piazza and Cavanagh, 2000). An understanding of the dynamic and structural complexities of both these joints is a prerequisite to understanding the difficulties of measuring these structures.

1.3.1 Anatomy of the Knee

The knee has been identified as the largest and the most complex joint in the body (Nordin and Frankel, 1989; Norkin and Levangie, 1992). It consists of 2 separate articulations involving 3 different bones, the tibia, the femur and the patella. The articulations formed by these bones, the tibiofemoral and patellofemoral joints, make the knee a joint *complex* rather than a *simple* joint (Norkin and Levangie, 1992).

The tibiofemoral joint has been described as a double condyloid joint i.e. it has 2 major articulating surfaces (one medial and one lateral), that has predominantly two degrees of freedom (flexion-extension and transverse rotation) (Norkin and White, 1995).

The superior aspect of the tibia has 2 separate concave facets, a larger medial facet, which is wholly concave and a smaller lateral facet, which is largely concave but becomes slightly convex posteriorly. The femoral surface is convex throughout and is divided into 2 large condyles. These condyles are situated medially and laterally and articulate with the respective tibial concavities.

The convex design of the femoral condyles continues anteriorly and superiorly until they unite to form an inverted "U". The patella articulates with this

surface when the knee is in extension, to form the patellofemoral joint (Soames, 1995).

The patella is the largest sesamoid bone in the body. It is a flat, roughly triangular shaped bone with its apex lying distally. The posterior surface has a smooth vertical ridge that divides it in two, the lateral surface being the lager. Both medial and lateral articulate with the respective femoral condyles. At the most medial aspect of the patella's posterior surface there is a narrow vertical strip, which articulates with the medial femoral condyle when the knee is in extreme flexion (Soames, 1995).

Within the joint capsule between the femoral and tibial articulations there is an accessory joint structure known as the meniscus. The meniscus is divided into medial and lateral segments which are designed to deepen the concavities of the tibial plateau into which the femoral condyles sit. The purpose of the meniscus is to increase articular congruence, to spread weight-bearing forces, reduce friction between surfaces (Nordin and Frankel, 1989), to act as a shock absorber and to allow rotational movement to occur at the knee (McMinn, 1994).

Although the meniscus provides increased surface area for articulation with the femoral condyle, the tibial plateau remains much the smaller of the 2 surfaces. This is an important consideration for knee dynamics. A sliding dynamic between the femoral and tibial surfaces must be present to cope with the small tibial surface. If the sliding component were absent, the femur would roll off the tibial plateau as the knee moves into flexion. In fact the femur must be allowed to slide and roll simultaneously if the normally large magnitude of movement is to occur.

The axis of rotation for flexion and extension passes through the femoral condyles and is oblique to the horizontal. Due to the complex arthrokinematics at the knee, its axis of rotation is not fixed but different for each angular position. This dynamic axis in termed the instant axis of rotation (IAR). The path of the IAR forms a semi-circle which moves posteriorly and superiorly to the femoral condyles with increasing knee flexion (Norkin and Levangie, 1992).

1.3.2 Normal Range of Movement

Flexion and extension are the main movements that occur at the knee. However, there is also a minimum degree of medial and lateral rotation that occurs at the tibiofemoral articulation. The knee is also subject to minimum displacement into anterior and posterior glide, and also into adduction and abduction movements. These movements are secondary to external forces to the knee and cannot be produced by the activity of muscle alone. Norkin and Levangie (1992) state that these accessory movements are due to the tremendous compromise between the mobility and stability that the knee joint provides.

Active transverse plane rotation of the tibia is most discernable when the knee is in 90° flexion. Active lateral rotation occurs when the biceps femoris pulls on the head of the fibula (which is outside the joint complex). Active medial rotation occurs when the semimembranosus and semitendinosus muscles contract and pull on their osseotendonous insertion at the medial aspect of the proximal tibia. Active rotation probably has a role in preparation for correct foot placement (McMinn, 1994).

Passive rotation is said to occur at end of range extension. Here the femur is said to internally rotate by a few degrees due to the shape of anatomical structures of the tibiofemoral joint. The medial femoral condyle is longer and narrower than the lateral femoral condyle. This anatomical difference forces a degree of joint rotation as the knee moves into extension. This action is known as the "screw-home" mechanism (McMinn, 1994). The screw home mechanism is said to "lock" the joint into extension. In order for the joint to move again it must be "unlocked" (Norkin and White, 1995). The unlocking of the knee is attributed to the rotational action of the popliteous muscle (Romanes, 1987). However, the presence or magnitude of the screw-home mechanism is unclear (Piazza and Cavanagh, 2000).

Movement at the knee is by far greatest in the sagittal plane, i.e. flexion and extension. Hyperextension at the knee is usual in young children. As a body ages, knee hyperextension usually disappears (American Academy of Orthopaedic Surgeons, 1994). A study by Boone and Azen (1979) revealed knee extension to be minus 2° ± 3° in a large population of healthy males. The same study showed knee flexion in the same group to average 141° ± 5.3°. This data is backed up by Nordin and Frankel (1989); they state that average knee flexion for the entire population is

about 140°. A large study by Roach and Miles (1991) report slightly differing results, suggesting that active knee flexion is just less at 132° ± 10°.

1.3.3 Anatomy of the Elbow

The elbow joint is described as the intermediate joint of the upper limb (Soames, 2003) connecting the upper arm (humerus) and the forearm (radius and ulna). The joint is formed by the articulations of the convex trochlea of the humerus with the concave trochlear of the ulna; and the convex capitulum of the humerus with the slightly concave, but comparatively flat, disc-like surface (fovea) at the head of the radius (Romanes, 1987). The articulation is held together by the joint capsule, which is thickened medially and laterally and to a lesser extent anteriorly and posteriorly. These thickened portions form the distinct ligaments of the anterior, the posterior, the ulnar collateral and the radial collateral ligaments (Soames et al, 1995).

Soames et al (1995) describe the elbow joint as including two articulations, the humeroulnar and the humeroradial, thus making it a compound synovial joint. They continue that the superior radio-ulnar joint should not be considered as a constituent of the elbow joint. Norkin and Levangie (1992) agree, but state that while both the superior and inferior radio-ulnar joints are separate to the elbow joint, they should both be considered as part of the elbow complex. Moore (1992) disagrees with the previous two publications and state that the proximal radio-ulnar articulation is to be considered as part of the elbow joint, underlining the fact that it is included in the same joint capsule as the other articulations.

These issues are academic, but become important when attempting to classify and describe normal range of movement (ROM) at the elbow. For the purpose of simplicity, this study will agree with Norkin and Levangie (1992) and Soames et al (1995) and consider the proximal radio-ulnar joint as separate to the elbow joint. Therefore the elbow joint can be considered as a uniaxial diarthrodial joint with 1 degree of freedom, that is it acts as a hinge and allows movement only in the one plane (Norkin and Levangie, 1992; Moore, 1992).

The humero-ulnar joint provides the largest articulation at the elbow and along with the capsule, ligamentus and other soft tissue structures are responsible for

the stability of the joint, the structure of the joint dictates the type movement available to it.

The humero-radial joint has a much smaller articulation and in fact when the elbow is full extension there is no contact between the capitulum of the humerus and the radius (Norkin and Levangie, 1992). Movement at the elbow joint occurs primarily by the sliding motion of the ulna about the trochlea, however a rolling motion may occur during the last 5° to 10° of flexion and extension (Norkin and Levangie, 1992). Extension is limited by the contact between the ulna and humerus at the olecranon, while flexion is limited by opposition of the soft tissues at the anterior aspect of the upper limb and the contact between the coronoid process of the ulnar and corresponding fossa of the humerus (Norkin and Levangie, 1992).

The IAR for flexion and extension is proposed as less variable, but more oblique than that of the knee (Norkin and Levangie, 1992). It passes through the trochlea and capitulum of the humerus, which is an individually varying oblique line, this IAR implies that flexion and extension occurs both in the sagittal and coronal planes.

1.3.4 Normal Range of Movement

Normal active flexion and extension at the elbow is approximately 145°, while normal excursion from supination to pronation is approximately 90° (Kapandji, 2001). Hyperextension is often seen at the elbow particularly in children and females. Another interesting angle at the elbow is the carrying angle (Norkin and Levangie, 1992). This is a description of the coronal plane orientation of the humerous and forearm. There is good agreement between the authors with regard to the magnitude of this angle. Stokdijk et al (1999), Moore, (1992 and Romanes (1987) all agree that the angle is of the magnitude of 165°, Soames et al, (1995) believe it to be slightly less at approximately 163°. Ellis and Dusseck (1995) distinguish between the carrying angle of the male and female. These authors suggest that the female has a carrying angle of approximately 165°, while the male has a corresponding angle of about 175°.

Accessory movements at the elbow are predominate in the coronal plane.

For normal function of the upper limb the elbow must have an available range of active movement of between approximately 30° to 130° flexion. If

available movement is less than this most upper limb functions will be limited (Norkin and Levangie, 1992). Soames (2003) reports that the maximum functional range of elbow flexion required is when speaking on the telephone. This task requires approximately 135° of flexion, while opening a door requires as little as 60°.

1.3.5 Considerations for Joint Measurement.

Both the elbow and knee joints certainly hold challenges to obtaining precision angular measurements. Difficulties common to both joints include their complex anatomical and biomechanical design, the presents of co-planar movement, kinematic crosstalk from adjacent joints, a dynamic joint axis, accessory movements caused by external forces, unique joint attributes based on different age and gender.

The normal movement for both joints is similar, the knee has approximately 140° available to it, while the elbow has approximately 135° of movement. Functionally both joints are very important to overall somatic performance and any loss of full ROM would have an effect on activities of daily living.

There is a greater spread of location for the IAR at the knee than at the elbow, which may make sagittal measurement more difficult. This maybe particularly true if the measurement tool used is restricted by a fixed rotational axis, such as that seen with the common universal goniometer (UG). Therefore, a measurement method that does not rely on a fixed rotational point may be more adept to such measurements. Tools that do not rely on a fixed rotational point include, inclinometers, bi-axial electrogoniometers and photographic based systems, such as the UM.

1.4 Discussion.

Most of the studies reviewed dealt with issues of validity and reliability. In this context validity describes the ability of the measurement tool to exactly reflect the angle between the bony structures, i.e. the true value, while reliability refers to the ability to consistently measure the same data when faced with the same angle.

Batterham and George, (2000) state that reliability may be thought of as a fundamental component of validity and is a prerequisite of an accurate measurement system. However, it may be more correct to say that validity and reliability are interdependent and should be considered together.

Most studies were able to clearly define the issues of validity and reliability and most undertook to examine one or both of them. Measurement error from different sources unquestionably effects both issues and it is data regarding these errors that should be thought of as highly valuable for measurement science and should be the primary focus of worthwhile research.

1.4.1 Measurement Error

Armstrongs (1998) believes that research into osteokinematic measurement tools should seek to answer three different questions, firstly how well do the examiners (clinicians) carry out the measurements, secondly how suitable are the measurement instruments to the applied tasks, and thirdly how well do the subjects (patients) lend themselves to being measured by the examiners and equipment. These relate to the sources of error outlined by Hopkins (2000a), i.e. examiner, instrument and biological error.

These 3 error sources effects the data in one of two ways, either predictably or unpredictably. Predictable error is classified as non-random error or systematic bias (SB), while unpredictable error is classified as random error or measurement noise.

In practice the procedures followed during the measurement process is an important consideration for error. It would seem logical that both random and non-random error may be minimised when examiners follow standard procedures. Consistent patient position has been shown to be an important factor during joint

angle measurements (Rothstein et al, 1983). Gogia et al (1987) had what appeared to be the most controlled position for their subjects'. When the extremely high reliability results demonstrated in that study are compared to results from similar studies that did not appear to control patient as stringently during their knee assessments (Boone and Azen, 1979; Rheault et al, 1988; Watkins et al, 1991; Brosseau et al, 1997; Armstrong et al, 1998) it leaves little doubt that controlling patient position during measurement is vital.

Maintaining correct alignment of the UG over the limb	Grohmann, 1983; Fish and Wingate, 1985; Petherick et al, 1988; Nicol, 1989
Identifying appropriate bony landmarks	Mayerson and Milano, 1984; Fish and Windgate,1985; Nicol, 1989
The inability of the tool to adequately cope with movement couples	Nicol, 1989; Norkin and Levangie; 1992, Ganjikia et al, 2000
Naturally occuring skin movement during changes in joint position	Nicol, 1989; Sati et al, 1996
The mismatch between the dynamic joint centre and fixed UG axis	Enwemeka, 1986; Nicol, 1989; Szulc et al, 2000
Variations in the joint size and individual (subject) joint variations	Fish and Windgate,1985; Williams and Callaghan, 1990; Bovens et al, 1990
The position in which the joint is assessed	Fish and Windgate,1985; Petherick et al, 1988; Williams and Callaghan, 1990
The lack of standardisation in nomenclature	Petherick et al, 1988
Errors arising from the different skill levels of the examiners	Fish and Windgate, 1985; Petherick et al, 1988; Williams and Callaghan, 1990; Goodwin et al, 1992; Marks and Karkouti, 1996; Karkouti and Marks, 1997
Errors arising from the characteristics of the instrumentation	Petherick et al, 1988; Williams and Callaghan, 1990
Errors arising from the characteristics of the patient	Mayerson and Milano, 1984; Williams and Callaghan, 1990; Hageman and Sorensen, 1991
Errors arising from the measurement procedures	Ekstrand et al, 1982; Stratford et al, 1984; Grohmann, 1983; Rothstein et al, 1983; Williams and Callaghan, 1990; Goodwin et al, 1992

Table 1.7 Challenges to Measurement Specified in the Literature.

Table 1.7. lists some of the difficulties mentioned in the literature as causative factors for measurement error. It should also be noted that some of the errors listed here have not been unanimously proposed. For example, the skill level of the examiner has been stated as a source of error by some authors (Fish and Windgate, 1985; Petherick et al, 1988; Williams and Callaghan, 1990; Goodwin et al, 1992; Marks and Karkouti, 1996; Karkouti and Marks, 1997). However, a number of other investigators disagree with this proposal, finding that profession, number of years experience or clinical speciality did not have an influence on reliability scores (Armstrong et al, 1998; Burton et al, 1999, Holm et al, 2000)

One specific aspect of examiner skill that was investigated by Williams and Callaghan (1990) was the ability to read the numerics from UG. This study showed that using a digital read-out UG improved reliability, implying mis-reading the analogue scale on the standard UG is an important consideration. This finding was backed up by Armstrong et al (1998), who showed much improved results for the NK goniometer (see Figure 1.6) over the UG. These devices are mechanically similar but the former method employs a digital read-out of the angle, which seems to make reading the device easier.

The fact that reading the UG may be a considerable source of measurement error may have negative implications for the methodologies of some of the other trials reviewed here. Blinding the UG to prevent bias is obviously important, but some studies went further and asked other independent examiners to read the instrument (Fish and Wingate, 1985; Petherick et al, 1988; Watkins et al, 1991; Burton et al, 1999). As mis-reading the instrument is a source of error, having the examiners negate this responsibility may have falsely improved their results and therefore not portrayed an accurate account of the actual expected reliability error. If the aim of a study is to produce clinically relevant reliability data for a measurement instrument, the methodology should ensure that the same examiner conducts all aspects of the procedures, this should be done in order to reflect what happens in an operational setting.

Gross osetochondral or articular irregularities may be common in some patient populations. This individuality presents further difficulties when attempting to design one tool to accomplish measurements for the entire population. Different joints inevitably have different characteristics, due to their size, predominant

movement, surrounding musculature and mechanical joint type. With this in mind, it would seem reasonable that different measurement tools may lend themselves differently to different joints (Fish and Windgate, 1985; Bovens et al, 1990). In an early study Boone and Azen, (1979) state that the UG has better reliability for joints in the upper limb measurement than for joints in the lower limb. Even different movements at the same joint will show different results, Greene and Wolf (1989) found best reliability for elbow and shoulder flexion and extension (95% confidence intervals (CI)= 2° to 4°), while the worst results were seen for transverse shoulder rotations (95% CI =15° to 17.2°). In short, measurement reliability may vary from one patient population to another (Leen T'Jonck et al, 1997; Hayes et al, 2001), from one joint to the next (Boone and Azen, 1979; Fish and Windgate, 1985; Clapper and Wolf 1988) and from one joint movement to the next (Fish and Windgate, 1985; Riddle et al, 1987; Hayes et al, 2001). Fish and Windgate, (1985) believe that research is needed at every joint if confidence is to be gained for a joint measurement method.

1.4.2 Scientific Research

Scientific research is the most appropriate way to assess the suitability of a measurement tool for a specific purpose. Such investigations must adhere to certain methodological fundamentals that will ensure the results are without bias and are able to provide reflective data regarding the research question.

The methodology should initially aim to take adequate sampling and provide a large degrees of freedom (df) for the data analysis. As with all quantitative research, a larger df implies stronger inferences for the study. Conversely, a limited amount of data implies a limited statistical power and limited implications from the study's findings (Hicks, 2004). Sampling should include the use of as many examiners and subjects as possible.

1.4.3 Study Populations

The subject population upon which the measurements are made should be also considered as a vital question for the methodology, whether the examiners are skilled or unskilled, or whether the subject groups are drawn from a healthy or patient population. These factors will be a major consideration for the conclusions

drawn by the study. A literature review by Karkouti and Marks (1997) underline the importance of adequate sampling if a study is to be considered worthwhile and able to stand up to peer review.

The majority of the studies reviewed here did not consider all of these issues. Some studies only used one healthy subject upon which to conduct repeated measurements (Mayerson & Milano, 1984; Fish & Wingate, 1985; Croxford et al, 1989; Williams & Callaghan, 1990; Rawes et al, 1996; Burton et al, 1999), while others only used one examiner (Clapper and Wolf, 1988; Greene and Wolf, 1989).

A lot of studies used a healthy population rather than a patient group as subjects for their research. It can be argued that this population will not introduce the comprehensive challenges to the measurement process that is needed to infer results to a patient population. Karkouti and Marks (1997) criticised a number of studies already discussed in this paper for homogenous sampling (Mayerson and Milano, 1984; Riddle et al, 1987; Youdas et al, 1993; Watkins et al, 1991).

Grohmann (1983) and Nicol (1989) proposes that UG measurement is more difficult when there is pain, immature scars, oedema or pathological structural deviation. All of which are not associated with a healthy population. This implies the need to test measurement equipment on a patient population in order to adequately challenge the system.

However, Burton et al, (1999) states that the nature of the subject has minimal effect on the *reliability* of the examiners techniques and that using a patient population that may be difficult to control leads to the introduction of more variability in experiments, i.e. biological error. If this type of random error is allowed to go unchecked, it may prevent a true assessment of the reliability of the examiner or instrument. Therefore the results may be subject to a type II error, i.e. the actual reliability will be better than the results indicate.

If the purpose of an investigation is to assess the reliability of the examiners technique, the measurement apparatus itself or more likely a combination of both of these criteria, then the best methodology must ensure that repeated measurements are made with biological error tightly controlled. This may be better achieved in a healthy population.

While there is controversy about which population is best suited to reliability research, there does not appear to be any justification for using a healthy population in a validity trial. Methodologies used for validity research typically only require a series of once off measurements made using different tools at approximately the same time. This implies variability in joint position during the one off exam may not be an issue, once the joint is stable during the examinations. Additionally, a validity study must show that a measurement tool can cope with the biological variations described earlier, otherwise the conclusions of a validity investigation will not be clinically relevant. For these reasons it is proposed that validity studies should exclusively use patient populations in order that the results of the study are to be truly reflective of the expected clinical performance.

1.4.4 Statistical Analysis

When incorrect statistical tools are used any conclusions drawn from the study will be flawed (Hicks, 2004). A limited understanding of data analysis tools and over use of correlation coefficients were probably the greatest fault with the reviewed data.

George et al (2000) state that the Pearsons product moment correlation coefficient (PPMCC) cannot account for systematic measurement bias and may consequently yield misleading results. Significance testing is also of limited benefit because validity and reliability are not all or non-phenomena (Altman, 1999; Rothstein 2001). This insight is exemplified by the confusing results seen in studies by Petherick et al (1988) and Rheault et al (1988) which claimed that reliability was good when the data was examined using the PPMCC, yet significant differences were present when the same datasets were subjected to t-tests. The value and interpretation of this non-significant correlation is not clear.

George et al (2000) also assert that while the Intraclass Correlation Coefficient (ICC) is a much better tool than the PPMCC, it too has its flaws for this type of analysis. A major drawback of limiting the data analysis to correlation coefficients is the fact that these tools do not display the magnitude of difference between the data sets or the type of error between the observations. In addition, they do not deliver the data in a relevant measurement unit. Therefore, results data may be changed from an original ratio set, to a less meaningful ordinal scale.

To illustrate these issues; if 2 different measurement tools had a constant disagreement of 10°, a PPMCC would still illustrate a perfect agreement between the methods, while an ICC (3,1) would display a moderate value (see Table 1.8). On the basis of the PPMCC the reader would falsely assume perfect agreement. The ICC value would show a limited agreement but the reader would not be able to determine the size or nature of the error.

Measurement	Degree	Degree
no. 1	100	10
no. 2	90	9
no. 3	80	8
no. 4	70	7
no. 5	60	6
no. 6	50	5
no. 7	40	4
no. 8	30	3
no. 9	20	2
no. 10	10	1
PPMCC	1	
ICC (3,1)	0.57	

Table 1.8. Correlation Table Between Two Cohorts with Large Bias

Walmsley and Tyler (1996) state that there are at least 6 different formulae to calculate the ICC and each formula will produce a different result on the same group of data. They recommend that when the ICC is used, the researcher should include the formula used and the rationale behind its use. Unfortunately, not all of the researchers reviewed in this paper reported this information which renders their results difficult to compare.

Another variable associated with the ICC is the way the formula is applied. It seems that the same formula may be used on the same data sets, but a different value may be gained. The confusion arises due to the fact that it is valid to make calculations from the means of the data sets or directly from the raw data itself (Eliasziw et al, 1994). These different calculation methods will result in a different coefficient despite the same ICC formula being used on the same data.

Hopkins (2000b) justifies the ICC (3,1) formula, 'ICC= (F - 1)/(F + k - 1)', (where k = (number of observations - number of tests)/(number of subjects - 1) (the F value should be based on a two-way ANOVA test (without replication)). However, contrary to Hopkins (2000b), Shrout, and Fleiss (1979) provide the formula for ICC (3,1) as 'BMS-EMS/BMS+(k-1)EMS'. However, these different expressions of the formula appear to be academic, both were tested on the same dataset and were shown to provide exactly the same figure to at least 7 decimal points.

Yet another issue relevant to the use of correlation coefficients is trying to gauge what their coefficient actually means. Theoretically, the closer to '1' the coefficient is, the better the agreement between measurements. When measuring humans some variables are expected over repeated tests, which implies that a perfect correlation coefficient is not likely. The logical next question is what coefficient value represents an acceptable performance. This figure is not clearly defined. Domholt (1993) has set out guidelines for interpretation of ICC results. These authors propose that an ICC score of between 0 to 0.19 = no correlation, 0.20 to 0.39 = trace correlation, 0.40 to 0.59 = fair correlation, 060 to 0.79 = good correlation and 0.80 to1 = high correlation. Fleiss (1986) and Armstrong et al (1998) proposed that a value of < 0.40 was deemed as representing poor reliability; 0.40-0.75 as fair to good reliability; and > 0.75 as excellent reliability. While Manly and Hopkins (1989) set a pass / fail limit for measurement apparatus at r=0.90. Atkinson and Nevill (2000) disagree and warn that even an ICC=0.90 does not ensure strong reliability. The discord between authors suggest that some of these guidelines may not be based on scientific fact, but more on opinion, resulting in confusion for the reader.

Relatively recent developments in statistical analysis have recommended agreement testing as the optimal means of gauging the accord between datasets. This type of analysis divides error as either random or non-random (Bland and Altman, 1986; Atkinson and Nevill, 1998; Rankin and Stokes, 1998; Altman, 1999; Hopkins 2000a; George et al, 2000; Batterham and George, 2000; Daly and Bourke, 2000; Bruton et al, 2000; Sprigle et al, 2003).

In the case of detecting consistent, non-random error or systematic bias (SB) between datasets, the most appropriate method suggested from other researchers is the 95% LOA test (see Appendix Q). This is achieved by comparing the averages difference between the cohorts (Bland and Altman, 1986; Hopkins, 2000; Batterham and George, 2000; Daly and Bourke, 2000). This average difference figure is unlikely to be exactly zero, but may have some posiive or negative value, a absence of the perfect zero is expected due to be some amount of random error in the measurements (Hopkins, 2000). Therefore, when attempting to detect a true SB a watershed limit must be set. Batterham and George (2000) propose that if the average difference between the datasets is less than 2.5% of the mean measured value, then this can be taken to represent no evidence of SB.

A number of tools that have been proposed as being suited to establishing levels of random error or measurment noise in the data. These include the 95% LOA, the Repeatability Coefficient (RC), Standard Error of Measurement (SEM), Standard Error (SE) and Typical Error (TE).

The 95% LOA test establishes the upper and lower limits of error over repeated tests, this analysis is informative as it portrays SB and maximum random error (for 95% of cases). The RC also portrays maximum error but assumes no SB.

It is more likely that SB will be present when measurements are made by different examiners or different techniques than between repeated assessments are made by the same examiner (Altman, 1999). This is why the RC is normally only suggested to be used for assessing random error for single examiners (Sprigle et al, 2003) and the LOA is recommended for inter-rater tests.

The RC allows the results to be reported as a single coefficient rather than a two figure result of an upper and lower limit as is the case with the LOA and is therefore easier to understand. If SB is not present in the data it seems reasonable to display the RC data rather than the LOA result.

The 95% LOA agreement test has been criticised by Hopkins (2000a) for a number of reasons which include the assertion that a 95% limit may be too stringent for analysis of some questions that may require the examiner to set their own specific detection limits. Another criticism is that the researcher is unable to detect a measure of expected error using only one set of data, i.e. the 95% LOA test is best suited to the assessment of 2 datasets.

In agreement with Stoke and Rankin (1998), Hopkins (2000a) recommends that a relatively large degree of freedom is required to accurately make inference from the LOA data, although they do not agree on the required size of the population. Stoke and Rankin (1998) propose that 50 participants are necessary to have optimal results although they do not justify this figure. Hopkins (2000a) shows mathematically that there is a bias of < 5% when there is more than 25 degrees of freedom (e.g. > 25 participants and 2 trials, or > 13 participants and 3 trials), however, this bias rises to 21% with 7 degrees of freedom (8 participants and 2 trials). This current trial would endeavor to employ greater than 25 degrees of freedom in the methodology.

Measurement	**A (Degree)**	**B (Degree)**	**A-B**
no. 1	200	10	190
no. 2	180	150	30
no. 3	160	70	90
no. 4	140	90	50
no. 5	120	100	20
no. 6	100	100	0
no. 7	80	150	-70
no. 8	60	50	10
no. 9	40	150	-110
no. 10	20	230	-210
Average	**110**	**110**	***0***
Std Dev.	**60.5**	**62.36**	***110.15***
95% Limits of Agreement (LOA): *-215.9, +215.9*			
Repeatability Coefficient (RC): *215.9*			
Typical Error (TE): 77.9			

Table 1.9. Example of the Calculations Made with Means and Standard Deviations.

Finally, the manner of displaying the 95% LOA data has also caused minor controversy, Atkinson and Nevill (1998) and Hopkins (2000a) recommend that the format should be 'SB +/- RC', where SB is the systematic bias and RC is the repeatability coefficient. However, it may be more appropriate to report the non-random error (SB) and the random error separately (Batterham and George, 2000).

Other authors (Russell et al, 2003; Sprigle et al, 2003; Nigel-Keenan, 2004) simply express this calculation as Lower -X, Upper X, where 'Lower –X' is the figure calculated for the lower LOA and is often a negative value, and 'Upper X' represents the figure for the upper LOA (see Figure 1.1). Expression of the data in this way should point the reader to the conclusion that during a subsequent measurement the random error will fall somewhere in that range. It does not imply that the subsequent measurements will differ by that entire range.

In addition to the maximum amount of random error detected between the datasets (for 95% of cases) which is calculated by the LOA and RC, it has also been recommended that the minimum random error also be analysed (Hopkins, 2000a).

The SEM, SE and TE have all been recommended as a means to predict the characteristic error of a measure (Hopkins 2000a; George et al, 2000; Batterham and George, 2000). However, there is controversy about which analsysis tool is best suited to this task.

In the literature there appears to be a plethora of different formulae that have been proposed for the calculation of the SE, SEM or TE data (Bland and Altman, 1986; Eliasziw et al, 1994; Lythgo, 1998; Hopkins, 2000; Batterham and George, 2000). One form of the SEM calculation is put forth as; $SEM=SD\sqrt{n}$, where SD is the standard deviation of the two cohorts and n is the sample size (Batterham and George, 2000).

Another proposed method of calculating the SEM is by using the formula $SD\sqrt{r-1}$, where SD is the standard deviation and r is the reliability coefficient for the entire test (Lythgo, 1998; Hopkins, 2000), this method appears to be the most widely proposed in the literature. However, Hopkins (2000b) states that this formula has limited use in statistical analysis and should be avoided as it can yield misleading results and does not provide accurate inferences.

Bland and Altman (1986) propose that the 'standard error' should be calculated on the data in conjunction with the LOA tests using the formula; $SE=\sqrt{SD^2/n}$, where SD is the standard deviation of the differences and n is the sample size. However, they propose this test as a means to gauge how precise the estimates provided by the LOA test are, i.e. it provides similar information to that of a confidence interval test, rather than detecting minimum typical error.

Hopkins (2000a, 2000b) proposes the use of 'Typical Error' (TE), which is calculated by the formula TE=SD/ $\sqrt{2}$, where SD is the standard deviation of the differences and 2 represents the number of cohorts in the estimate. That author feels that all of the above formulae seek to detect the within-subject error and correspond to the variability between a single individual's measurement scores, which are useful clinically because they allow the clinician to determine when a real change has occurred as opposed to a change attributable to measurement error. Therefore the SE, SEM and TE are all terms that should be thought of as synonymous.

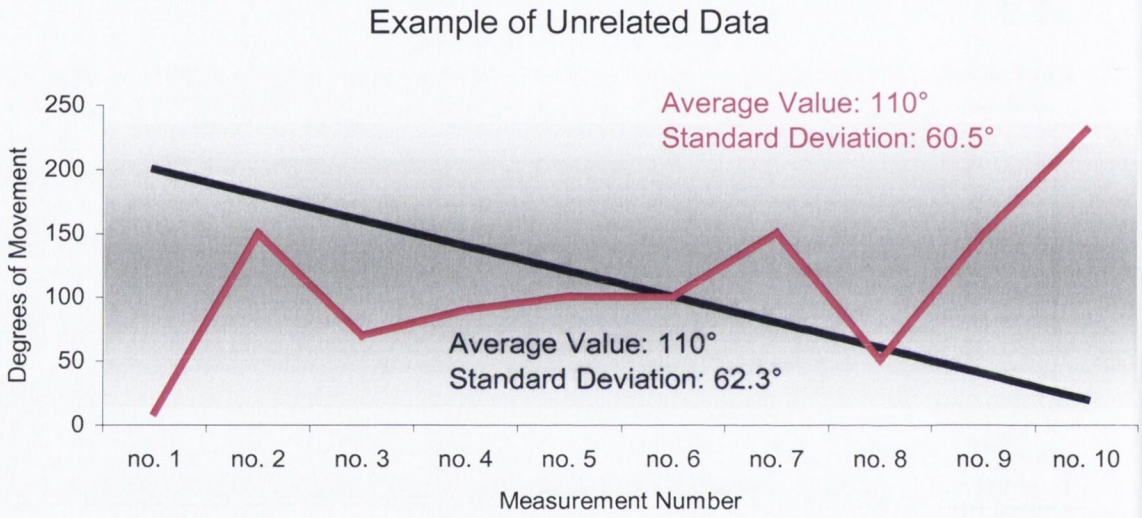
Methods of calculating TE for both bi-variant and multiple-variant datasets are displayed in Appendix Q, while the calculations of confidence limits for TE data have been published (Hopkins, 2000b).

Descriptive analysis or analysis of central tendency (Coxford et al, 1998) in addition to correlation coefficients have been employed in the majority of studies reviewed here to help explain results. This type of analysis is necessary to help illustrate the extent of error that was experienced by the researchers. This information will be very beneficial to the clinician. However, the descriptive analysis in some of the literature has not always been reported exhaustively and can sometimes be misleading.

Some studies used only mean differences to reveal magnitudes of error (see Table 1.4; 1.5) or may have only reported the standard deviations (SD) of the individual datasets. It should be noted that without publication of SD of the *differences* between datasets, the spread of the data will remain unknown and as a result knowledge is severely limited. Conclusions based on mean data only must be considered inadequate and ambiguous. This issue can be illustrated by the data in Tables 1.8. In this example the data are completely unrelated. Dataset A is homogenous and dataset B is without form (heterogeneous), yet mean differences and the SD of the datasets are comparable. The reader only gets an impression of the poor relation between the cohorts when the SD of the differences between the datasets is examined (Table 1.9.; Graph 1.1).

It may be said that the majority of studies discussed in this review had limited data analysis methodology, which is said not to be uncommon in the medical

literature (Daly and Bourke, 2000). Some of these studies are therefore seriously limited, leaving interpretation of their results, including comparisons between similar studies difficult. Due to the presence of vastly different methodologies and data analysis systems that have been used in the literature, final conclusive information regarding the osteokinematic measurement tools seen daily in clinical practice can not be drawn with full confidence. This is despite much time and resources invested over the years. Encouragingly, the trend of heavy reliance on correlation coefficients seen in older studies now seems to be changing.



Graph 1.1 An Illustration of Unrelated Data Maintaining Means and Standard Deviations (from Table 1.9).

To prevent wasted resources on poor research designs, it may be that there is a need for professional bodies to recommend and make readily available standardised methodologies and data analysis approaches to the clinicians who are embarking on research projects. This may help minimise the simple pitfalls that were discussed here and thereby improve the quality of methodologies and thereby strengthen the conclusions that are drawn from research. It may even make research more accessible to clinicians who are inexperienced in this field.

1.4.5 Osteokinematic Measurement Tools

It is not surprising that the most commonly investigated tool in the literature is the UG, as it is probably the most common measurement device used in clinical practice (Coxford et al, 1998; Burton et al, 1999).

The electrogoniometers (EG) approach to joint measurement has seen enormous development since initial conception. Probably the most common type of EG used in research or clinical practice today is the flexible/ strain gauge EG. The more recent evolution of this device has the advantage of being tri-axial, i.e. it has the ability to detect joint movement in 3 planes (Biometrics, Gwent, UK). Nicol (1989) holds that the EGs should be primarily used to assess dynamic ROM. Norkin and White (1995) agree, they also report that EGs are expensive, take time to properly calibrate, and that they are best suited to the research process.

Inclinometers use gravity as a constant reference against which to assess joint movement. They are usually attached directly to the limb and are commonly used to assess the total excursion of a joint rather than static angular measurement between adjacent body segments. It seems to be vital that the rest of the body is properly stabilised when using this device in order to properly isolate the joint movement being measured.

Measurements taken from still or moving images derived from photographic or cinemagraphic techniques allow the examiner to assess a hard copy 2-D image away from the patient, at any time and allow assessment of 1 or more joint angles simultaneously. Calculations may be done using a common mathematical protractor directly over a photograph or by using one of a number a computer interfaces.

X-ray has advantages over photography as it can map the exact position of the under-lying bones and does not rely on surface anatomy. However, it has been shown that all of these imaging methods are subject to significant error if strict photographic/ radiographic procedures are not carefully controlled (Jackson and Waugh, 1984, Gray et al, 1991, Dunlevy et al, 2000). Fish and Wingate (1985) agree and state that obvious sources of error associated with photography include parallax error, optical error and digitisation errors. These authors advise that photography is the preferable choice when trying to detect small changes in joint ROM, although this assertion seems based on opinion only.

Fish and Wingate (1985) also found a consistent error between UG and photography and propose that this was due to different viewing perspectives of the

examiner when placing the UG. Karkouti and Marks (1997) held that if photography could be made more time efficient and practical it should be recommended for clinical practice. Norkin and White (1995) state that photographic measurements are most suited to research and not clinical practice presumably taking this view because of the traditional expenses of both time and money associated with photography. However, new developments in the area of digital photography may offer a change to these previously experienced limitations. Even in 1997 Balis foresaw that new and exciting developments in the fields of communication and science would be made possible by the new digital camera. It is now possible that photographic based measurement techniques may offer a viable addition to the science of joint angle assessment, such as Russell et al's (2003) IBG.

Visual estimation (VE) of a joint angle is commonly used by clinicians as it is very quickly done (Watkins et al, 1991; Youdas et al, 1993; Norkin and White, 1995). In fact its speed of application is probably why it is used by about 50% of occupational and physiotherapists (Burton et al, 1999). There does not appear to be any recommended approach or technique towards conducting VEs, rather it is left to the inherent ability of the examiner to make the estimations.

Vision is very important when conducting a general physical assessment, however a lot of the information gained from this sensory mechanism may be defined as subjective information. Some authors found VE of joint angles was relatively reliable compared to other techniques, while others refute this claim. There appears to be stronger evidence to support the use of an incremented scientific measurement device over VE when attempting to gain osteokinematic data. VE of a joint angle can be thought of as a subjective impression of objective data.

1.4.6 Challenges to Static Osteokinematic Measurement.

The standard UG placement outlined in the literature usually direct the examiner to employ specific anatomical landmarks as reference points from which to base a measurement. However, as previously stated, these static landmarks do not acknowledge the presence of a dynamic joint centre, which seems a particularly important point when assessing the knee (Norkin and Levangie, 1992).

Watkins et al (1991) found that the reliability of UG improved from ICC=0.85 to ICC=0.92 for knee flexion and from ICC=0.84 to ICC=0.87 for knee

extension when they used a standard placement for their UG. Rheault et al (1988) asked their examiners to place the UG at the approximate joint centre rather than an anatomical landmark. Their reliability results were slightly lower (r=0.87) than similar studies that used the lateral epicondyle as a placement point for the UG (Rothstein et al (1983), r=0.97; Mayerson and Milano (1984), r=0.97; Gogia et al (1987), r=0.99). From this data it could be suggested that using fixed reference points may be a more reliable method for UG measurement, even thought it may theoretically be a less valid approach (Szulc et al, 2000).

This point highlights a dilemma facing the science of human measurement, it is a question of which is fundamentally more important validity or reliability. Is it better to have a more valid model, i.e. try to allow for a dynamic joint centre at the expense of reliability, or try to have a more reliable model, which uses anatomical landmarks and will ignore the true nature of the joint. The latter model will undoubtedly hold some amount of SB in the measurement results. In fact it would be unwise to ignore either one of the validity or reliability issues for a measurement tool. The best practice approach must consider both issues together.

Despite the popularity of the UG it has been shown that this device has various inherent limitations (Croxford et al, 1998; Szulc et al, 2000). One limitation may be the inability of the UG to compensate for the complex nature of knee joint arthrokinematics (Enwemeka, 1986; Norkin and White, 1995; Brosseau, 1997, 2001). Szulc et al (2000) substantiated this theory finding that there were significant differences between the fixed axis of movement of the UG and the constantly changing axis of the knee joint. Indeed, those authors found that the lateral epicondyle of the femur, which is commonly used to indicate the joint axis can be over 48 mm away from the true centre of joint rotation.

However, Brosseau et al (1997, 2001) investigating the parallelogram goniometer (PG), which is a similar device to the UG but it can accommodate a dynamic joint axis by its design (see Figure 1.2). In neither study did this device show improvement in either validity or reliability over the UG for measurements at the knee. This finding may suggest that either the design of the PG does not adequately cope with the dynamic joint centre of the knee or that this issue is not the only problem facing joint measurement.

In agreement, Grohmann (1983) compared two different UG placements, one standard lateral placement and the other along the extensor surface of the flexed elbow, thus leaving the UG fulcrum completely remote from the joint movement axis and negating the influence of a dynamic joint centre on the UG placement. This study found no differences between the reliability of placement methods, again suggesting that dynamic joint axis alone is not the main source of measurement error. Indeed, some joint measurement methods do not rely on the position of the joint axis (imaging techniques, VE and EG) but these are also subject to validity and reliability measurement errors.

Static joint angles are by their nature difficult to predict and pose yet another challenge to the measurement process. The end-of-range position of the two bone shafts is subject to its surrounding soft tissue. This soft tissue is in turn subject to the unpredictable nature of pain, spasticity, nutritional status, fatigue and the viscoelastic properties of the tissue (Sterigou, 2004). Therefore, as these properties change due to environmental or physiological factors so too may the available movement.

Passive ROM assessments are useful when differentiating between causative factors of joint restriction. These passive movements require an external force to find the end point of a joints movement. However, a common difficulty with this practice is how much force is required. It has been proposed the a gentle force that will place minimum stress on the opposing soft tissue should be classified as 'R1', while a large force, or over-pressure, on a joint that will stretch the opposing tissue should be termed 'R2' (Norkin and White, 2003). Consistency in the technique for this passive measurement may be a further challenge to minimising error.

1.4.7 Specific Challenges to Photographic Measurement Systems.

In a study by Szulc et al (2000) the term *photometry* was used to describe joint angle measurements taken from photographs. However, in a paper on international symbols, units and nomenclature (SUN), Palmer (1999) states that the discipline of photometry is the science of measurement of light which is defined as "electromagnetic radiation which is detectable by the human eye". Defining angular

measurements taken from photographs as a photometry a term already defined by the International Standards Organization (Palmer, 1999), seems inappropriate and would certainly lead to confusion.

There are numerous advantages to photogrammetric measurement methods in osteokinematic science. A photograph can capture multiple joint movement simultaneously at any instant during dynamic movement (Nicol, 1989). The photograph freezes the image allowing multiple or single joint position analysis. Most clinical osteokinematic assessment tools are used to make a once off only measurement that cannot be verified at a later stage by a different examiner, whereas the photographic image can be held indefinitely and analysed repeatedly by one or more examiners. Other obvious benefits include the fact that it is a non-contact technique and there is no need to touch or interfere with the subject being assessed. This may help to lessen bias to a measurement introduced by the examiner and be more reflective of the actual function at the joint. The technique is non-invasive and therefore does not carry any health risk for the patients that are associated with repeated invasive imaging procedures, such as X-ray.

However, it should be recognised that there are some shortfalls to photogrammetry measurement procedures. Conventional photography is costly in terms of both time and money. There can be a lengthy delay from the time of photography until the images are returned from chemical processing for analysis of the hard copy photographs. The cost of chemical processing of the film and purchase of replacement film is also a considerable ongoing cost (Nicol, 1989). These practical issues are a major obstacle toward the use of conventional photography in a clinical setting.

There are also considerations intrinsic to photography that may introduce errors to the measurement process. Primarily, errors may be introduced by a misrepresentation of the true image on the photograph, which may originate from a number of sources. One such consideration is the fact that photographs capture a 3 dimensional (3-D) structure and record it as a 2 dimensional (2-D) image. Three-D vision allows a perception of depth, which allows the viewer to be aware of the orientation (rotation) of an object with respect of the view-point. A 2-D image, such as a photograph or a radiograph does not provide this type of information. An illusion of an angle becoming more acute or obtuse is created when a 3-D angle is rotated with respect to a fixed view-point. Due to the loss of depth perception, this

illusion cannot be detected on the 2-D picture. Therefore, the amount of rotation between the object and the photographic lens can contribute to increased error during angular calculations (Gray et al, 1991).

Secondly, a standard camera lens will carry slight flaws due to imperfections inherent from the manufacturing process. It is possible that lens imperfections may cause photographic distortion and therefore result in measurement error (Scholz, 1989). The design of the cameras and lens may also introduce flaws to the recorded image. The fact that the cameras' photographic plate is flat while its lens is convex leads to a fundamental mismatch in the design of the camera. A convex lens refracts light, i.e. it bends light rays as they travel through it allowing light emitted or reflected from an object to be focused on a photographic plate. This distortion of the natural straight path of the light rays may alter the exact nature of the true image. The resulting distortion of the photograph is known as the 'pin cushion' or 'pillow' effect (see Figure 1.7). This term illustrates the presence of some amount of image distortion, which occurs at the 4 sides of a photograph. This phenomenon may imply that if a photograph is to be used for angular measurement, the body segment in question must be in the centre of the lens view-field (Dunlevy et al, 2000).

Another important reason for the image to be in the centre of the lenses view-field is due to the danger of introducing parallax error. Dunlevy et al (2000) investigated parallax error by moving a digital camera to different positions, so that an angle on a target object was situated in different positions of the lens view-field. The camera was positioned 2 meters away from the target at all times. They discovered that when the camera was moved more than 0.9 meters in any direction from the centre (at 2m perpendicular distance), average measurement error increased from less than 1° to over 3°. They conclude that further research is needed into camera - subject position and the photographic procedure in order to completely evaluate sources of error associated with osteokinematic photogrammetry.

A further potential source of error may be introduced during the photographic procedure from inappropriate movement. The photograph may be distorted by movement of the camera when the lens shutter is open and the photosensitive plate is exposed. Movement of the camera at this critical point will blur the image making it difficult to discern finite points, such as anatomical landmarks. Movement of the camera when the lens shutter is open may occur when

the photographer presses the trigger button to take the picture and accidentally moves the whole camera. To overcome this problem a photographer may use a tripod to steady the camera. There does not appear to be any published research that has investigated the error involved with free hand versus tripod mounted photogrammetry.

The calculation procedure may also be a source of error. The traditional procedure used to assess angles from the photographs is done by placing a protractor over the photograph to calculate the joint angles. This may limit the accuracy of the procedure and introduce error when reading and recording the measurements. The examiner may discern between increments of 1°, however, these may be difficult to read as most protractors are only numbered in increases of 10°. Exact placement of the scale on the protractor over the relevant body segment can also be difficult as neither the protractors scale nor the photograph are adaptable. These factors may introduce errors to the measurement procedure by simple misreading of the scale on the protractor (Mayerson and Milano, 1984).

An illustration of the type of distortion that may occur at the edges of a photograph. Note the outer lines do not appear as straight. This may be due to the effect of the convex lens and the flat photographic plate.

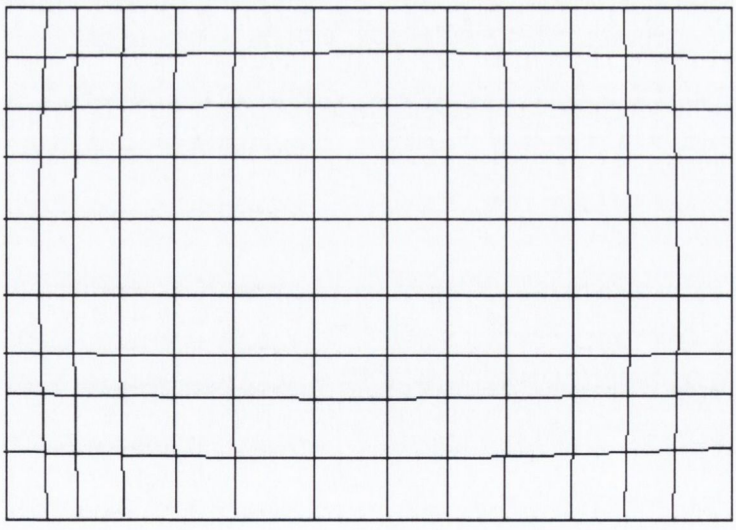


Figure 1.7. The Pillow Effect.

Another disadvantage of photogrammetry is the fact that the photograph of the relevant body segment is generally a fraction of the size of the actual limb,

which may make it difficult to identify reference points, such as superficial bony landmarks. The size of the average developed photograph is usually in the order of 15cm x 10cm. This results in an enormous reduction in the apparent size of the limb, which may lead to a decrease in measurement sensitivity. The decreased sensitivity during the measurement procedure may lead to the presence of an increased measurement error when calculating the joint angle.

The labelling of anatomical landmarks with adhesive skin markers has been used to try to improve the reliability of measurement techniques, such as photography (Fish and Wingate, 1985). However, it clearly does not eliminate reliability error. The main limitation of this technique appears to originate from the inability of examiners to accurately replace markers over subsequent assessments. Selfe (1998) found that intra-rater reliability of placing adhesive markers on the exact area of skin showed a 95% confidence interval of within 1.5 cm. Fish and Wingate (1985) and Armstrong et al (1998) suggest that reliance on anatomical labelling presents a major additional source of reliability error. Additionally, the accuracy of the replacement of these markers also seems to be related to time. Marks and Karkouti (1996) and Karkouti and Marks (1997) found that as time between replacing the labels increased so too did the error associated with the practice.

A further validity issue also associated with adhesive landmarking involves the physiological movement of skin over the bony landmark during joint movement. Sati et al (1996) found that there may be over 2 cms of movement between the skin and underlying bone during knee movement from 0° to 65°, which alone could lead to miscalculation of true angular measurements by about 2.3° (see Appendix A), and may lead to larger errors when combined with other error sources such as the misidentification of the true joint axis (Szulc et al, 2000) and the changes in repeated marker placement errors already discussed (Selfe, 1998).

1.4.8 Current Knowledge of Measurement Error

From some of the studies that had reasonable methodologies it is possible to quantify some of the attributes of these measurement tools. Beginning with measurements of the lower limb, Coxford et al (1998) considered the inter-rater reliability of the UG and visual estimation (VE) from measurements of ankle

dorsiflexion and plantar flexion. This methodology was reasonable, but may be criticised for a small population size (n=1) that the repeated measurements were made on. These authors suggest the 95% CI for VE between examiners may be 11.1° while the UG scored a 95% CI of 5.5°. Holm et al (2000) may be censured for apparently failing to isolate their data groups, otherwise there methodology seems to have been sound. These authors found that the intra-rater reliability for the UG for 6 different measurements at the hip ranged from a 95% CI of 5.5° to a 95% CI of 26.1°.

Sprigle et al (2003) had an excellent methodology, a minor shortfall was a different analysis methods used between the intra-rater and the inter-rater trials and a moderate sample size (n=20). The measurements made with the UG were of the hip flexion angle. The repeatability coefficient (RC) reported in this study imply that there is an expected disagreement of 6.32° when the same examiner repeats a measurement. When different examiners make the measurements their 95% Limits of Agreement (LOA) was -10°, +14.6°. These authors also examined the pelvic goniometer for the same measurement, this device scored slightly better with an intra-rater error (RC) of 5.5°. The inter-rater error for this device was between -4° to +8.4° (95% LOA). In addition the study also investigated the validity of the pelvic goniometer and found that the measurements differed between data taken from X-ray by between -18° to +8° (95% LOA). These findings seem to provide sound evidence for the expected validity and reliability of these tools at this joint.

Analysing Enwemekas validity data may suggest that discord between X-ray measurements and UG at the knee would be in the region between 9.5° and 14.4°.

Leen T'Jonck et al (1997) investigated measurements of shoulder movements, they used a UG and an electronic inclinometer (EDI). They report a SEM between measurements by the same examiner for shoulder abduction of 2.6° for the UG and an SEM 2.8° for the EDI. The standard deviations for these devices were 10.1° and 11° respectively.

Another study that used the SEM to assess shoulder abduction was investigated by Hayes et al (2001). These authors compared UG, VE and measurements made using photography. Intra-rater SEMs were much different than those published by the previous authors. According to Hayes et al (2001) the UG has an SEM of 23° for shoulder abduction, while VE has a SEM of 21° and photography

has an SEM of 22°. The inter-rater SEMs were 21°, 19° and 23° for UG, VE, and photography respectively. The population size for this study was small with only 9 patient participants. The large disagreement for the SEM of shoulder abduction between these 2 studies underlines the need for further research into this area using these and other analytical tools.

Burton et al (1999) investigate the inter-rater reliability of assessing the metacarpophalangeal (MCP) joint using both the UG and VE. They establish that the UG has an RC of 4.4° for flexion and 5.9° for extension, while VE had a much larger RC of 44.5° for flexion at the MCP and 51° for extension.

Elbow movement was investigated by a number of authors. Fish and Windgate (1985) found that expected inter-rater reliability error for 95% of measurements (2 x SD) would be less than 8.5°. The results published by Armstrong et al (1998), were not dissimilar despite different data analysis methods. These latter authors found that the 95% CI for elbow flexion when measured with the UG was 9.5° while elbow extension was 10.2°. They also investigated a digital goniometer, which improved inter-rater reliability results to a 95% CI of less than 6.1° for both measurements. The intra-rater reliability for both these devices were comparable. The UG had a 95% CI of between 5.9° to 6.6°, while the digital goniometer had a 95% CI of between 5.2° and 6.3°.

Agreement between Fish and Windgate (1985) and Armstrong et al (1998) regarding an expected error of less than 10°, for 95% of measurements, between examiners for UG measurements at the elbow may imply some legitimacy to this figure. Especially, as both studies had fair methodologies, with the exception of Fish and Windgates' (1985) small population number (n=1).

Further evidence for error of this magnitude may be provided by two different studies. Both Bovens et al (1990) and Walker et al (1984) examined multiple joint measurements of both upper and lower limbs. Bovens et al (1990) stated that they expect error between examiners to be less than 10° for same joint measurements, while Walker et al (1984) displayed a SD of the differences between examiners of 5°, which may indicate a 95% limit of agreement of approximately 10° (2*SD) (see Table 1.5)

One issue that all of the literature reviewed agreed on was that intra-rater reliability is always higher than inter-rater reliability. Clapper and Wolf (1988) and Greene and Wolf (1989) agree that for best results the same measurement device should be used by the same examiner for any particular joint measurement. The fact that the accuracy of measurements improves when there is an absence of different assessors and devices implies that all types of joint measurement have an amount of systematic error associated with them. When different examiners repeat assessments their individual techniques will add to measurement noise. With this in mind it may be recommend that measurement protocols must recommend that repeated measurements should be made with the same device and by the same examiner when possible. When measurements are made by different examiners strict protocols should aim to promote homogenous techniques.

The interchanging of different devices for the same measurement has been investigated by some authors (Rothstien, 1983; Riddle et al, 1987; Rheault et al, 1988; Bovens et al, 1990; Rome and Cowieson, 1996; Brosseau et al, 1997; Armstrong et al, 1998; Brosseau et al, 2001). Rothstein et al (1983) and Coxford et al (1998) recommended that different types of UG may be used interchangeably. Rheault et al (1988) disagree and conclude that both UG and FFG are both valid and reliable, but warn that these measurements should not be used interchangeably. Norkin and White (1995) agree with Rheault et al (1988) and state that the FFG may be a useful tool, but it should not be used interchangeably with a UG or any other type of goniometer. Rome and Cowieson (1996) investigated three devices, the UG, FFG and EG. These authors also agree and state that these measurement tools cannot be used interchangeably and furthermore repeated measurements should only be made by the same examiner.

It may also be considered that acceptance of one particular method as an absolute gold-standard may be flawed. Lantz et al (1999) report that no procedures or technologies, including radiography, have been tested thoroughly enough to be titled an undisputed gold-standard, and that all measurement methods have some degree of fallibility associated with them. This implies that all concurrent validity research of osteokinematic measurement tools are open to some degree of

interpretation. Burton et al, (1999) agree and feel that there is no absolute gold-standard for joint angle measurement.

1.4.9 Implications of Measurement Error

While it appears that every method of osteokinematic measurement may have some amount of error associated with it (Lantz et al, 1999; Burton et al, 1999), consideration must be given to the magnitude of this error. It must be contemplated as to what amount of this inevitable error is acceptable with specific reference to research and clinical practice. Ishii et al (1995) suggest that when measuring joint angles using X-ray anything less than a 2° error should not be considered clinically relevant. Mayerson and Milano (1984) report that in 1949 Hellebrant et al suggested that a 3° variation in repeated measurement is unimportant. Rowe et al (2001) agree with this figure and purpose that when dealing with musculoskeletal problems errors of the order of 2° or 3° are clinically acceptable. Boone and Azen (1979) feel that only a change of between 3° or 4° should be considered as a figure to indicate a true change in joint movement, implying error less than this is unimportant. Enwemeka (1986) proposes that an average validity error of less than 4° is acceptable. Mayerson and Milano, (1984) consider that an inter-rater variability of 5° is acceptable, while Nigel-Keenan et al (2004) suggest that when reliability is examined using the Bland and Altman 95% LOA, an agreement of less than 10° is satisfactory.

Nicol (1989) states that in terms of overall motion, 1° error is clinically insignificant. The author illustrates this by stating that for an average male, 1° of motion at the elbow results in only 0.8 cm of fingertip displacement, i.e. the furthest aspect of the limb segment from the joint is subject to a very small positional change as a consequence of 1° of joint movement. Even with the relatively large lever of the forearm the impact of 1° of movement at the elbow is small.

This current author has derived a similar figure of 0.87 cm for foot displacement when 1° of change is seen at the knee. This figure increases to 1.74 cm of foot displacement for 2° at the knee, and 4.36 cm for 5° (see Appendix A). These figures are a clear illustration of the impact of error propagation, while 1° error has an equivocal effect on overall movement, an accumulative amount of error from

different sources can quickly provide misleading information. This supports the need for tight constraints of measurement error for all sources.

1.4.10 Areas Requiring Further Research

Despite being the most generally used tool in physiotherapy practice (Coxford et al, 1998; Burton et al, 1999) validity and reliability for the UG remains poorly defined. Part of the reason for this is that the majority of research carried out on this device to-date has failed to demonstrate through meaningful data analysis the expected error associated with this type of measurement.

There is agreement that the UG will perform differently at different joint but the exact behaviour of this error is not well understood. As the device is used so frequently, definitive research that includes appropriate data analysis must be seen as urgent.

The literature concerned with validity and reliability issues for a photographic based approach to measurement is sparse. This presents a clear lack of knowledge with regard to the performance of this approach to osteokinematic measurements (Fish & Wingate, 1985).

When using photographic methods of joint angle assessment, the literature suggests that adhesive skin markers are necessary. However, there is controversy regarding the validity and reliability of using these markers (Marks and Karkouti, 1996; Sati et al, 1996; Karkouti and Marks 1997). It remains unclear how the error associated with marker placement influences the over all error associated with the photographic measurements or whether the placement of these skin markers is necessary at all.

When the studies that dealt with the upper limb are inspected it becomes apparent that no investigation has attempted to validate a measurement tools using a gold-standard. This presents a further gap in the knowledge of osteokinematic measurement science, specifically for the upper limb.

1.5 Conclusion.

The use of efficient outcome measures are pivotal to the delivery of a physiotherapy service that aspires to a best practice model. There are many outcome measures available to the clinician which attempt to measure issues such as pain, function, strength, balance, joint range of motion (ROM), etc. The measurement of joint ROM allows inferences toward the performance of a specific joint and general somatic function (Greene and Heckman, 1994). Objective joint ROM measurement is one of the most important measures to the physiotherapist (Petherick et al, 1988).

It is vital that health professionals who assess joint angles do so with both valid and reliable measurement tools. There are numerous methods currently used by clinicians and researchers to assess joint angles, examples of some of these methods include; universal goniometry (UG), visual estimation (VE), electrogoniometers (EG) and photography. Invasive imaging measurement procedures, such as X-ray, can also yield information regarding joint angles. No joint angle measurement method has shown perfect accuracy, all seem subject to some level of error (Lantz et al, 1999; Burton et al, 1999). It is the responsibility of the clinician to be aware of the limitations of the measurement tools they use (Rothstein, 2001). These limitations can only be investigated by standard scientific research. It is vital that any researcher who undertakes to assess the performance of a clinical measurement tool conducts their research using standard scientific principles, which will include adequate sampling, appropriate exhaustive data analysis and sufficient control of sources of error.

There seems to have been a poor appreciation of the importance of sampling and informative analytical tools. However, this trend does appear to be changing in more recent studies.

Identified problems that are associated with the science of osteokinematic measurement are difficulties allowing for movement couples when try to isolate movement in only one plane. The dynamic joint centre, especially of larger joints may also be difficult to contend with. The different sizes and pathologies seen between individual patient joints also make it problematical to design one tool to

deal with all of these variables. Tools that obtain measurements based on the external surface anatomy also face the difficulties of correct placement of the tool and may be subject to errors from the independent mobility of the superficial soft tissues.

It seems inevitable that some amount of error will be present in joint angle measurements, however, it must be asked what amount of measurement error is considered clinically significant. It has been proposed the errors of 5° (Mayerson and Milano, 1984) to errors of 2° (Ishii et al, 1995) are clinically insignificant. While, there appears to be a some consensus regarding how much error is tolerable, the research has been unable to give a clear account of the actual error associated with our measurement tools. In fact, due to the widely different methodologies and data analysis used in the literature, simple comparisons between the studies are difficult (Burton et al, 1999). There is some evidence to suggest that errors of about 10° should be expected over repeated examinations by the UG, although the exact amount if likely to change from joint to joint.

The only major consistent finding between the studies reviewed here is the point that intra-rater reliability is constantly better than inter-rater reliability. This finding may suggest that generally there are substantial variations between examiners approaches to all of the different measurement methods. Additionally, most studies agree that for optimal results different assessment tools should not be used interchangeably and repeated clinical measurements should only be made by the same examiner.

There is general concensus that all osteokinematic measurement error can be classified as originating from either of 3 different sources. These are errors that arise from the patient (biological), the examiner or the measurement tool itself (instrument). Identification and controlling these errors would seem to be the core of good measurement practice.

Chapter 2

METHODS: THE DEVELOPMENT AND APPLICATION OF THE UILLINN METHOD©

2.1 Introduction

In 2001, the Irish Department of Health and Children (DOH, 2001) proposed that information technology (IT) 'has the power to streamline and improve the healthcare system'. Worldwide, the health care industry has already benefited enormously from the information revolution with the development of many diagnostic devices (Thorp, 1998). No one would argue that the technological innovations of the last century have been critical to advancement of health care (Sandelowski, 2000). This relationship seems set to continue at pace into the future (Borchers & Kee, 1999).

Physiotherapy practice has certainly seen benefits realated to IT. Personal Computers (PC) are now commonly used for such activities as word processoring, data store, home exercise prescription via applications such as 'Physio Tools' [Microsoft, Redmond, Washington, USA] and they may be used to manage research data.

Diagnostic, rehabilitation and treatment tools have also been developed using IT, examples are isokenetic machines, which provide highly specialised data regarding muscle function. Diagnostic ultrasound machines can demonstrate activity in the deep musculature and provide biofeedback to the patient, foot pressure plates linked to PCs provide information about the forces and timing during the stance phase of gait and may provide information regarding the need for orthotics. There are also a number of electrotherapy devices, such as therapeutic ultrasound machines that now rely on IT to improve their performance by providing a clear display of treatment dosage.

Measurement science is of central importance to the development of an evidence base for physiotherapy. The practice of joint angle measurement has taken advantage of developments in IT in an attempt to improve accuracy and influence. The electrogoniometer (EG) employs IT to record, calculate and present its joint dynamics data. Sophisticated gait laboratories employ a great deal of computer hardware and software to collect objective 3- dimensional human motion data. The detailed assessment that such systems provide are highly beneficial when attempting to perform precision movement analysis (Nigel-Keenan et al, 2004).

The literature review (1.2) clearly showed that our ability to accurately measure static joint angles with traditional tools was limited. Recently IT has

presented new ways to measures static joint angles with such tools as the IBG (Russell et a, 2003) and the NK goniometer (Armstrong et al, 1998). Such new approaches to this practice were extremely welcome. This chapter was written to describe the development process of a new type of computer driven static joint angle measurement tool, the Uillinn Method© (UM, version 2.x).

The development and experimental testing of this new system of static joint angle measurement using virtual measurement arms (VMAs) to form a virtual goniometer was undertaken as a PhD thesis.

2.1.1. Background to Developing the Uillinn Method©.

Postural analysis systems based on conventional photography and IT have been used by several authors in the past (Braun and Amundson, 1989; Braun, 1991; Raine and Twomey, 1994,1997). The UM version 1.x, was developed by the author to improve upon this type of analysis by using digital photography and a custom built software application.

Dunlevy et al (2000) described this system and its reliability for analysis of the craniovertebral angle (CV) of the head and neck (see Figure 2.1, 2.2). This type of measurement required the examiner to be able to assess the angle between the true horizontal, as represented in the photograph, and a line between the tragus of the ear and cervical vertebra 7 (C7) (see Figure 2.1; 2.2), this task was successfully achieved using the computer software. Testing this technology was submitted as an under-graduate research project.

The Craniovertebral angle is represented by 'x'.

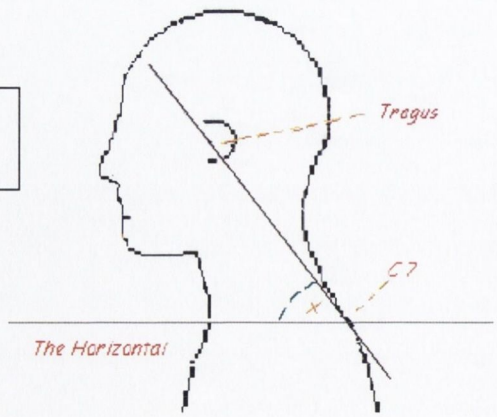


Figure 2.1. The Craniovertebral angle.

At the conclusion of the undergraduate study it was felt that the system maintained only a minor amount of examiner and instrument error, and hence this technology was judged to be reliable. However, the CV angle itself was subject to considerable biological error. It was recommended that a system based on the UM (version 1.x) could be developed as a virtual goniometer for the purpose of static joint angle measurements.

2.2 Software Development

The digital camera uses a microprocessor and a photosensitive electronic plate to capture and store an image digitally, which is in turn designed to be accessible to a standard personal computer (PC) as a 'Joint Photographic Experts Group' (JPEG, pronounced "Jay Peg") file. Computer files of this format end with the file suffix of ".jpg" and are referred to as JPEG files (www.jpeg.com, 2004). The formal standard for the JPEG file format is ISO/IEC IS 10918-1 | ITU-T Recommendation T.81 (www.iso.ch, 1994). It is relatively easy for a software program to access and manipulated computer files such as these.

The novel approach to angular measurements proposed by the UM required new computer software that could render a JPEG computer file, i.e. display a digital photograph, and allow a user to identify coordinates from which an angle could be measured. It was felt that the most optimum computer language that could complete such a task was the Visual Basic (VB, version 6.0) language.

VB is an 'object oriented' computer language was developed by Microsoft [Redmond, Washington, USA]. It was released in order to allow users to create software applications for the Microsoft Windows Operating Systems (OS). As this language was designed specifically for the Microsoft Windows environment, it has an extensive library of pre-defined "objects" which represent many of the standard features of a Windows based application. Examples of these standard 'objects' (also called 'controls') include, 'Drop Down Boxes', 'Menus' and 'Radio Buttons'.

The key advantages of using VB for the development of the UM project were twofold. Firstly, the construction of the software is greatly simplified when it is developed to run specifically on the Windows environment. The language provides predefined libraries of functions for most of the complicated data 'event handling' or 'image processing' tasks. If this feature were not available, the Windows OS would need to be provided with detailed instruction on every intricacy of the data processing tasks. This requirement would have greatly exaggerated the development time and may have been outside the technical expertise of the developers. Therefore the use of VB for the development of the UM project allowed attention to be focused on the functional aspects of the software rather than the need to write

enormous amounts of detailed computer code associated with controlling the OS subsystems.

Secondly, as the Windows controls supplied in the VB's libraries and have been extensively tested by Microsoft. This meant that it was unlikely that any program errors in the system would be related to these controls. Any errors experienced during the software programming stage, would be relatively easy to locate in the customised functional subsystems that were written during the developed process. These advantages would not be present if the software had been developed in another computer language, such as the prevalent alternative language of C++ [AT&T Labs - Research, Middletown, NJ].

The disadvantage of using VB was that this language would be unable to cope with complex image handling and data processing functions that would be essential if the project were to evolve towards an elaborate or multi-level functioning system. Essentially, if the system were to be written in VB it would need to be limited in terms of the complexity of the functions it could perform. If, during the development stage of the UM, it were to be decided that the project needed to evolve into a more complex system, a more powerful programming language would need to be considered.

However, it was felt that clinical physiotherapists did not currently require a more complex operational design. In fact, such a system may intimidate some clinical users and therefore limit its general appeal. The intentions behind the development of the UM project was to provide a simple, user-friendly graphical interface that would allow the majority of users to quickly become familiar with its operation and therefore allow examiners to easily obtain clinically relevant joint angle measurements.

2.2.1. The Uillinn Method© Version 1.x

The construction of the software initially dealt with the design of the graphic user interface (GUI) (see Figure 2.2). It was identified that this area had to provide a simple and obvious platform for interaction with the user. The design would incorporate a map of the file system that would allow the user to easily identify and select their computer files and folders.

Two manoeuvrable measurement lines were added to the software which allowed the examiner to select coordinates on the graphical image. These were termed the 'Reference line', which was to be placed against the true horizontal and the 'User line', which was to be placed on the subject's tragus and C7. These lines could be manipulated by clicking on different 'radio buttons' in the software's control area (see Figure 2.2).

In short, the key elements that were identified as being the basic required components of the system were;

- A graphical, mouse driven user interface.
- An obvious and easily navigable file system based on a standard Windows model.
- A large image (photographic) display area.
- Automatic calculations of the relative angle between selected coordinate points.
- An on-screen numeric representation of the measured angle, in degrees.

The Microsoft Windows Operating System (OS) calculates graphically based angles in terms of radians, with a pre-designated definition of the zero vector. All angular estimations made by the UM had to be defined relative to the OS's impression of zero. In addition the software had to resolve this data from radians into the more recognisable angular degrees. This data was continuously computed and delivered, in real time, to a display area within the software's window. A version tracking system was used to monitor, trouble shoot and modify the different development stages of the UM's software.

2.2.2. Operational Procedures for the Uillinn Method©

The measurement procedure for the UM was divided into 3 sections, firstly the *photographic component*, which involves positioning the subject and camera, and then taking the actual photograph.

The second part of the procedure is termed the *file transfer component.* This entails transferring the digital image from the camera to the PC via either direct cable linkage or a Smart Media drive (see Figure 4.3, page 139). This process also involves the files being correctly named and stored.

The final stage of the process is *the software component* which involves resolving the image with the UM software and manually using the software to calculate the joint angle.

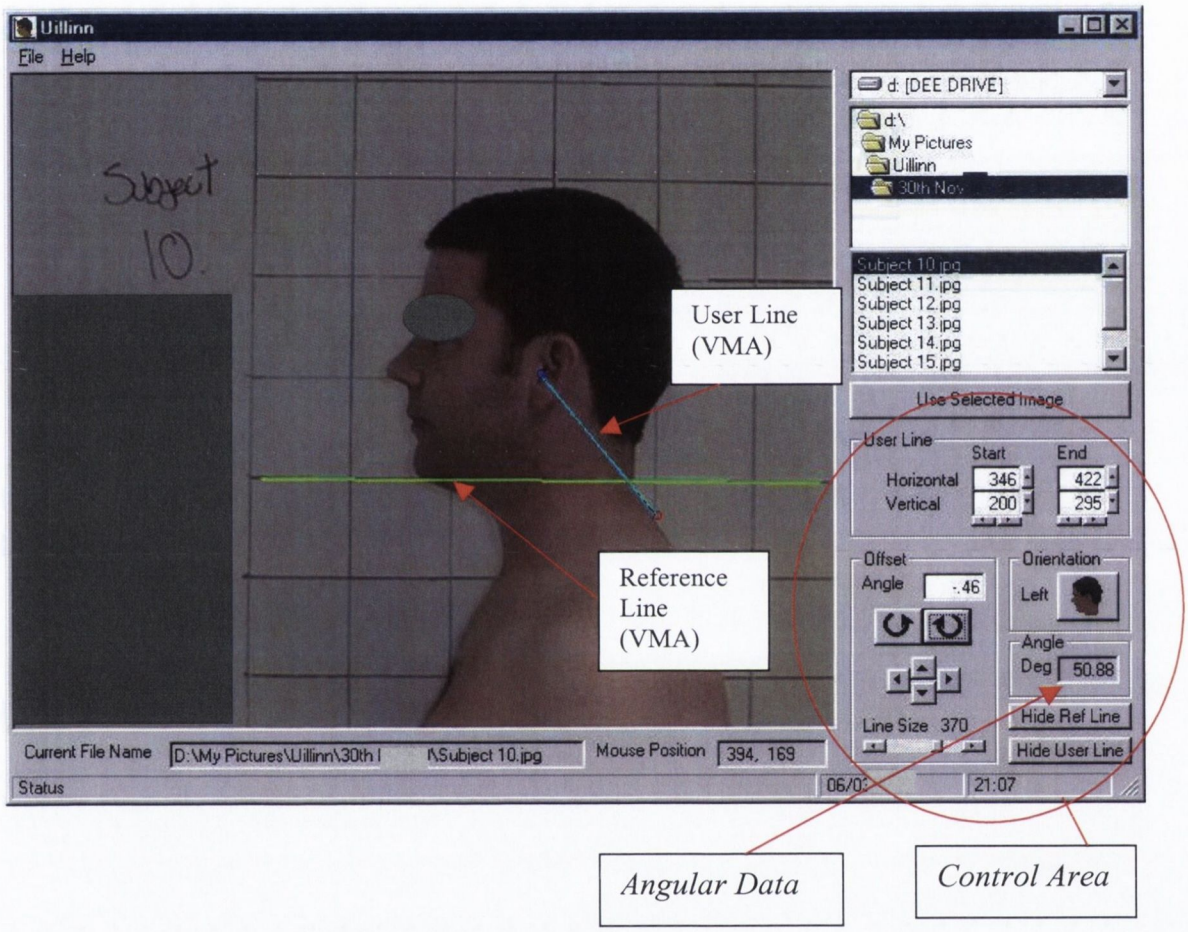


Figure 2.2. Version 1.2 of the Uillinn Method©.

2.2.3. The Rationale for Further Development

On completion and submission of the undergraduate project that used UM version 1.1.9 (March, 2000). A number of problems were highlighted with the system. Some of these problems were subsequently addressed and re-complied as version 1.2 (see Figure 2.2). A summary of the identified problems included system crashes upon opening a second JPEG file, difficulty identifying the correct orientation of the User line with respect to the Reference line, system crashes when attempting to change the orientation of the image within the software, the inability to zoom and scroll within the software and the limitation of the size of measurable angle (< 90°).

Most importantly it was also felt that the application of this type of information technology should be developed further. It was envisaged that the system may be more useful if it were developed to respond to the need for new ways of assessing static joint angles, rather than being limited to merely assessing postural angles. This could only be achieved if the system was completely re-designed with specific reference to this new potential application.

Initially, the 'User line' and 'Reference line' of UM version 1.2 were re-designed as virtual measurement arms (VMA) (see Figure 2.3). This system configuration was the defining and differential feature of the new system. The concept of 2 VMAs with 'point and click' arms that could be manipulated independently, but remained relative to each other was innovative. Their construction formed a virtual goniometer which at that time was unlike any device previously reported.

This virtual goniometer concept was titled UM version 2.x and was considered as a completely different type measurement tool than that of the previously mentioned postural analysis tool (UM, version 1.x). Both having two distinct functions.

As version 2.x developed the system became unrecognisable from its version 1.x origins, with a completely new graphical user interface (GUI) (see Figure 2.2; 2.6), use of virtual measurement arms (VMA), different file importation and management technique and different mathematical modelling.

2.2.4. Developing the Uillinn Method© Version 2.x.

Development of the system from postural analysis system (version 1.2), to the current latest version of the joint angle measurement system (version 2.4) took place over a three-year period. The largest departures from the original system was creation of the VMA and their angle calculation operation. Other developments were based on practical challenges experienced when using the system. Each stage of development was met with major or minor software 'bugs', which are a normal process during software development. All of these were successfully overcome. The major evolutions of the system all outlined below.

The graphical user interface (GUI) was redesigned on multiple occasions during the course of the programme. This redevelopment also involved a new file access and storage features and resizing the measurement data display area.

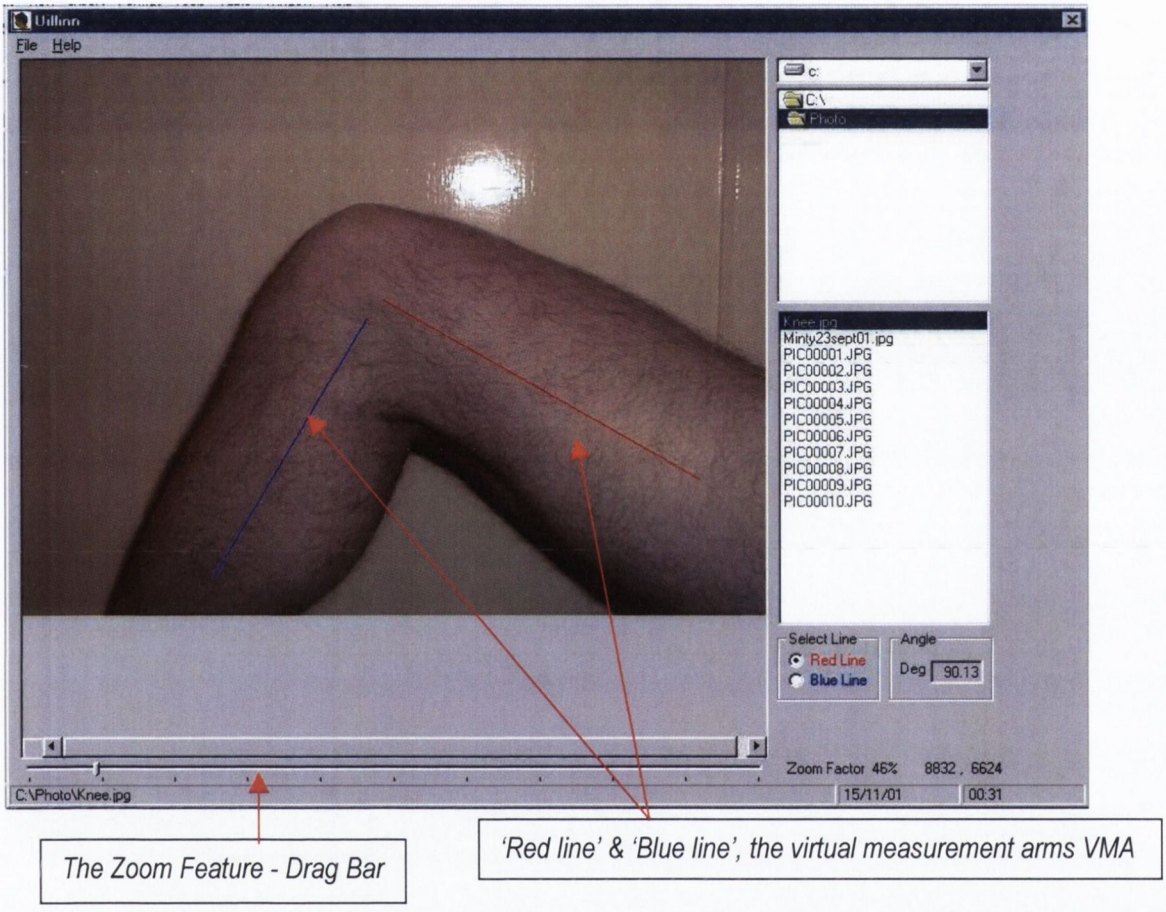


Figure 2.3. Version 2.1. of the Uillinn Method©.

The coordinate system used in version 1.x was based on measurement lines (Reference and User lines). The configuration used to manipulate these lines would not have lent itself well to the placement of the VMAs, which had to be much more versatile, due to the wide range of angles it would be likely to face. A red and blue VMA were developed to satisfy this demand. These could be selected individually and manipulated using a point and click system on the photograph (see Figure 2.3).

The design of the VMAs themselves changed as the project advanced. Distinguishing features were added to proximal and distal ends of the VMAs to assist in comprehending their orientation. The addition of a broken green line (see Figure 2.4) between the ends of each of the VMAs was added to versions above 2.2 to improve operations and appropriate placement about the joint.

An image zoom system was developed that allowed the user to magnify the photograph and scroll to different points of the picture. This enabled the user to scrutinise any aspect of the image closely.

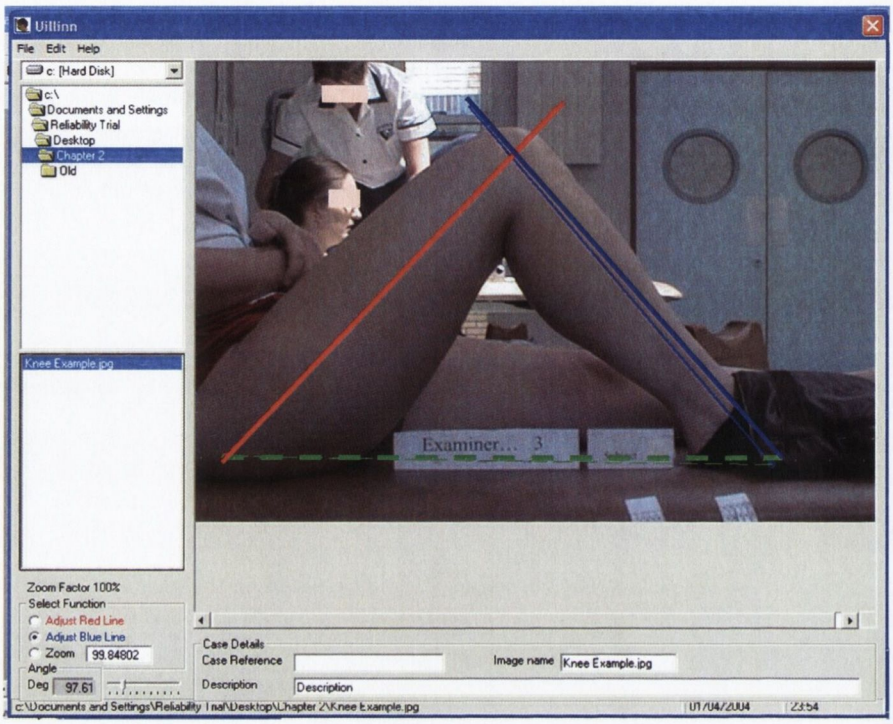


Figure 2.4. Version 2.2. of the Uillinn Method© (Parallel Approach).

A major limitation of version 1.x was that the system could not measure angles greater than 90°. This issue was related to the fact that VB calculates angles in radians. This is not a problem inherent to mathematics based on radians, but rather it was a function of the method in which the OS expressed radian angles (between zero and $\pm\pi$). This meant it could not easily express the obtuse angle between two lines and would, by default, express only the acute angle. This issue had the potential to produce and ambiguous or incorrect result and would have to be addressed if the system were to be applied to general somatic joint measurements which may commonly be greater than 90°.

The 'Radian to Degree Function' of the software was rewritten completely to initially determine what quadrant of the circle each of the VMAs were in and apply logic to correctly calculate their value in angular degrees relative to the OS's version of the zero vector. This step had to be completed before the relative angle between the 2 lines could be computed. This calculation was made difficult when the angles of 0^{o}, 90^{o}, 180°, 270^{o} and 360^{o} were involved as these angles could be represented as

2 different radian values, due to the fact that they lie on the boundary between the different quadrants. A subroutine was written to manage this eventuality.

In version 1.x, identification of the correct photograph was deemed unpractical and laborious, as the user could only see the JPEG file name. To overcome this problem a file preview function was developed which allowed the user to obtain a 'thumbnail' view of any selected JPEG file prior to final rendering of the image.

The standard Minimise and Maximise taskbar functions were added to make it more consistent with the typical Windows format, this would improve its usability when there was more than one software application running on the PC simultaneously.

Finally, the latest version of the software (version 2.4) includes the addition of 3 virtual assistive lines that enable the user to select the mid-point of a limb. These black lines each of which has a different coloured dot associated with it (yellow, green and blue) were added to guide the user when applying the VMAs to the limb segment (see Figure 2.6). The coloured dot in each of the lines is programmed to remain in the exact mid-point of the line itself, irrespective of how long the user adjusts the lines to be on the screen. The rationale for this feature is discussed below (2.3.5).

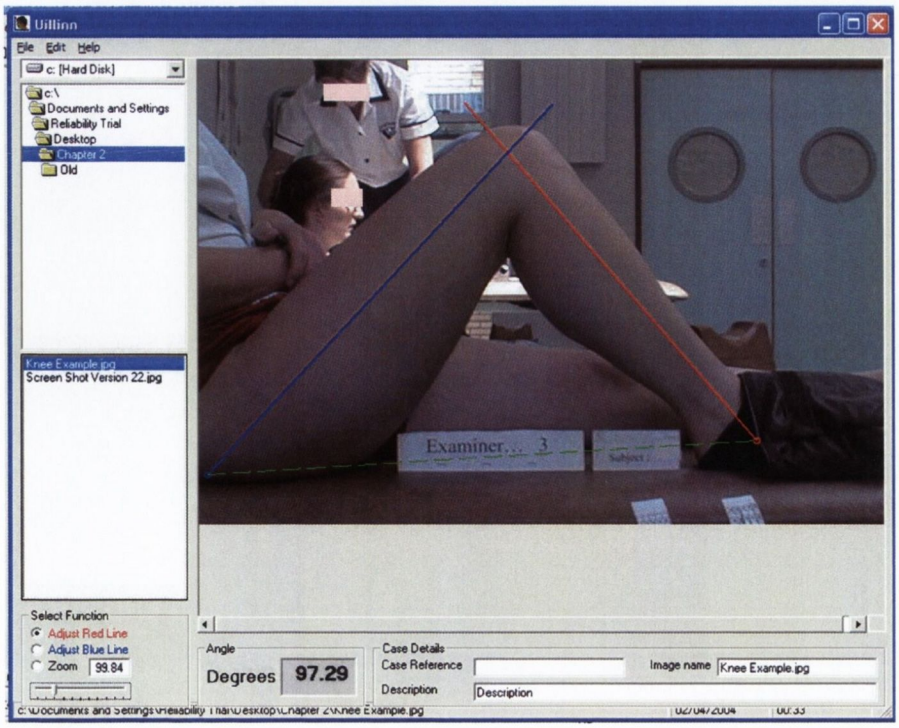


Figure 2.5. Version 2.3. of the Uillinn Method© (Parallel Approach).



Figure 2.6. Version 2.4. of the Uillinn Method© (Mid-point Approach).

2.3 Applying the Uillinn Method© to the Human Model

Previous researchers who have employed photographic based methods of joint angle assessment have invariably employed adhesive skin markers to highlight anatomical landmarks on their photographs (Fish and Wingate, 1985; Cameron et al, 1993; Gajdosik et al, 1994; Marks and Karkouti, 1996; Marks, 1996; Marks and Quinney, 1997; Karkouti and Marks, 1997). This practice has been done to highlight specific anatomical reference points from which, it has been proposed, measurements should be made (Norkin and White, 1995; 2003). However, the literature has not clearly shown that adhesive skin markers actually improve the performance of photographic based measurements methods. On the contrary, some studies have called into question the validity (Sati et al, 1996) and reliability (Marks and Karkouti, 1996; Karkouti and Marks, 1997) of using adhesive skin markers. Due to the lack of evidence for this practice it was decided that other approaches should be considered to assist the examiner during the positioning of the VMAs on a photograph.

A number of different approaches toward the application of VMAs at the knee and elbow were contemplated. After consideration it was felt that at least 5 different methods would merit further investigation, 2 of these methods used adhesive skin markers and 3 did not.

2.3.1. The 3 Marker Approach (3Mks)

Norkin and White (1995; 2003) recommend that certain anatomical points be used as reference coordinance when assessing segmental joint angles with the universal goniometer (UG). At the knee these points are the greater trochanter of the femur, the lateral femoral condyle and the lateral malleolus at the ankle. At the elbow the acromion process of the scapula, the lateral epicondyle of the humerus and the styliod process of the radius are recommended (see Figure 2.8).

These points are not directly detectable on the 2-dimensional photographic image, to overcome this problem a landmarking technique has been used. As there are a total of 3 points each for the upper limb and lower limb, this approach to applying the UM was called the '3 Marker approach' (3Mks).

Three adhesive skin markers are placed over anatomical landmarks. This system allows clear reference points for the angular measurement from the photograph.

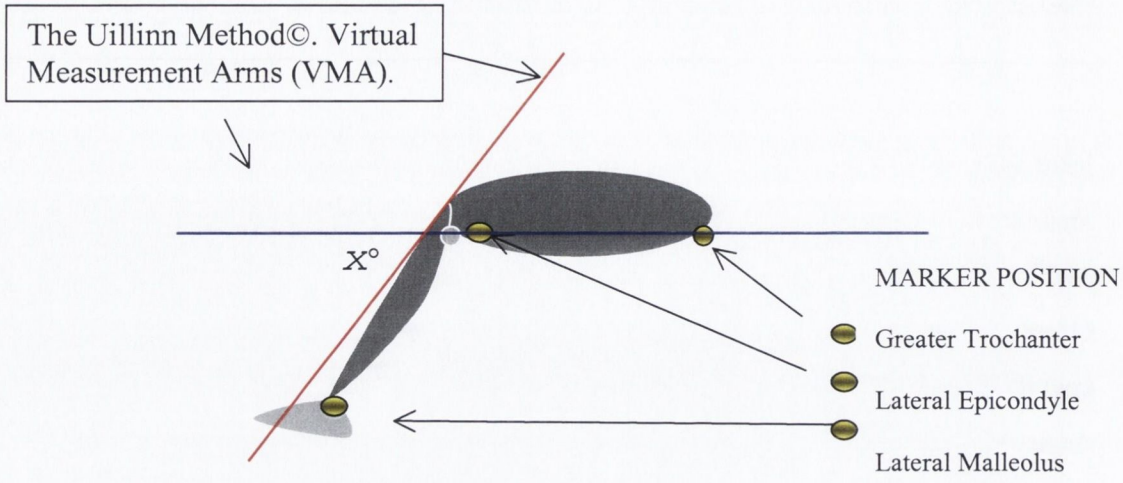


Figure 2.7. The '3 Marker Approach' to Applying the Uillinn Method©.

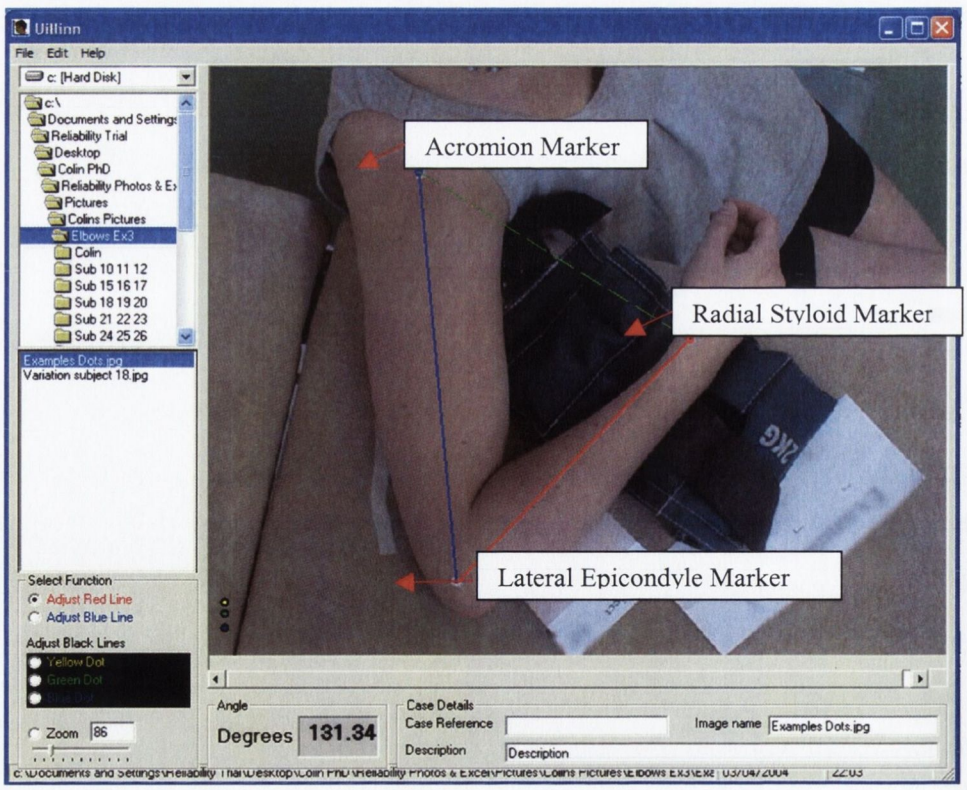


Figure 2.8. The '3 Marker Approach' at the Elbow.

2.3.2 The 4 Marker Approach (4Mks)

There has been some controversy in the literature regarding the identification of the dynamic joint centre of the knee (Norkin and Levangie, 1992) and how the changeable nature of this point effects static joint angle measurement (Enwemeka, 1986). Because of this controversy it was felt that an alternative approach should be considered that negated the requirement of the joint centre as a reference point, usually identified as the lateral epicondyle of the femur.

Joint angle measurements indicate the angle of incidence of two rigid bony segments, if two reference co-ordinance are placed on each of the rigid segments, the exact location of the axis of movement becomes unimportant. Calculations of joint centres have been done in this manner previously via a 'technical marker system' or a 'CAST' system (Besier et al, 2003). It may be possible to make measurements based solely on the orientation of the two segments (see Figure 2.9). If one distal and one proximal point were highlighted on each of the body segments an angular relationship between the segments could be calculated.

This approach would use 4 different adhesive skin markers, at the elbow the landmarks would be the acromion process and the lateral epicondyle for the proximal segment, and the head of the radius and the radial styloid process for the distal segment (see Figure 2.10). At the knee the proposed points would be the greater trochanter and the lateral epicondyle of the femur and the lateral malleolus and the head of the fibula (see Figure 2.9). As this method used 4 adhesive landmarkers it was termed the '4 Marker approach' (4Mks).

Four adhesive skin markers are placed over anatomical landmarks. This system allows clear reference points for the angular measurement from the photograph.

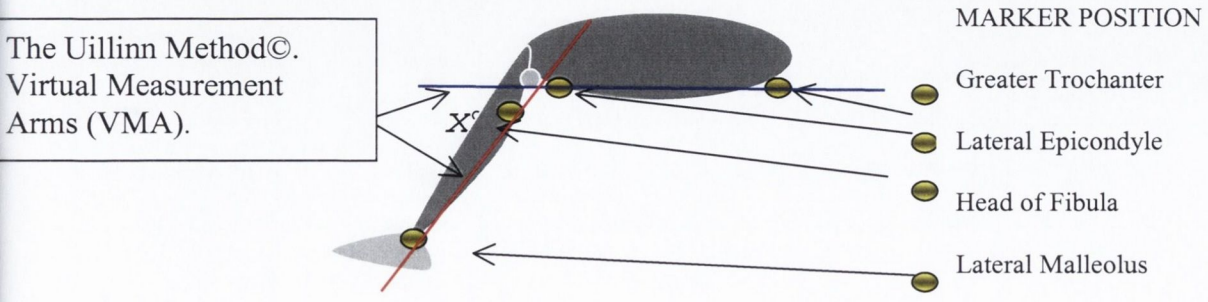


Figure 2.9. The '4 Marker Approach' to the Uillinn Method©.

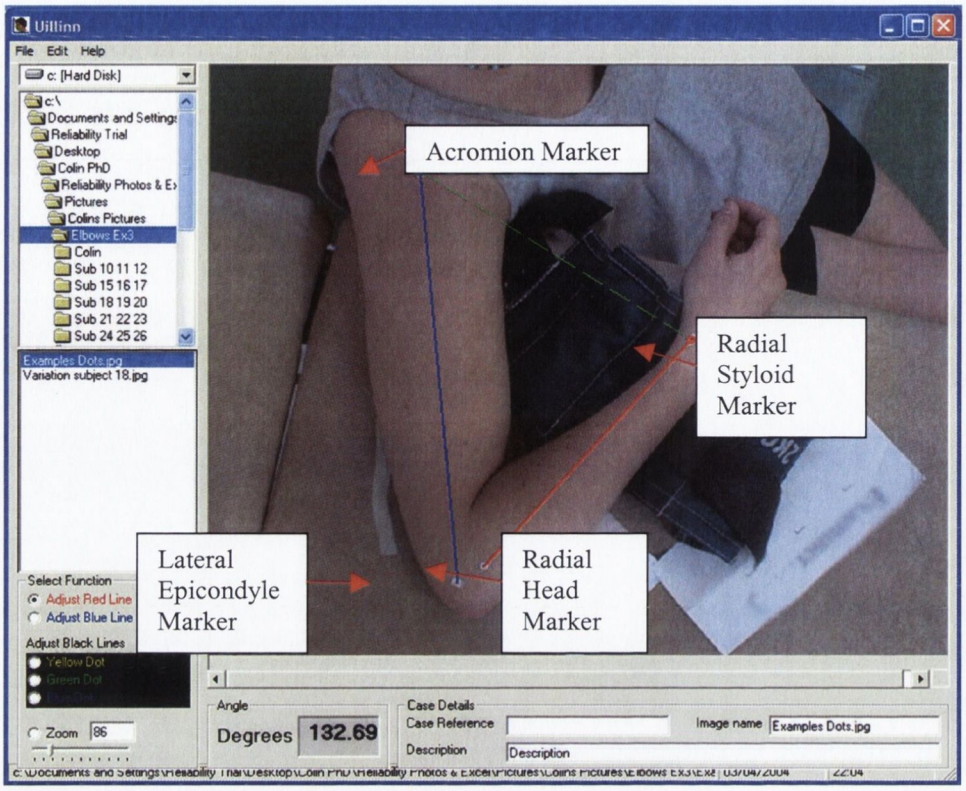


Figure 2.10. The '4 Marker Approach' at the Elbow.

2.3.3. The No Marker Approach (NMks)

As previously discussed there is little evidence to support the use of adhesive landmarkers as a means of efficiently identifying specific anatomical landmarks. However, there does not appear to any research published that fully investigated reliability and validity issues associated with photographic measurements that did not use adhesive skin markers. Such an approach may allow the skill of the user decide where to place the reference points for an angular calculation. Due to the lack of research into this area it was decided that it was legitimate to investigate such an approach to the application of the VMAs.

For the purposes of standardisation, the anatomical reference points proposed by Norkin and White (1995; 2003) would be used to calculate the joint angle. The users employing this approach would be asked to use their anatomical knowledge to identify these points on the photographs. The user would not have the benefit of adhesive skin markers to guide their placement of the VMAs (see Figure 2.11).

The Measurement arms of the Uillinn Software are aligned to be approximately at the Greater Trochanter, the Lateral Epicondyle and the Lateral Malleolus.

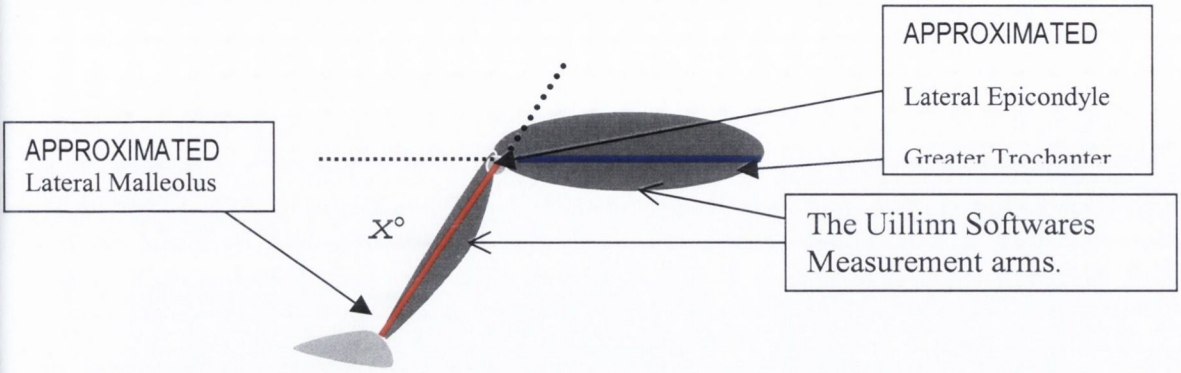


Figure 2.11. The 'No Marker Approach' to the Uillinn Method©.

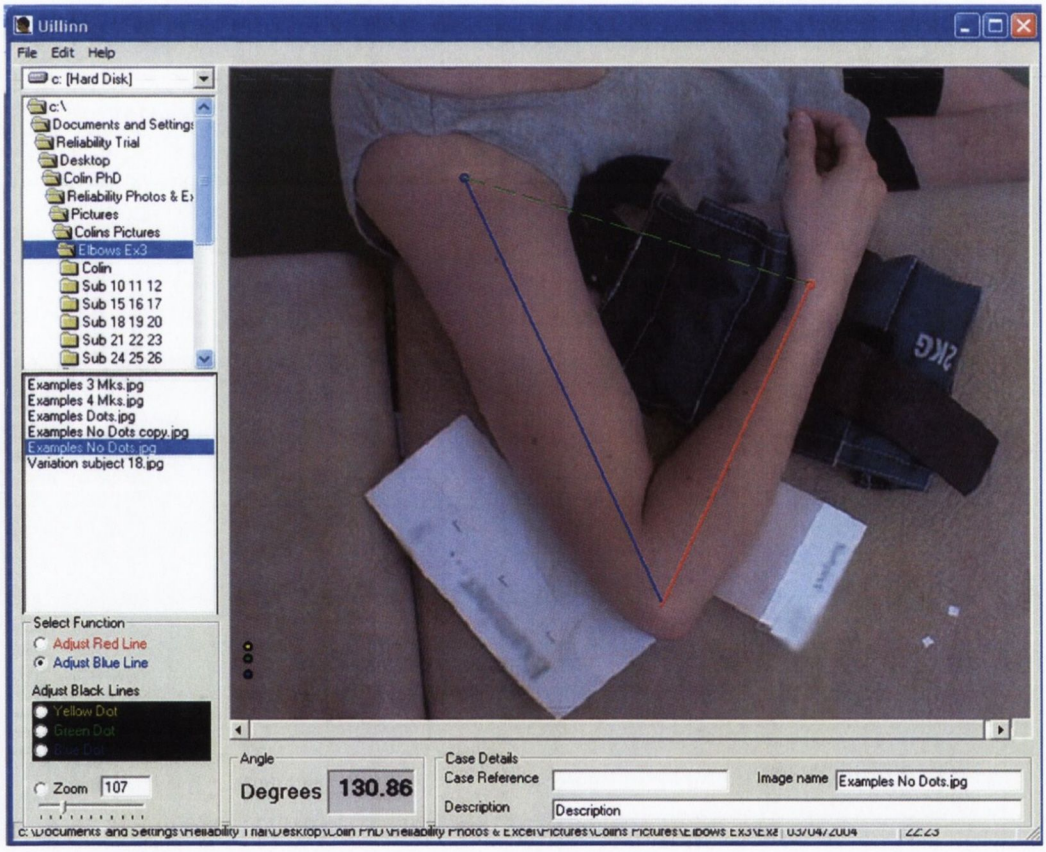


Figure 2.12. The 'No Marker Approach' at the Elbow.

2.3.4. The Parallel Approach (Parallel)

Most approaches to joint angle measurement described to date recommend the use of specific anatomical reference points to guide the examiner. However,

Grohmann (1983) compared two different methods of assessing joint angles using the universal goniometer (UG). One approach investigated was the standard lateral approach that used anatomical reference points, while the other was the 'over-the-joint' approach. This method ignored specific reference points, including a joint axis, instead placed the arms of the UG on the dorsal aspect of the limb. Both approaches displayed equal reliability implying that approaches to joint angle measurement that do not use specific anatomical reference points may hold potential for the placement of the VMAs of the UM.

The 'Parallel' approach is one such method of applying the VMAs of the UM software to the photographic image. This approach demands that the user place the VMAs along the longitudinal axis of the limb segment, therefore aligning it to be approximately parallel to the underlying bone shaft. No specific reference points are recommended, rather the user must try to judge where exactly the longitudinal axis of the limb segment is situated (see Figure 2.13, 2.14).

The Measurement arms of the Uillinn Software are aligned to be parallel to the limb segments above and below the knee. No specific reference points are used. The anterior surface is used as a guide.

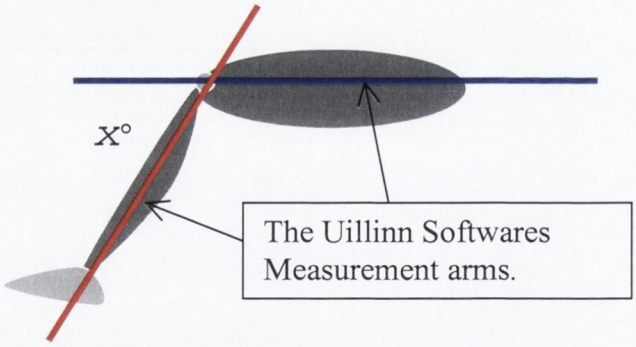


Figure 2.13. The 'Parallel' Approach to the Uillinn Method©.

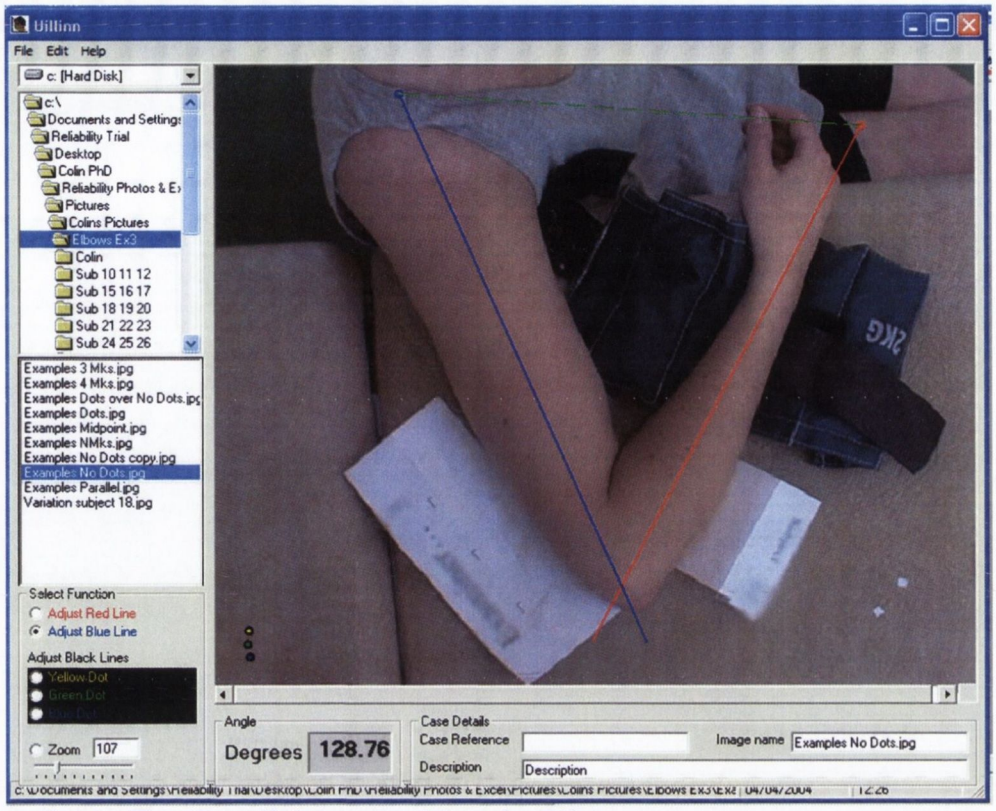


Figure 2.14. The 'Parallel Approach' at the Elbow.

2.3.5. The Midline Approach (Mid-Point)

The last approach to applying the UM to human joint measurement that was considered was based in some part on the Parallel approach. It was experienced during early testing that when conducting multiple measurements without using adhesive landmarkers there was a larger demand on the examiners mental concentration than when using those approaches that had adhesive skin markers. This was probably because the examiner had to continuously rationalise the exact position in which to place the VMAs, whereas when the examiner was guided by the adhesive markers all the examiner had to do was use the mouse to click on those points. It was envisaged that another approach should be developed that helped guide the examiner during the VMA placement. A concept was devised and programmed that provided the examiner with 3 adjustable guide-lines each with a coloured 'Dot'. When each of the guidelines were placed on the photograph the dot would indicated the exact mid-point of that guide-line. Therefore the examiner could now be precise about the location of the mid-point of the limb segment and could use this technique to position the VMAs. This tool was built into the UM software version 2.4.

The procedure recommended for using this 'Mid-point' approach was; at the elbow to place one of the mid-point lines across the proximal upper arm roughly at the level of the insertion of the deltoid muscle, another mid-point line was to be placed across the wrist, and the last mid-point line, would be placed about the elbow joint between the most posterior aspect of the olecranon and the edge of the cubital fossa, the VMAs could then be aligned directly to the mid-point dots (see Figure 2.15).

At the knee the suggested method of application was that the first mid-point line should be placed on the proximal thigh segment just distal to the approximate level of the femoral triangle, the most distal mid-point line would be placed at the level of the lateral malleolus, the final line would be positioned about the knee joint itself between the most anterior prominence of the patella and the edge of the popliteal fossa, again the VMAs would then be aligned to the dots (see Figure 2.6, page 101). It was hoped that using this tool the examiner could be sure that VMAs were placed along the mid line of the limb segment.

One difficulty experienced with this method during a preliminary test was that when subjects were measured with their joints approaching end range flexion the outline of the joint fossa would be indistinguishable making it difficult to accurately place the mid-point guide at the joint. This problem was more evident in subjects with excess adipose tissue. In these cases the line of the skin fold would be used instead of the edge of the joint fossa (see Figure 2.15).

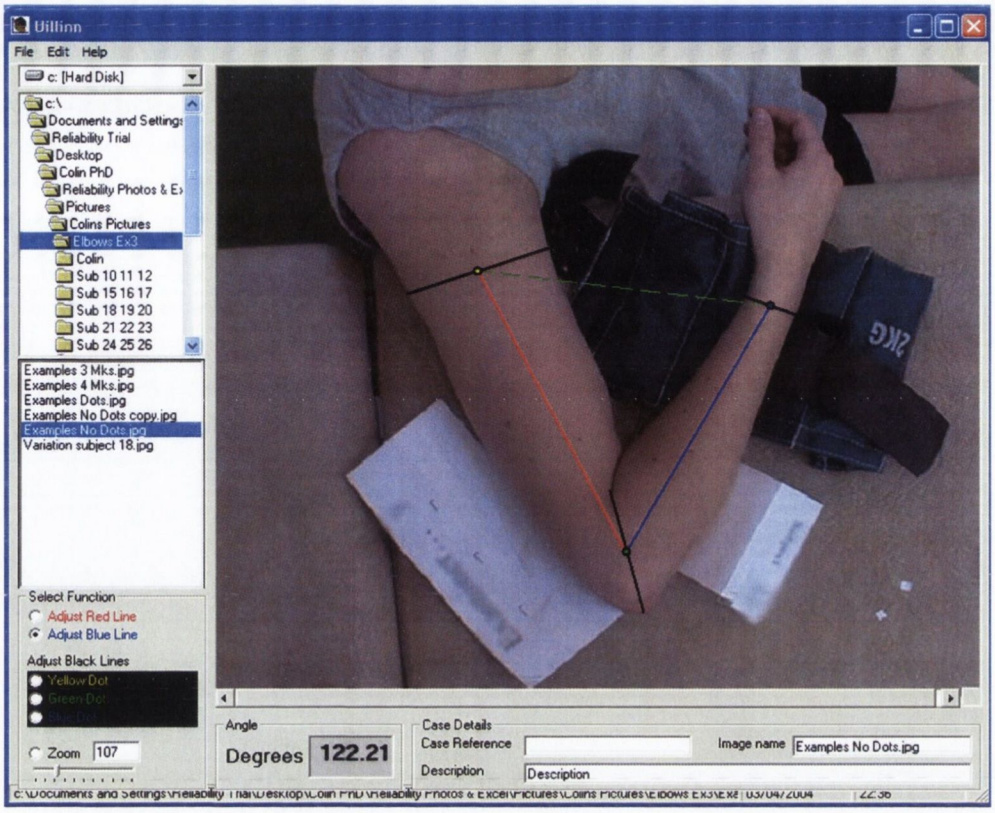


Figure 2.15. The 'Midline Approach' at the Elbow.

2.4. Discussion

It had been identified in the literature that there are inherent difficulties to assessing static joint angles with conventional tools, such as the universal goniometer (UG) (Grohmann, 1983; Enwemeka, 1986; Brosseau et al, 1997; Szulc et al, 2000; Brosseau et al, 2001). Therefore, there was a clear need to improve upon our current ability to measure static joint angles.

Conventional photographic based system have been reported as was being well suited to measuring segmental angles (Fish and Wingate, 1985; Braun and Amundson, 1989; Braun, 1991; Cameron et al, 1993; Gajdosik et al, 1994; Raine and Twomey, 1994,1997; Marks and Karkouti, 1996; Marks, 1996; Marks and Quinney, 1997; Karkouti and Marks, 1997). However, they have also been criticised as being limited towards clinical measurements (Norkin and White, 1995; 2003). Nonetheless, it was felt that due to dramatic technological advancements in photography, i.e. the invention of the digital camera, the value of such a system should be revisited as an option for clinical measurement.

A photographic based system seemed to offer several benefits including the ability to capture both dynamic and static joint motion angles, to measure multiple joint movements simultaneously and to freeze the specific instant of end-of-range movement allowing repeated measurements by the same or different examiners. In addition, photography is a hands-off technique which may help to lessen examiner bias during a measurement. The photograph provides hardcopy evidence of the movement ability of a subject at the time of photography, which may be beneficial when performing repeated measurements on a subject over time.

However, it was also noted that there were some probable disadvantages to photography as a means of measuring joint angles, which included the conversion of the actual 3-dimensional status of the subject to a 2-dimensional photographic image. This transformation results in the loss of depth perception and therefore the loss of the rotational perspective of the subject. Gray et al, (1991) showed miscalculations of angular data due to the effects of rotation between an object and an X-ray image, this finding would almost certainly have implications for a photographic based measurement system. Imperfections in camera design, including the pin-cushion effect (see Figure 1.7) or poor skill when taking the photograph

resulting in image shake or parallax error may also lead to image misrepresentation and therefore lead to measurement error (Scholz, 1989; Dunlevy et al, 2000). Another issue associated with photography is the fact that the photograph provides an image that is a fraction of the size of the actual limb. It was felt that a loss of scale could make measurements less sensitive and therefore may hold relatively more error.

The fast and cheap system of photography presented by digital photography could offer a new dimension to clinical and research methods of static measurement. The UM was developed to exploit this new technology. It was originally conceived as a tool for measuring postural angles of the head and neck, but quickly progressed as a proposed tool for the measurement of joint angles generally. The system involves the digital photography of a joint, the transfer of the digital image to a PC and then the operation of a software application to analysis the joint angle. The digitisation of data may help the accuracy of the measurement process, and may help to eliminate reading and recording errors, which have been linked to measurement error (Mayerson and Milano, 1984; Williams and Callaghan, 1990). However, there may be other issues related to photographic measurements that give rise to measurement error, it remains unclear how the performance of this tool compares to the UG.

The software for the UM went through a formal version development programme over a 4 year period. Probably the most difficult question posed during the development of the UM was the how best to apply the measurement arms (VMA) of the software to the 2-dimensional image of the joint. It is standard practice in the research literature to attach skin markers on specific anatomical landmarks for the purpose of subsequent joint angle measurements (Fish and Wingate, 1985), However, there does not seem to be strong evidence to suggest that this technique will improve results (Sati et al, 1996; Marks and Karkouti, 1996; Karkouti and Marks, 1997). Therefore, it was felt that other approaches should be investigated.

Five different approaches were devised to tackle the problem of how best to apply the UM to human measurement, 2 of the approaches required the attachment of adhesive skin markers prior to photography, and the remaining 3 did not. One

major advantage of 3 of the 5 different approaches to the UM (the Parallel, the Mid-Point and the 4-Marker), over other joint measurement methods, was that they do not use a fixed point as a reference for the angular calculations but allow for the true dynamic nature of the joint axis.

Of the published literature that used photographic based measurement systems, few undertook to investigate the validity or reliability of this type of clinical measurement (Fish and Wingate, 1985; Marks and Karkouti, 1996; Szulc et al, 2000; Hayes et al, 2001). Fish and Wingate (1985) was one study identified as dealing with these issues, they showed that photography performed significantly better than the UG, however, Hayes et al (2001) disagreed, their results showed similar levels of performance between visual estimation (VE), UG and a conventional photographic method, as judged by the standard error of measurement (SEM). Unfortunately direct comparisons between these studies may not be appropriate as they used significantly different methodologies and data analysis tools.

Further investigation is warranted due to the fact that the disagreement presents a large deficit in our professional knowledge as to the expected performance of a photographic based measurement system. It was felt that was necessary to scientifically investigate the levels of validity and reliability of a newly proposed measurement tool such as the UM.

2.5. Conclusion

Information technology (IT) plays an increasingly important role in the provision of health care. New systems are constantly being developed that purport to improve the delivery of health care services. Physiotherapy already uses IT extensively to improve upon their health care provision. Joint angle measurements are an essential part of this practice. Current tools available to physiotherapists to perform joint angle measurements have not shown satisfactory levels of validity and reliability.

The Uillinn Method© (UM) (version 2.x) was developed as a potential method to improve the practice of joint angle measurements in a clinical setting. It employs a custom-built computer software program that can render a JPEG file allowing a user to manipulate virtual measurement arms (VMA) and place them anywhere over that image. When the VMAs are aligned appropriately over limb segments a calculation of their angular orientation in relation to each other can infer the angle of the limb segments and therefore provide information on the joint angle.

The simple image capture using digital photography, the file transfer from camera to PC, and data calculation using the software VMAs form the operational design of the UM. To date the error associated with these 3 components was largely unknown.

Chapter 3

METHODS: RATIONALE FOR THE EXPERIMENTAL WORK.

3.1 Introduction

The development and delivery of health care should be based on a scientifically proven evidence-base. A best practice model will ensure that a successful and correct management strategy for a particular dysfunction is accomplished. Appropriate clinical research is the best way to establish an evidence base.

The assessment of joint range of motion (ROM) is an essential aspect of the diagnostic and management process for different patient groups and is used to infer levels of function (AAOS, 1994). Therefore the tools used to obtain this type of joint angle data are an essential part of the delivery of optimal health care. It is vital that clinicians understand the limitations of the tools they use and that there is clear knowledge of the expected error associated with each of these tools (Rothstein, 2001). Only well designed scientific research can furnish the clinicians with this information.

The concepts of validity and reliability are fundamental components to the provision of an evidence base for clinical measurement tools (LoBiondo-Wood & Haber, 1998). The concept of validity, of which there are many different types, basically indicates whether a tool is measuring what is it supposed to measure (Sims and Arnell, 1993). Testing validity for a tool that is used to calculate objective clinical data is best achieved by assessing the agreement between it and a gold-standard measurement method (Atlman, 1999; Berrymann Reese and Bandy, 2002; Norkin and White, 2003), if the assessments are made at approximately the same time concurrent validity may be claimed (Sims and Arnell, 1993; LoBiondo-Wood & Haber, 1998). A strong concurrent validity provides the best evidence for the overall validity of a measurement tool (Norkin and White, 2003).

The term reliability denotes how repeatable measurements are, under the same set of circumstances. There are two main types of reliability, intra-rater reliability and inter-rater reliability. The former term refers to nature of agreement between repeated tests when they are made by the same examiner (int*ra*-rater). Int*er*-rater reliability deals with the concordance between measurements when they are made by different examiners (Sims and Arnell, 1993).

The most commonly used joint angle measurement tool in clinical practice is the universal goniometer (UG) (Coxford et al, 1998; Burton et al, 1999), despite

much published research there is a poor definitive account as to the performance of this tool in terms of validity and reliability. This controversy represents a clear deficit in our knowledge base.

The most common problem associated with previous research that investigated the performance of measurement tools, especially the UG, has been the use of incomplete statistical analysis. Other common methodological problems included limited sample size or poorly controlled study procedures (see Table 1.2 to 1.6). Any of these common pitfalls greatly limit the relevance of the findings from previous studies.

A new tool designed to perform clinical measurements, such as the Uillinn Method© (UM), should be investigated for issue of validity and reliability in order to provide an evidence base for that tool. It is vital that a study designed to perform such an investigation should avoid the common methodological errors seen in previous research.

One of the main objective of this thesis was to fulfil the requirement for clear data regarding the validity and reliability of both the UG and the new UM. This chapter will outline specific deficits in our professional knowledge to justify a rationale for this current investigation.

3.1.1 Gold-Standard Measurements

When a new measurement tool is introduced it is important to compare its performance against the system that is already in standard use (Bland and Altman, 1986). However, if the standard system does not fulfil all the criteria needed to deliver an unquestionable measurement, such as strong evidence for validity and reliability, it may not be classified as a gold-standard. When Rowe et al (2001) investigated the clinical validity of the electrogoniometer (EG) against a strong standard in the field of motion analysis, the Vicon motion analysis system, they identify a limitation of their study as being the fact that both systems track the overlying soft tissue about a joint rather than the a direct analysis of the dynamics of the bony tissues themselves. Contrary to this Rheault et al (1988) and Petherick et al (1988) believed that their studies investigated the concurrent validity of the UG and the fluid filled goniometer (FFG) by comparing them against each other, they proposed the UG as the gold-standard. Russell et al (2003) also identified the UG as

a gold-standard in their validity investigation of the internet based goniometer (IBG). However, proposing the UG as a gold-standard does not seem to stand up, firstly because the UG does not have the ability of to directly consider bony movement, and secondly because the literature has questioned its performance for valid and repeatable measurements. Norkin and White (2003) agree with Rowe et al (2001) and state that only imaging systems that depict the exact orientation of the bones themselves should be considered as the gold-standard against which the validity of joint angle measurement tools should be judged. Therefore the design of this study aspired to use a radiographic method of joint angle measurement as a gold-standard against which to assess the validity of the UG and the UM.

3.1.2 Comparisons Between Methods.

Nonetheless, there remained a benefit in comparing the performance of the new UM measurement tool against the more established standard UG (Bland and Altman, 1986). This comparison was justified for a number of reasons, firstly, it was necessary to establish if these tools would be inter-changeable, but more importantly it was necessary to assess which of the tools offer the examiner the most precise and repeatable measurement options.

3.1.3 Deficits in Knowledge for Universal Goniometry.

It seems unavoidable that some amount of error will be present in joint angle measurements (Hopkins, 2000; Rothstein, 2004), different researchers have put forth figures of clinically acceptable error, which range from Mayerson and Milano's (1984) proposal of approximately 5° to Ishii et al's (1995) figure of 2°. Nigel-Keenan et al (2004) set their limits of clinically acceptable error at 10°, as measured by the Bland & Altman's limits of agreement (LOA).

Academically, it may be the case that clinicians should not perceive any error as 'satisfactory' but be constantly striving to minimise error. Therefore setting 'limits of acceptable error' may be improper. Nonetheless, it appears that some amount of measurement error is unavoidable and the recognition of this error is important. The literature states that measurements of different joints will have different levels of error associated with them (Boone and Azen, 1979; Fish and

Windgate, 1985; Riddle et al, 1987; Bovens et al, 1990; Watkins et al, 1991; Youdas et al, 1993; Nigel-Keenan et al, 2004). Furthermore different measurement techniques hold different properties (Clapper and Wolf, 1988; Armstrong et al, 1998). Rothstein (2004) states that no measurement has perfect validity or reliability, rather these qualities sit somewhere along a continuum.

It seems that as error must be thought of as unavoidable, clinicians should endeavour to arm themselves with knowledge regarding the specific quantities of this unavoidable error. This in turn could be used to check their own measurement practice and thereby improve their clinical skills. This aspiration implies that there is a need to establish 'levels of *inevitable* error' rather than setting 'limits of *acceptable* error' for each measurement tool, at each joint motion. Currently, this data is not clearly interpretable from the literature with regard to most measurement tools. Therefore there is a need to establish 'levels of inevitable error' for both the UG and indeed the new UM.

Aside from the lack of firm data for the performance of the UG in general, it appears that some aspects of joint angle measurement have never been investigated. While there have been concurrent validity studies, based on radiography for the lower limb (see Table 1.9) there does not appear to be any such study that considered measurements of the upper limb.

Boone and Azen, (1979) state that the UG has better reliability for measurement of the joints in the upper limb than for the lower limb. Comparisons between the levels of validity of upper limb versus lower limbs measurements have not been possible due to the absence of upper limb data. This presented a need to establish concurrent validity for the UG at the upper limb using a radiographic methodology and compare levels of validity between upper limb and lower limb measurements.

Enwemeka (1986) controversially reported that measuring knee extension is less valid than assessments of knee flexion. However, this claim appears to warrant re-investigation as it seems the reported data could be interpreted differently to support an opposing view (1.2.1). Additionally, it is not clear whether this proposed phenomenon of decreasing validity for joint extension also is true for the elbow. Therefore, a study to investigate a proposed diminishing validity for joint extension at the knee and elbow is justified.

In summary, the knowledge deficits highlighted here include the lack of strong data regarding the expected error of the UG and other measurements tools for the knee and elbow, and to examine relative performances of the joints in different ranges.

3.1.4 Deficits in Knowledge for Photographic Measurement Methods.

The literature concerned with validity and reliability issues for a photographic based approach to measurement is sparse. In 1985, Fish & Wingate proposed that there was a clear lack of knowledge with regard to the performance of photographic methods of osteokinematic measurements, unfortunately since then there have only been a few studies that have attempted to fill this gap (1.2.9). An extensive 'Medline' literature search conducted in April 2004 was unable to find any study that investigated the concurrent validity of photographic measurements using a radiographic methodology. A combination of keywords were used in the search including 'Validity', 'Photography', 'Radiography' and/or 'X-ray'.

Karkouti and Marks (1997) report that their results into the reliability of conventional photographic measurements were favourable enough to warrant further investigation for the clinical use of photography (see Table 1.3; 1.5). They continue that this type of analysis may be more practical if an instantaneous method of photographic delivery was available. However, this claim has not been substantiated and it is not clear whether clinicians would find this type of measurement tool of value to their practice.

A potential problem for photographic based systems is the difficulty associated with the identification of anatomical reference points on the limb being measured. The main solution to this has been the use of adhesive skin markers placed on the landmarks prior to photography, however, the precision associated with this technique has been questioned (Karkouti and Marks, 1997; Sati et al, 1996).

No published studies were found that had attempted to conduct either reliability or validity measurements without the use of adhesive skin markers. Therefore, it is not clear if these markers are necessary to improve the accuracy of measurements. Comparisons between the five different approaches to the application

of the UM (2.3) would provide valuable information regarding this question and also establish which of these methods perform best overall.

Most studies that have used photographic measurements systems indicate clearly the height that the camera was set at and the distance between it and the subject. While it seems appropriate to predict that parallax error or incorrect angular perspective will have an effect on measurements, it is not clear how these issues will actually effect the angular measurements. Gray et al (1991) investigated this issue with regard to radiography, however, there does not appear to be any research into the effects of inappropriate camera to subject alignment for a photographic measurement system.

An investigation that aimed to quantify how a photographic measurement system, such as the UM, would respond to different procedural stresses, i.e. the introduction of rotation or parallax error, would provide information that could help to establish the limitations of such a system.

This type of investigation would first need to establish how the sensitive angular measurements made with the UM are without any stresses placed on it. This data would then form a baseline which would represent the most optimal measurement result that could be expected from the system. The impact of stresses introduced by the photographic procedure could then be gauged by comparing measurements made under the most optimal photographic conditions against those made under different procedural variations. This type of study would need to be completed prior to other clinical investigations of validity or reliability as the information gained from this initial study would provide vital data upon which to base a sound operational procedure for the system.

3.1.5 Clinical Usage of Measurement Technology.

Some of the more high-tech static joint angle assessment tools that have been investigated in the literature do not appear to have been widely adopted by physiotherapy departments generally. This despite promising research regarding their performance. The reason for this poor uptake may be multifactorial, although one proposed reason may be that physiotherapists feel there are barriers to using technology during assessments. These barriers may include, time restraints, lack of confidence using technology, distrust of the equipment, etc. It seems appropriate that

a new measurement tool should attempt to assess barriers to its use in a working environment. Such a goal may be achieved by quizzing physiotherapists who have experience of using the system clinically.

3.2 Specific Aims

This study was initially devised to determine the appropriateness of the new Uillinn Method© (UM) for the measurement of human joint angles. When the literature was reviewed, it was determined that levels of validity and reliability for currently used measurements tools, such as the UG, were unclear. In order to address these uncertainties and therefore provide meaningful comparison data for the performance of the UM, specific aims for an investigation were drafted. These aims were grouped into 4 different question categories; a *Calibration* question for the UM, a *Validity* question for the UM and the UG, a *Reliability* question for the UM and UG, and finally a *Clinical Use* question for the UM.

A series of investigations would be required to address these questions.

3.2.1 Outline of the Study Design

The study proposed to deal with each of these 4 different questions under separate experimental categories;

i. A calibration trial for the Uillinn Method©.
ii. A reliability trial for the Uillinn Method©.
iii. A concurrent validity trial for the Uillinn Method©.
iv. A clinical usage trial of the Uillinn Method©.

Initially, trails (i, ii, iii) were proposed as quantitative investigations that were to provide their results numerically. The latter trial (iv) was proposed as mainly a qualitative trial which was to facilitate feedback for therapists who had used the system clinically.

Trials (i) and (iii) would clearly define a control data against which to base comparisons of the performance of the UM and UG. The 2nd trial (II), a reliability trial, would seek to outline the relationship of 2 or more sets of repeated test data.

Fundamentally, all of these 3 trials would be similar in that they would each seek to assess the agreement between two or more datasets and draw there conclusions from the observed levels of accord between them. The future performance of the UM would be predicted from this data.

Therefore, it was vital that appropriate inferential statistical analysis were used in the methodology. In order to identify the most optimal analysis tools for this

purpose the published literature was examined (1.4.4). It was decided that the data should be scrutinised in terms and random and non-random error under the specific aims listed below.

The Aims of a Calibration trial for the UM;

1. To determine the potential accuracy of the UM under optimal conditions.
2. To compare error for the UM when tripod mounted versus free-hand photography is used.
3. To investigate the influence of the pillow effect, lens imperfections and parallax error on the UM.
4. To investigate the influence of camera or target rotation on the UM.
5. To draw up procedural guidelines for the photographic component of the Uillinn Method© based on researched evidence.

The Aims of a Reliability trial;

1. To compare the levels of int*ra*-rater reliability for the 5 different approaches of the UM and the UG for joint angle measurements at the elbow and the knee.
2. To compare the levels of int*er*-rater reliability for the 5 different approaches of the UM and the UG for joint angle measurements at the elbow and the knee.
3. To report the reliability data with appropriate descriptive and statistical tools, including the Bland and Altman 95% limits of agreement, the repeatability coefficient and measurement typical error (TE).
4. To report on issues regarding the reliability of re-applying adhesive skin markers by the same examiner for the purpose of measurements.

The Aims of a Validity trial;

1. To investigate concurrent validity of both the UG and the UM, for joint angle measurements of the elbow and knee by making comparisons with a radiographic procedure.

2. To compare the performance of each of the 5 different approaches to the UM and the UG against the radiographic data and to elect the most valid measurement method for both the elbow and the knee.
3. To report the validity data of the UM and UG with appropriate descriptive and statistical tools, including the Bland and Altman 95% limits of agreement and measurement of typical error (TE).
4. To investigate whether applying adhesive skin markers to anatomical landmarks improves the validity of photographic based measurement.

The Aims of Clinical Usage Trial;

1. To facilitate a 3 month trial of the UM in a physiotherapy outpatient department.
2. To provide training to a group of physiotherapists on the operation of the UM.
3. To question the physiotherapists regarding their experiences of using the system.
4. To conduct an audit of the number and type of patients assessed during the trial period.

3.2.2 Data Analysis

Agreement would be tested by determining random and non-random error using Bland and Altman's repeatability coefficient (RC) and the 95% limits of agreement (LOA) tests. Both of these tests assume the data is from a normal distribution and that they are homosedastic, i.e. the magnitude of measurement value does not influence the size of the difference between the measurements (Atkinson and Nevill, 1998; Altman, 1999; Hopkins, 2000). A limit of 2.5% of the averaged measured values would be used as a limit to detecting the presents of SB (Batterham and George, 2000). If this limit wasn't exceeded the RC would be used in place of the LOA. These tests would be accompanied by appropriate graphs to illustrate the findings.

As Bland and Altmans (1986, 1996) other systems of analyses (RC, LOA) were already elected for use in this current study, it was sensible to use the SE formula proposed by them. Therefore, all SE data put forth in this study were calculated using the formula $SE = \sqrt{SD^2/n}$, and was used to present the confidence in the data.

Due to the debate in the literature regarding the SEM data calculation (Hopkins, 2000a), the computation of 'typical error' (TE) would be conducted using the spreadsheet made available by Hopkins (2000b). This value would be taken to represent the normal expected error of the measurement tools.

Descriptive data analysis or measures of central tendency and dispersion were also to be employed in the current study in order to present the basic data to the reader and to give an account of the averages, standard deviations and ranges of the data gathered.

In summary, the primary method of data analysis that would be used to compile the results of the initial 3 quantitative experiments would be based on either Bland & Altmans 95% LOA test, Bland & Altmans RC (if there was no systematic bias between the cohorts) and TE.

Data would be presented in a number of LOA scatter plots which can be used to illustrate the spread of the data from the experiments and also to predict where the agreement between future measurements would be likely to fall, i.e. between the LOA lines (95% of the time) (Bland and Altman, 1986).

TE would be used to express minimum expected random error and the confidence intervals for this data would be calculated by a spreadsheet provided by Hopkins (2000b) (Microsoft Excel [Redmond, Washington, USA]).

Standard error (SE) would be calculated using the formula put forth by Bland & Altmans (1986) and would be used to predict the continuity of the sample data. Finally, the data would also be portrayed descriptively using means, SD and the coefficient of variation.

Data calculations for the entire study would be done using 2 software programs, a number of standard Microsoft® Excel 97 SR-1 [Redmond, Washington, USA], spreadsheets were complied or downloaded (Hopkins, 2000b) to collect, sort and to perform some of the basic descriptive data analysis, to calculate the Pearsons product moment correlation coefficient (PPMCC) or even perform some statistical analysis. A specialist statistical computer application, KyPlot©, 2000 [Koichi Yoshioka], was also used to perform all other statistical and descriptive analysis and especially to plot data graphs.

3.2.3 Ethics

A research proposal was initially drafted in the summer of 2000, and amended in winter 2002. Both the initial proposal and the amended proposal were submitted to the 'Joint Research Ethics Committee of the St James's Hospital And Federated Dublin Voluntary Hospitals'. All of the research was to be carried out in the Adelaide and Meath, incorporating the National Children's Hospital (AMNCH).

Approval was granted by this committee prior to undertaking any of the experiments that involved either healthy or patient subjects.

3.3 Discussion.

For osteokinematic measurements to be meaningful the device used to calculate the data must be both valid and reliable (Norkin and White, 2003). The concepts of validity and reliability are not all or none phenomena (Rothstein, 2004). These issues should be discussed in terms agreement rather than correlation coefficients or hypothesis testing (Bland and Altman, 1996; George et al, 2000; Bruton et al, 2000).

This approach to investigations of the validity and reliability of measurement tools has only recently been taken on board in the physiotherapy literature. This has presented a large deficit in practical knowledge regarding our commonly used measurement tools. Clinicians need to know the amount of error they can expect when using a specific measurement tool for a particular joint angle movement.

In addition to inappropriate statistical analyses, deficient methodologies have also been a feature of previous research. It was identified that there was a need to re-investigate the performance of the UG using appropriate statistical analysis and large numbers of subjects which included a patient population for validation studies. The methodology should use blinded instruments yet allow the examiner to read and record the data themselves as this task has previously been implicated a source of measurement error (Williams and Callaghan, 1990).

The UM is a new approach to photographic based osteokinematic measurements. Before the UM can be recommended for use in research or clinical practice, scientific investigation is needed to evaluate its levels of validity and reliability in a similar fashion to the requirements still facing the UG. Additionally, Bland and Altman (1986) hold that it is beneficial to compare a new system of clinical measurement against an older more established method. As the UG is the most commonly used osteokinematic tool in clinical practice (Coxford et al, 1998; Burton et al, 1999), it is appropriate to compare the performance of the UG against the UM. An investigation of this nature would allow a judgement as to which method had the best levels of validity and reliability, and may help to establish the strengths and weaknesses of the methods.

Dunlevy et al (2000) previously carried out limited work on a postural analysis version of the UM and suggested that more work is needed in this area. The investigation should seek to determine whether the UM can improve on current

measurement methods or whether it has anything to add to osteokinematic science in general.

An investigation into this system would have to take account of the 5 different proposed approaches to this systems use (2.3). It would be hoped that this type of experiment would identify one of these approaches as the most optimal way to use the UM at the elbow and knee. However, other authors have identified that the performance of the various measurement techniques change at different joint motions (Fish and Windgate, 1985; Clapper and Wolf, 1988; Bovens et al, 1990; Watkins et al, 1991; Youdas et al, 1993; Armstrong et al, 1998; Nigel-Keenan et al, 2004). Therefore, the same approach to the UM may not be the most appropriate for both the elbow and the knee.

This chapter has identified a number of issues that warrant scientific investigation. A number of specific aims have been identified and categorised, 4 different sets of experiments were drafted that would endeavour to provide scientific data that could be used to address each of the aims. The first set of experiments were to be carried out in a non-clinical environment that essentially aims to establish the sensitivity and specificity of the UM towards the calculation of different angles. The effect of procedural error would also be investigated and used to draft guidelines for the optimal use of the UM.

The second and third sets of experiments aimed to provide data for the validity and reliability of the UM for the measurement of human joint angles, which are arguably the most important aspect of a measurement system. The final set of aims sought to provide data on the nature of usage that the UM got over a three month period in a clinical setting.

3.4 Conclusion.

The dearth of strong evidence regarding the validity and reliability of the universal goniometer (UG) is due to either inappropriate statistical analysis or flawed methodologies. Whatever the reason, there is a clear and urgent need to conduct an investigation that will yield tangible, clinically relevant data regarding the levels of expected error associated with this frequently used tool.

The Uillinn Method© (UM) presents a novel approach to the science of osteokinematic measurement. It is based on previously used photographic techniques. However, the literature also presents a dearth of information regarding the performance of photographic measurement methods. This lack of knowledge presents the need to provide evidence for the validity and reliability for photographic measurement methods.

Furthermore, as the UM is based on new technology an investigation was warranted to gauge its performance towards measurement. It remains unclear how best to apply this system towards the measurement of human joint angles. Five different approaches to its use have been proposed and there was a need to establish which of these systems was the most appropriate. It is necessary to design a study using a sound methodology that could establish the levels validity and reliability for the UG and each of the 5 different approaches to the UM.

Chapter 4

THE ACCURACY OF THE UILLINN METHOD© CALCULATIONS:

A NON-CLINICAL TRIAL.

4.1 Introduction.

The newly developed method of assessing static joint angles, the Uillinn Method© (UM) had taken advantage of new technologies including digital photography and newly engineered software, neither of which had been tested together as a means of calculating angles. The amount of accuracy and error associated with the newly developed system remained unknown.

It was not known whether the digital photograph was of high enough quality to provide an accurate representation of an angle. There were uncertainties regarding the influence of photographic procedures for issues such as parallax error or rotational misalignment.

Placement of the virtual measurement arms (VMA) and the system of angular calculations, were issues that had potential to create substantial error, but were so far untested. This was particularly true for angles larger than 90° which had been a challenge in earlier versions of the UM software.

This set of investigations was conducted in order to examine the question of how well the system was calibrated for the calculation of angular measurements and to understand the how photographic procedures influence measurement error associated with the UM.

This study did not aim to infer towards the total error associated with the clinical environment, but rather to investigate the performance of the system at a more fundamental level. Therefore, in order to allow for a highly controlled setting it was decided that this study should be conducted in a non-clinical environment. This could be justified as the investigation did not have the objective of relating its findings directly to human measurement but rather to help establish a baseline for the potential accuracy of a calculation system.

4.2 Specific Aims of the Study.

1. To determine the potential accuracy of the UM under optimal conditions.
2. To compare error for the UM: tripod mounted versus free-hand photography.
3. To investigate the influence of the pillow effect, lens imperfections and parallax error on the UM.
4. To investigate the influence of camera to target rotation on the UM.
5. To draw up procedural guidelines for the photographic component of the Uillinn Method© based on researched evidence.

4.2.1 Justification of the Aims.

Aim (1): To determine the potential accuracy of the UM under optimal conditions.

The amount of error associated with angular calculations was unknown. Error for the system could arise from the calculation system of the software, or through factors intrinsic to the camera itself, which may cause a misrepresentation of the target angle.

In order to isolate measurement error to these sources tight procedural controls would have to be observed which would ensure the data was not contaminated with error from other sources, such as parallax or photographic blur caused by the unsteady hand of the photographer.

Such a study could then provide information regarding the agreement between a known control data (CD) and the software's mathematical calculations. Indeed if inconsistencies between the CD and the experimental data (the UM's calculations) were found, it may have been possible to use this data to alter the calibration of the software and thereby optimise the system.

The results of this study would form a baseline against which the subsequent experiments, which placed various stresses on the system, could compare their results.

Aim (2): To compare error for the UM: tripod mounted versus free-hand photography.

In addition to potential hazards for the UM that were intrinsic to the camera design, consideration must be given to factors associated with the operation of the camera itself. These factors relate to procedural aspects of the UM which may have a negative effect on measurement accuracy. For example, during the photographic procedure the photographer must hold the camera perfectly steady in order to avoid image distortion or blur. Such a blur artefact would misrepresent the image and thereby lead to measurement error. Often the photographer will use a tripod to overcome this souce of potential blur. It is not clear whether the use of a tripod is essential to maintain the optimum accuracy of the UM.

This issue could be achieved by comparing the error seen during a tripod mounted photographic trial against the error seen for a trial without a tripod.

Aim (3): To investigate the influence of the pillow effect, lens imperfections and parallax error on the UM.

The exact dimensions of the target angle may be incorrectly represented due to a number of specific photography related issues. Parallax error may be introduced as the result of a lateral rather than a perpendicular view point of an object (Koper, 2001) and can be defined as the perceived difference of an image seen from one viewpoint against that seen from a different viewpoint, that lie in the same plane. Misalignment of camera to target angle may introduce parallax error to the recorded image and may present another sources of error to the UM.

It can be said that the fundamental design of a camera is slightly flawed, light rays that are emitted or reflected from an object naturally travel in straight lines. However, in order for a camera to focus these rays on to a photosensitive surface, a convex cameras lens must refract the light; i.e. bend the light rays. This is done in much the same way that the human eye focuses light on the retina, however the human eye is highly evolved to this task and does not share the same limitations of the camera (Sherwood, 1997).

In the case of photography, this distortion of the natural straight path of the light rays results in some amount of distortion of the true dimensions of the image. The distortion results in a phenomenon known as the 'pillow effect' which presents

as a mild bowing at the edges of the photograph (see Figure 1.7, page 80). The light travelling through the centre of the lens does not bend at all and therefore the centre of the photograph should show the truest representation of the image (National Instruments Corporation, 2002).

In this study design it would be impossible to distinguish between error arising from parallax error and error arising from the pillow effect, as both issues would be maximised in an off-centre area of the photograph.

Aim (4): To investigate the influence of camera or target rotation on the UM.

Digital photographs like conventional photographs record only 2-dimensions (2-D) of a real 3-dimensional (3-D) structure. The 2-D image does not portray depth and cannot give an account of rotational aspect of the camera or object at the instant of photography. As a rotational component is added to the photographic process, the angles on the 2-D picture may appear more acute or obtuse than they truly are (Scholz, 1989; Gray et al, 1991). This illusion of a larger or smaller angle may have a substantially negative effect on the accuracy of the UM data. Therefore, a set of experiments was needed to properly analyse this issue.

Aim (5): To draw up procedural guidelines for the photographic component of the Uillinn Method© based on researched evidence.

When the above aims were scientifically analysed, valuable information would be gained regarding the best operational parameters for the photographic component of the UM. This information could be used to avoid unnecessary error when using the system and thereby improve clinical results. It was necessary to compile these findings and draft operational guidelines for the photographic component of the UM. These operational guidelines would be recommended as those that must be followed by examiners who intend to use the UM.

4.3 Methodology.

Four related experiments were designed in order to achieve the studies aims. Each compared the experimental data, i.e. measurements made using the UM, against a control data (CD). Different components of the UM were then stressed during the latter experiments in an attempt to determine how important various procedural factors were to the accuracy of the UM.

4.3.1 Calculation of the Control Data.

The experiments would use 2 datasets; the experimental data measured by the UM and the CD. It was necessary to formulate a highly accurate method of calculating 2 dimensional angles that would act as the criterion, gold-standard method and present the CD.

The subject matter for the study would be a series of triangles of various dimensions. These triangles would be depicted on a whiteboard with 3 small arbitrarily placed dots, the dots were made with a non-permanent marker and were approximately 2mm in diameter. These dots depicted the 3 corners of a triangle.

Each dot was labelled either A, B or C. These 3 uppercase letters represented an angle, e.g. angle 'A' represented the angle made by the lines 'c' and 'b' (see Figure 4.1) and was opposite to the line 'a'. The length of each line was random, and was constructed by the examiner, who subjectively attempted to create a wide range of angles using different triangular dimensions. It was decided that the strongest model for calculation of the CD would be done using the mathematical model of the Cosine rule.

This method would require the exact linear measurement of each of the lines in the triangle. These measurements would be calculated using a Vernier callipers [RS Components Limited]. The callipers used were capable of measuring from 0 mm to 131 mm and are accurate to 0.02 mm. All of the linear measurements were made using this device.

In practice reading this device was difficult and presented possible error from misreading the scale, to control this potential source of error all linear measurements were made twice. The values of these lines were used to assess each of the angles in the triangle using the Cosine rule.

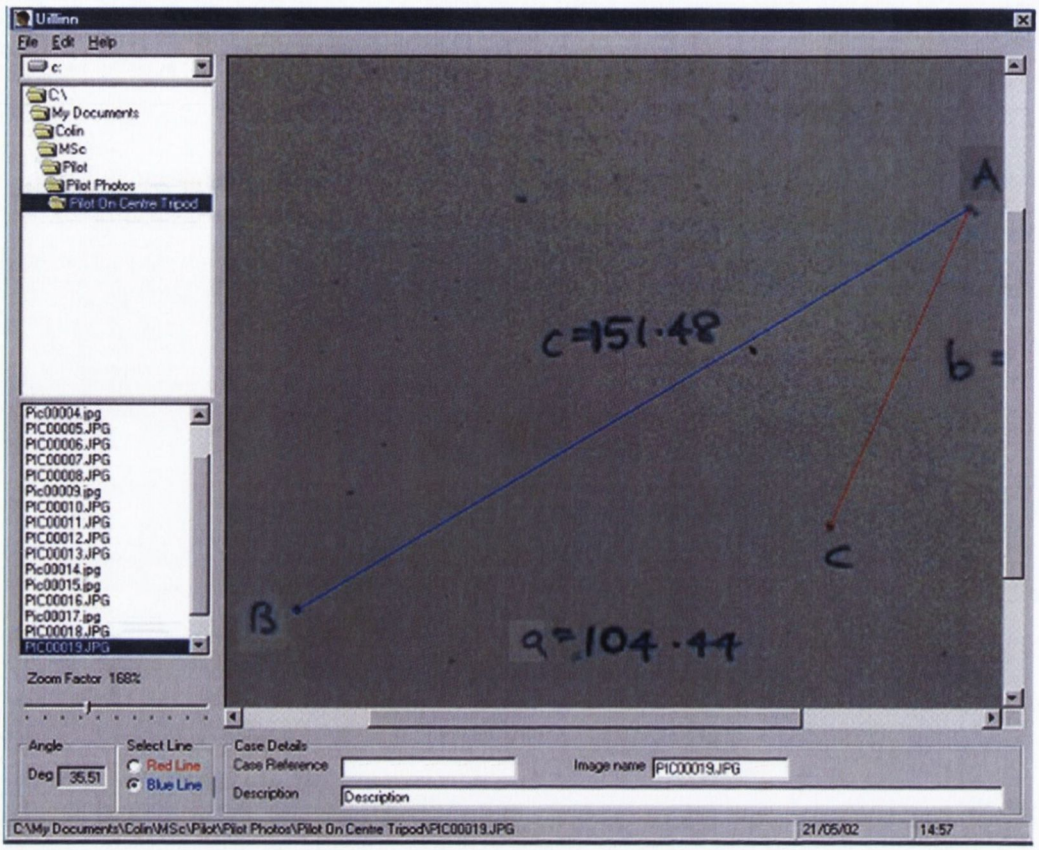


Figure 4.1. The Uillinn Method© used to assess angle 'A' from one of the triangles.

The Cosine rule states that; $a^2 = b^2 + c^2 - 2 \times b \times c \times \cos(A)$.

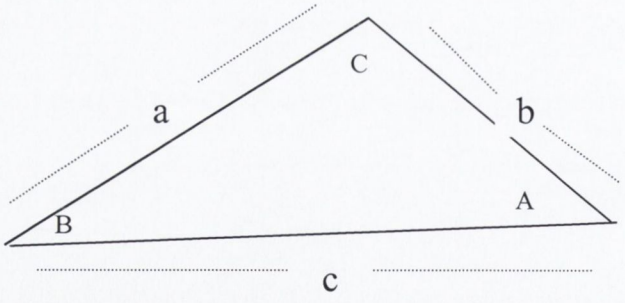


Figure 4.2. Illustration of the Cosine Rule.

To remove human error and hasten the calculation process, a spreadsheet was designed that calculated the results automatically. The Cosine rule formula of '$a^2 = b^2 + c^2 - 2 \times b \times c \times COS(A)$' (see Figure 4.2) was programmed into a Microsoft® Excel 97 SR-1, spreadsheet [Redmond, Washington, USA]. The values

of each of the lines were inputted into this sheet and the final angles were presented automatically.

One photograph per triangle was taken. Each photograph was used to calculate 3 angles (A, B & C). The experiments produced 2 sets of data; the CD and the experimental, UM data. The calculation method for the CD was the same throughout the different experiments, i.e. using the Cosine rule. The UM data was assessed by only one examiner using the UM software over a number of sittings.

4.3.2 Data Analysis.

Data analysis included Bland and Altman 95% limits of agreement (LOA) to assess random and non-random error. A limit of 2.5% of mean measured data was set as the limit for detection of SB (Batterham and George, 2000). Because the subject matter for the investigations was to be all three angles in a triangle, this would imply the average CD figure would be 60° for all of the experiments. Therefore, the limits of SB should be set at 2.5% of 60°, i.e. 1.5°. When SB was rejected 95% agreement limits were expressed through the repeatability coefficient (RC).

Typical error (TE) would be calculated to give an account of the minimum expected random error in the data. A calculation of standard error (SE) would be used to check the predicted trend of the data sets, i.e. if the data was gathered continuously how much further it would be likely to vary. If this figure were low it would strengthen the inferences posed by the 95%LOA, RC or TE.

Descriptive statistics would be used to allow the reader to access the data at a basic level. Appropriate graphs and scatter plots would be used to clearly illustrate the result data.

4.3.3 Equipment

The equipment used to assess the angles with the UM was an Olympus D-340L 1.3 Megapixel digital camera [Olympus Optical Co., London, UK], a personal computer (PC) with a 233 MHz processor [Intel Corporation, Hillsboro, OR, USA] and the UM software installed onto it. A SmartMedia card drive manufactured by SanDisk [Sunnyvale, CA, USA.] was used to transfer the data from the digital

camera to the PC (see Figure 4.3). During photography the camera was kept on a 'High Quality' resolution setting.

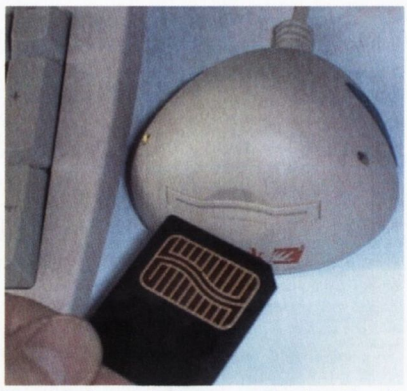


Figure 4.3. The Smart Media Disk and Smart Media Disk Drive.

4.4. The Experiments & Results

4.4.1. Experiment 1: To determine the potential accuracy of the UM under optimal conditions.

The camera was mounted on a tripod and placed 1m away, perpendicular to the whiteboard. This photographic procedure would omit image distortion due to the pillow effect, parallax error and rotation between the camera and the triangle. The viewfinder crosshairs of the camera were on the approximate midpoint of each of the drawn triangles. It was hoped that the tripod would remove the possibility of shake occurring during lens exposure.

A total of 20 triangles, of random dimensions were created and the lengths of their sides were measured as per the procedure described above. Each of the triangles were numbered, their individual angles labelled as either A, B or C and the length of each side of the triangle was also written on the whiteboard and recorded on a blank spreadsheet.

When this was done a digital photograph was taken and the image then transferred to the PC. Each photograph contained 3 angles; A, B and C, which resulted in a total of 60 different angles from the 20 triangles. All of these angles were assessed by the UM, version 2.2, and were recorded as the UM data (see Table 4.3). The examiner who assessed each of the 60 angles was blind to the CD, which had not yet been calculated.

4.4.2 Results: Experiment 1.

This experiment aimed to investigate the potential accuracy of the UM under optimal conditions. The average difference between the UM and CD was low at 0.07° (±0.46°), within a range of between –0.86° to +0.95°. The PPMCC value was high (r=0.99) and a graphed slope showed high a correlation (see Graph 4.3).

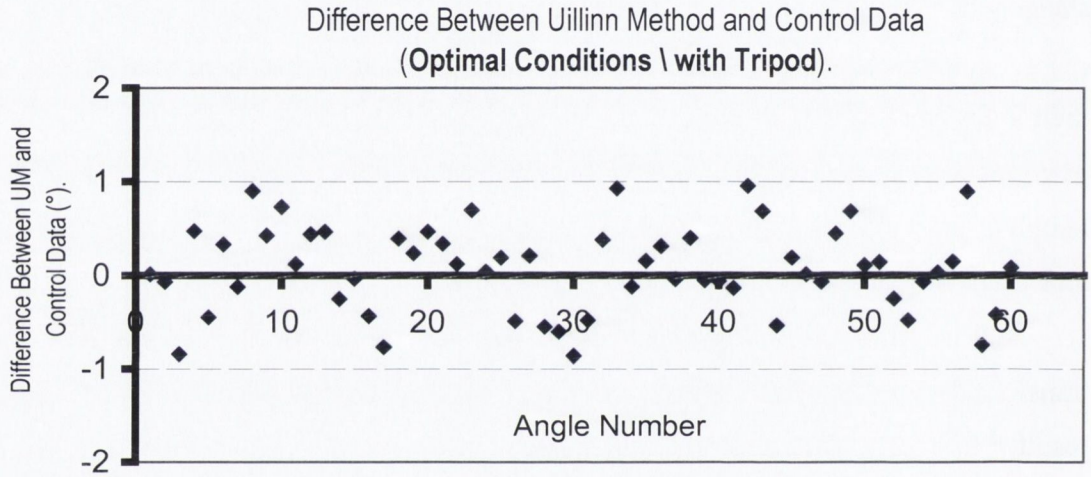
Graph 4.4. illustrates an even spread of data, indicating homoscedasticity or normality in the spread of the data, which implied the Bland and Altman 95% LOA and RC tests were appropriate.

The mean difference of 0.067° was well below the limit of SB (1.5°) and therefore indicated that no SB. The RC was 0.89° and the corresponding 95% LOA

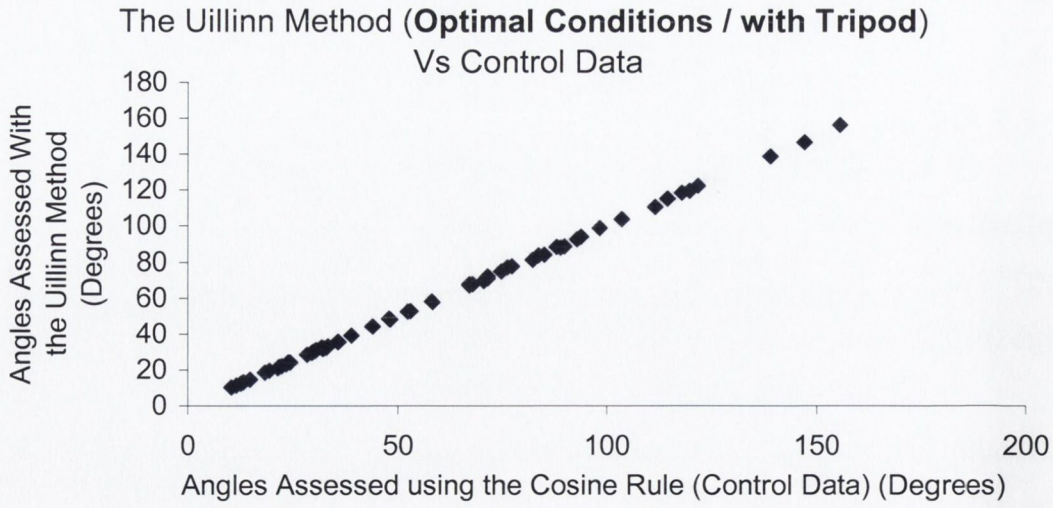
was +0.91°, -0.88°. The TE was calculated as 0.32° with a lower confidence limit of 0.28° and an upper confidence limit of 0.38°.

The results show that only a quarter of the 60 angular data of the CD and UM were different above 0.5° (see Table 4.3) and that agreement would consistently be less than 1° (see Table 4.3 / Graph 4.2).

The standard error (SE) for the differences between the CD and the UM was low at 0.06°, indicating that the data would not be likely to vary further than that amount (Bland and Altman, 1986).



Graph 4.2.



Graph 4.3.

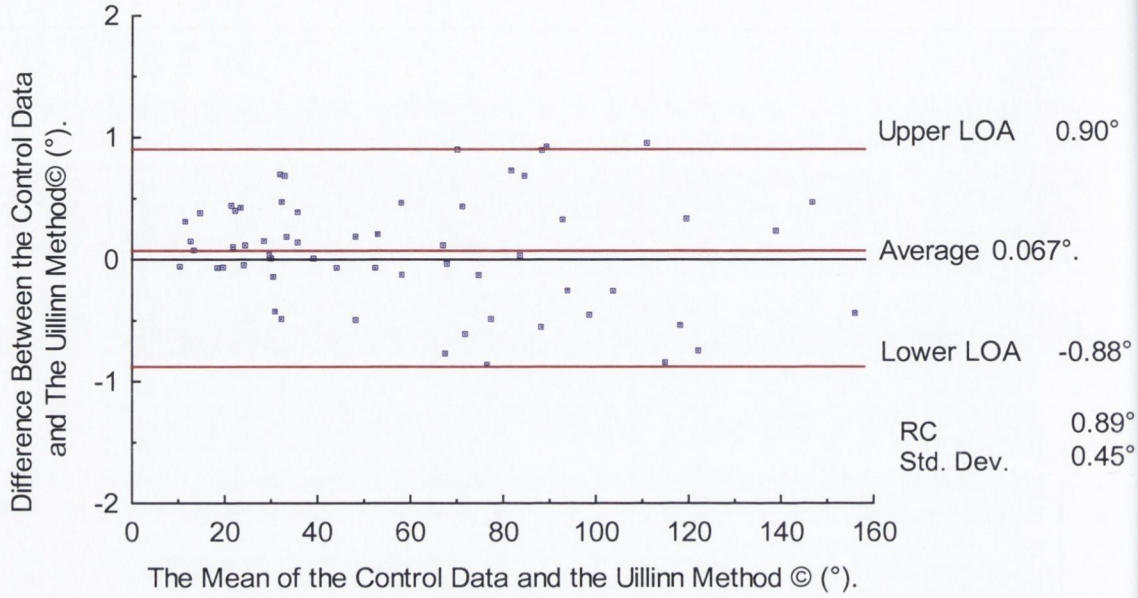
Tripod Mounted Photography

	Length of Sides (mm)			Cosine Rule (°) Control Data (CD)			Uillinn Method (UM) (°)			Difference Between CD and UM (°)		
Trial	Side a	Side b	Side c	Ang A	Ang B	Ang C	Ang A	Ang B	Ang C	Ang A	Ang B	Ang C
1	61.78	107.5	62.58	29.97	119.6	30.4	29.96	119.3	30.54	0.009	0.333	-0.14
2	82.28	49.04	110.1	44.06	24.49	111.5	44.13	24.37	110.5	-0.07	0.117	0.953
3	132.1	77.52	79.58	114.5	32.29	33.26	115.3	31.59	32.57	-0.84	0.698	0.685
4	61.74	56.72	101.5	32.53	29.61	117.9	32.06	29.57	118.4	0.471	0.036	-0.54
5	97.18	73.24	54.12	98.35	48.22	33.44	98.8	48.03	33.25	-0.45	0.187	0.186
6	112.4	83.54	71.04	92.94	47.92	39.14	92.61	48.42	39.13	0.327	-0.5	0.009
7	113.9	94.3	93.48	74.68	52.99	52.33	74.81	52.78	52.4	-0.13	0.207	-0.07
8	122.5	129.8	47.82	70.55	87.85	21.6	69.65	88.4	21.16	0.901	-0.55	0.442
9	47.34	112	117.7	23.62	71.49	84.88	23.2	72.1	84.2	0.425	-0.61	0.684
10	132.5	129.8	49.72	82.1	76.08	21.82	81.37	76.94	21.72	0.73	-0.86	0.103
11	82.76	87.58	52.5	67.11	77.13	35.76	66.99	77.62	35.62	0.117	-0.49	0.14
12	112.6	30.42	118.5	71.42	14.84	93.74	70.98	14.46	94	0.435	0.38	-0.26
13	101.6	119.6	63.38	58.22	89.76	32.01	57.76	88.84	32.5	0.464	0.924	-0.49
14	109.9	95.98	35.66	103.5	58.08	18.38	103.8	58.2	18.45	-0.26	-0.12	-0.07
15	93.3	48.08	100.1	67.87	28.51	83.62	67.9	28.36	83.59	-0.03	0.152	0.033
16	132.5	64.86	70.66	155.7	11.64	12.7	156.1	11.33	12.55	-0.44	0.311	0.149
17	121.3	53.76	131.7	67.11	24.09	88.79	67.88	24.14	87.9	-0.77	-0.05	0.895
18	104.4	67.84	151.5	35.94	22.41	121.7	35.55	22.01	122.4	0.387	0.4	-0.75
19	134	36.62	104.2	139	10.32	30.64	138.8	10.38	31.07	0.234	-0.06	-0.43
20	116.9	72	49.7	147.1	19.57	13.37	146.6	19.63	13.29	0.468	-0.06	0.076

Differences Between CD and UM	
Mean (systematic bias)	0.067°
SD	0.456°
SE of Difference ($\sqrt{SD^2/n}$)	0.059°
Typical Error ($SD/\sqrt{2}$)	0.32°
Repeatability Coefficient	0.89°
Coef. Var.	6.83%
Range of Difference	-0.86° to 1°

Table 4.3. The Accuracy of the UM for Angular Measurements (Optimal Conditions) (n=60).

Bland & Altmans 95% Limits of Agreement Data Plot.
The Uillinn Method© (Optimal Conditions / With Tripod) Vs. The Control Data.
(r=0.99 / n=60)



Graph 4.4

.

4.4.3 Experiment 2: To compare error for the UM when tripod mounted versus free-hand photography is used.

This experiment was very similar to experiment 1, the only difference was that a tripod was not used during the photography of the triangles, instead, the photographs were taken free-hand. Twenty randomly constructed triangles were depicted on the whiteboard, their sides were measured using the callipers and labelled as was done previously. The lens of the camera was held so that it's was approximately parallel to the whiteboard and the viewfinder crosshairs centred in the middle of the triangle. The photographer stood 1m away from the whiteboard.

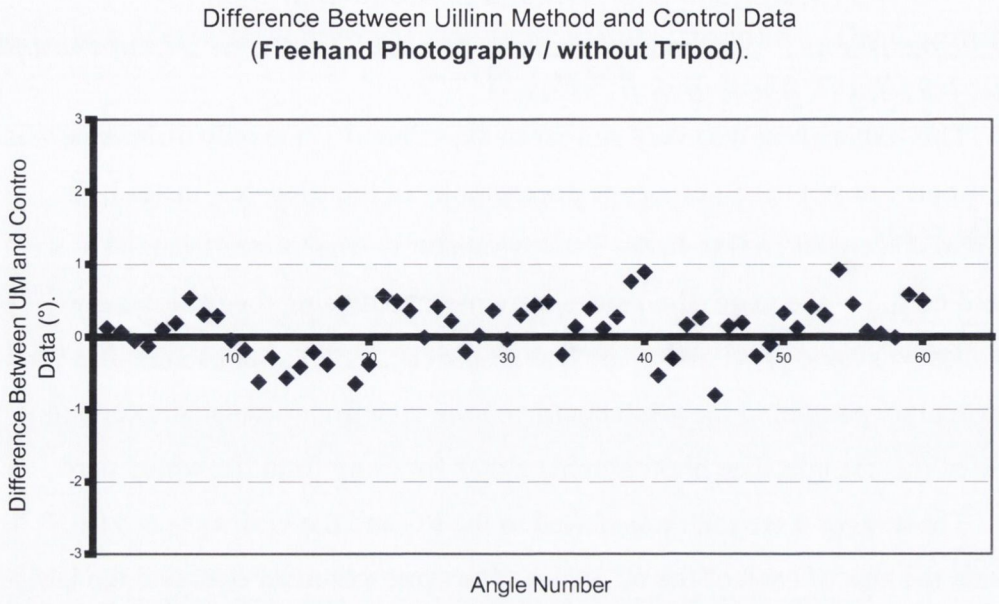
The images were then transferred to the PC and the UM was used to calculate the size of each of the 60 angles. The same examiner operated the UM software during the measurement process. Again, the Cosine rule was also used to calculate the control data, but only after all the UM data was calculated.

4.4.4 Results: Experiment 2.

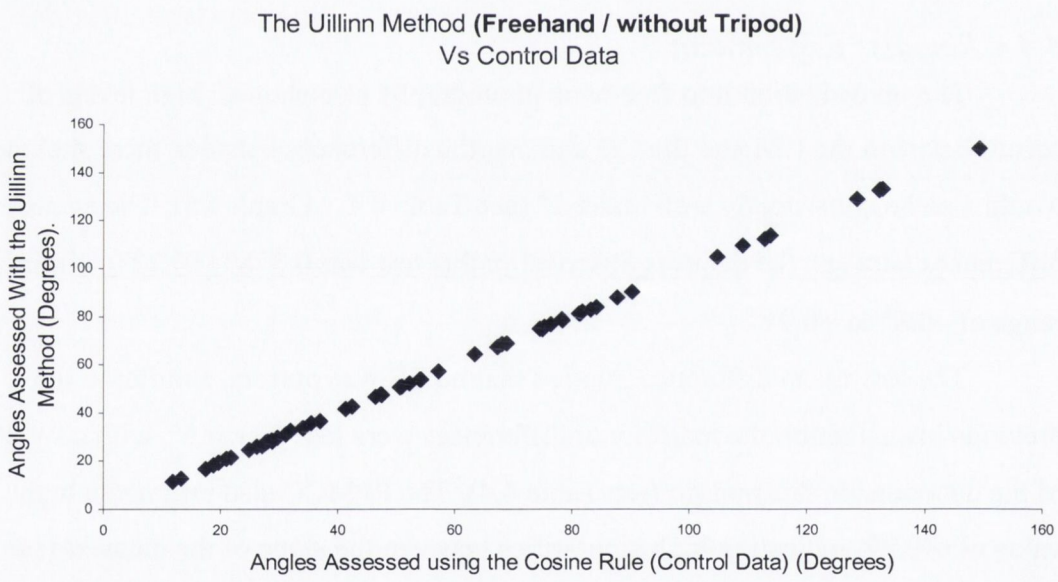
The investigation into free-hand photography also showed high levels of accord between the UM and the CD and that the differences between these methods would also be consistently well under 1° (see Table 4.4. / Graph 4.5). The average differences between the datasets observed in this test was 0.378° (±0.22°) within a range of –0.8° to +0.91°.

The low mean difference implied that no SB was present. Similar to the previous investigation the majority of differences were less than 0.5°, with only 20% of the data outside this margin (see Table 4.4). The PPMCC also returned a high value of r=0.99, indicating high similarities between the slope of the datasets (see Graph 4.6).

The data was homoscedastic when graphed (see Graph 4.5) and the RC was calculated as 0.741°. The TE was calculated as 0.27° with a lower confidence limit of 0.23° and an upper confidence limit of 0.31°. The SE in the data was calculated as 0.048°, which indicated that the data was reflective and unlikely to change.



Graph 4.5.



Graph 4.6.

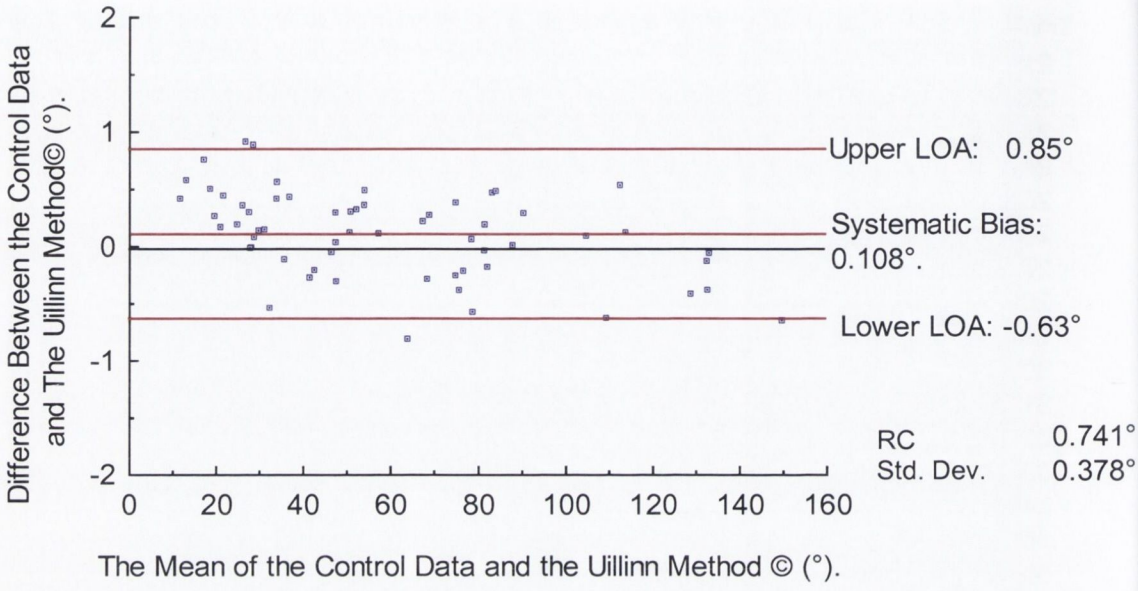
Freehand Photography

	Length of Sides (mm)			Control Data (CD) (°)			Uillinn Method (UM) (°)			Difference Between CD and UM (°)		
Trial	Side a	Side b	Side c	Ang A	Ang B	Ang C	Ang A	Ang B	Ang C	Ang A	Ang B	Ang C
21	88.66	54.46	51.34	113.8	34.19	31.99	113.7	33.62	32.52	0.126	0.568	-0.53
22	90.58	75.06	68	78.4	54.26	47.34	78.33	53.77	47.64	0.066	0.495	-0.3
23	110.5	66.44	53.98	132.8	26.16	20.99	132.9	25.8	20.82	-0.05	0.363	0.172
24	109.6	69.64	49.84	132.3	28.06	19.67	132.4	28.06	19.4	-0.12	-0	0.27
25	87	18.47	80.46	104.7	11.85	63.45	104.6	11.43	64.26	0.097	0.42	-0.81
26	100.4	93.68	52.5	81.46	67.38	31.15	81.27	67.16	31	0.194	0.223	0.153
27	89.58	65.54	40.78	112.7	42.44	24.83	112.2	42.64	24.63	0.538	-0.2	0.196
28	70.84	91.14	60.2	50.96	87.74	41.3	50.65	87.73	41.57	0.306	0.013	-0.27
29	87.84	71.16	51.06	90.35	54.11	35.54	90.06	53.74	35.65	0.295	0.365	-0.11
30	64.14	87.52	70.1	46.41	81.25	52.34	46.46	81.28	52.01	-0.05	-0.03	0.329
31	92.74	69.08	72.42	81.86	47.51	50.63	82.04	47.21	50.5	-0.18	0.299	0.127
32	87.86	55.92	51.98	109	37	34.02	109.6	36.57	33.6	-0.62	0.434	0.419
33	81.5	87.36	40.74	68.15	84.21	27.64	68.43	83.72	27.34	-0.28	0.487	0.303
34	92.38	90.92	43	78.33	74.55	27.12	78.9	74.8	26.2	-0.57	-0.25	0.92
35	102.2	64.84	48.4	128.5	29.76	21.75	128.9	29.62	21.3	-0.41	0.142	0.449
36	85.92	85.36	42.46	76.38	74.92	28.7	76.59	74.53	28.62	-0.21	0.386	0.084
37	66.76	58.06	50.84	75.30	57.27	47.44	75.67	57.15	47.4	-0.38	0.116	0.041
38	82.72	77.72	38.82	83.29	68.93	27.78	82.82	68.65	27.79	0.473	0.277	-0.01
39	100.5	59	45.19	149.3	17.46	13.28	149.9	16.7	12.7	-0.64	0.76	0.583
40	92.34	60.34	40.22	132.3	28.89	18.79	132.7	28	18.28	-0.38	0.89	0.506

Differences Between CD and UM	
Mean (systematic bias)	0.107°
S.D.	0.376°
SE of Difference ($\sqrt{SD^2/n}$)	0.048°
Typical Error ($SD/\sqrt{2}$)	0.27°
Repeatability Coefficient	0.736°
Coef. Var.	3.48%
Range of Difference	-0.81° to 0.92°

Table 4.4. The Control Data and the Uillinn Method© (Free-hand Photography, n=60).

Bland & Altmans 95% Limits of Agreement Data Plot.
The Uillinn Method© (Free Hand Photography) Vs. The Control Data.
(r=0.99 / n=60)



Graph 4.7.

4.4.5 Comparing Free-Hand and Tripod Mounted Photography

When comparisons were made between this experiment (Free-Hand Photography) and experiment 1 (Tripod Mounted Photography) an unpaired t-test revealed no significant differences ($p=0.05$) for the difference between the CD and the UM datasets (see Appendix A).

The 95% LOA data plots showed similar spreads between the two experiments (see Graph 4.4 & 4.7) and the TE measures were of a similar magnitude (see Table 4.5). These findings indicated that there were no actual difference in the performance for the UM when the photographic stage was controlled with a tripod or the camera was held free-hand. The minute differences that were seen in the data, for example, the additional 0.04° of difference in the means (see Table 4.5) were attributed to random error.

Freehand Vs Tripod Mounted Photography		
	Free-Hand	Tripod Mounted
Mean (systematic bias)	0.108°	0.067°
S.D.	0.378°	0.458°
SE of Difference ($\sqrt{SD^2/n}$)	0.048°	0.059°
Typical Error ($SD/\sqrt{2}$)	0.27°	0.32°
Repeatability Coefficient	0.741°	0.916°
Coef. Var. (100xSD/mean)	3.48%	6.83%
Range of Difference	-0.81° to 0.92°	-0.86° to 0.953°

Differences Between CD and UM

Table 4.5. Comparison of Freehand Vs Tripod Mounted Photography.

As there was no difference in the performance of the UM due to the use of a free-hand photographic procedure, data from experiment 1 (tripod) and experiment 2 (free-hand) were coupled in order to achieve a larger degrees of freedom ($n=120$) for the statement into the expected error for this system's angle calculations (see Table 4.6).

The results from this re-calculation revealed the mean difference between the CD and the UM data was 0.09° (±0.41°), which was again well below the limit of SB of 1.5° and gave a strong indication that SB was not present.

The PPMCC was calculated at $r=0.99$ which showed a strong correlation between the cohorts. Graph 4.8 shows that the spread of the data was homoscedastic. The RC between the CD and the UM was calculated as 0.81°, which indicated that

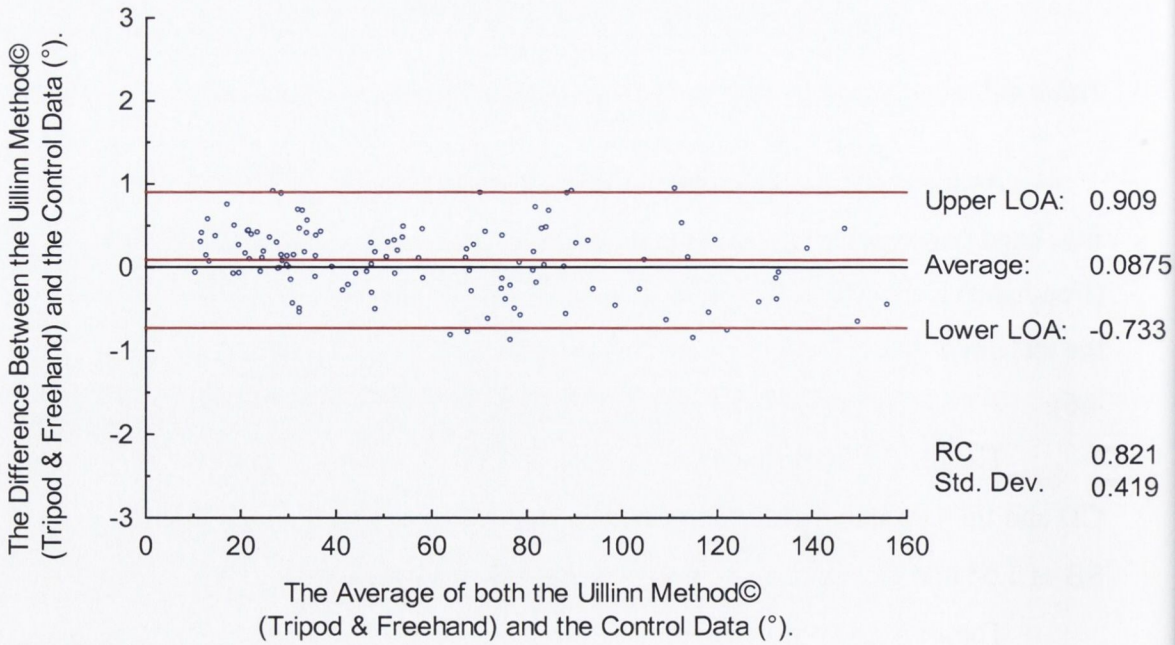
for 95% of repeated measurements calculations made by the UM would agree to within 0.81°. Typical error (TE) between the datasets was 0.29° with a lower confidence limit of 0.29° and an upper confidence limit of 0.33°. The SE for the data of the difference between the CD and UM data was 0.038°.

These statistics could basically be interpreted as showing the expected difference between angles calculated by the Cosine rule and the same angles calculated by the UM would agree to within 0.81° for 95% of the measurements, but typically the difference between the two measures would be only 0.29°.

Differences Between CD and UM (n=120)	
Mean (systematic bias)	0.088°
SD	0.416°
SE of Difference ($\sqrt{SD^2/n}$)	0.038°
Typical Error ($SD/\sqrt{2}$)	0.29°
Repeatability Coefficient	0.816°
Range of Difference	-0.86° to 1°

Table 4.6. The Accuracy of the UM for Angular Measurements (n=120).

Bland and Altmans 95% Limits of Agreement Data Plot.
The Uillinn Method© (Tripod & Freehand) and the Control Data.
(r=0.99 / n= 120).



Graph 4.8.

4.4.6. Experiment 3: To investigate the influence of the pillow effect, lens imperfections and parallax error on the UM.

It was felt that if the UM could show the same level of accuracy for angles calculated in any part of the cameras field of view there would be no basis to suppose that the area of the photograph in which the angles appeared should be a considered a source of error for the UM, from either parallax or rotational perspective errors.

The methodology employed would make it difficult to logically differentiation between error which was due to the pillow effect, parallax error or even other lens imperfections. Therefore, all 3 sources of error were examined using the same experiment. The simple objective was to determine whether differences increased error between the CD and the UM data across the different areas of the photograph.

The experiment divided the white board with 6 vertical and 6 horizontal lines, which produced a grid containing 25 boxes. Each box was of the dimension 10 cm x 8 cm. Three random dots were placed in each of the 25 boxes resulting in a triangle of different dimensions in each of the boxes (see Figure 4.5). The length of each side of the 25 triangles was measured using the Vernier callipers and this data was again used to calculate CD using the Cosine rule as described previously.

The whiteboard was photographed with the camera mounted on the tripod, which was exactly 63 cm away from the centre of the grid. This was the distance that totally filled the photograph with the grid on the whiteboard (see Figure 4.5). A total of 72 different angles were measured in the photograph, 48 different angles in the outer 16 boxes of the camera's field of view and 24 different angles in the 8 boxes in the camera's off-centre field of view (see Figure 4.4). The triangle in the centre of the camera's field of view was ignored as this single triangle only yield 3 angles which was too small to have analytical value as a control.

The same examiner used the UM to assess each of the 72 angles in the photograph, this data set was recorded as the UM data. The datasets were divided into UM and CD for both the 'Periphery' and the 'Off-centre' data.

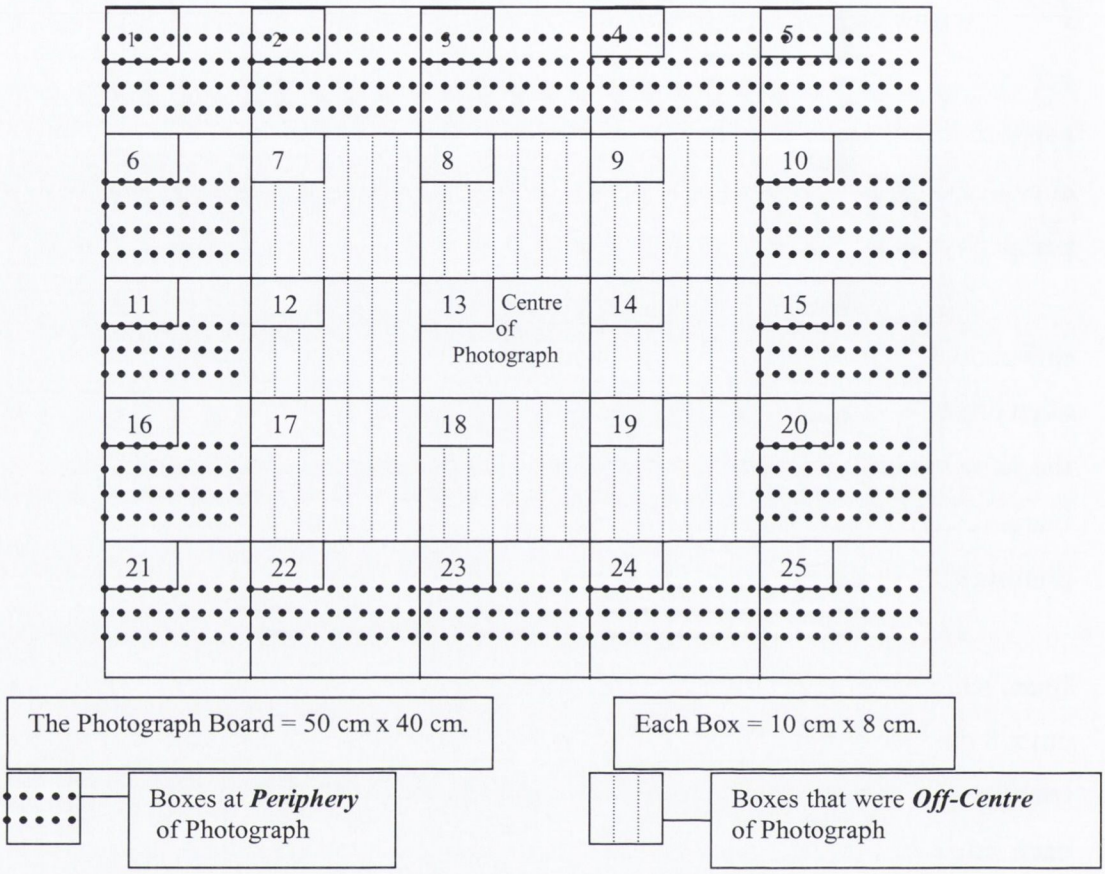


Figure 4.4. A description of the lay out of each photograph.

4.4.7. Results: Experiment 3.

Sixteen boxes were at the edge of the photograph and were dealt with as the 'Periphery' data, there were 3 angles in each of these boxes which resulted in a total of 48 cohorts in this set.

The 8 boxes that were located within the ring of peripheral boxes, closer to the centre of the photograph were labelled the 'Off-Centre' angles (see Figure 4.4). There were also 3 angles in each of these boxes which yielded a total of 24 angles being assessed from this region of the image.

Comparison between the CD and the data calculated by the UM showed that the PPMCC remained high for both regions (Off-Centre, r=0.99; Periphery, r=0.99) (see Graph 4.9; 4.10). The average difference between the UM and the CD for the Off-Centre data was 0.1° (± 0.73°) while the Periphery data had a mean difference

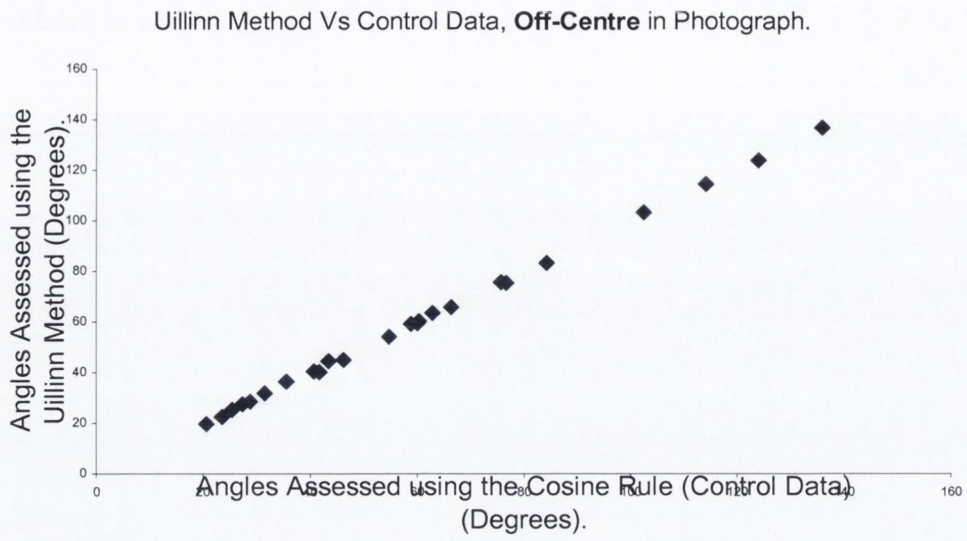
was 0.05° (± 1.02°). The limits of SB were again set 1.5° and inspection of the data implied no SB.



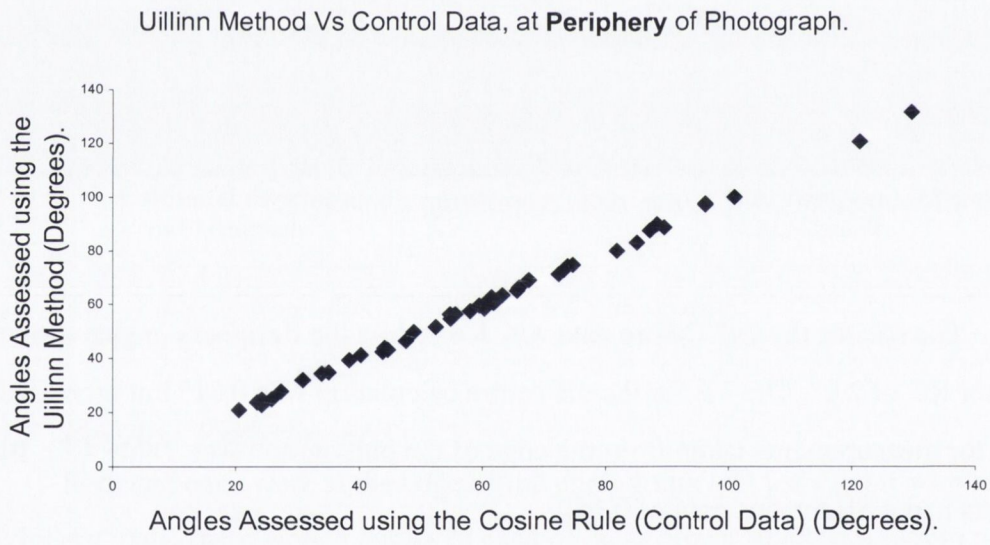
Figure 4.5. A grid of 25 squares, each containing a triangle.

The RC for the Off-Centre data was 1.4° while the Periphery angles showed a higher RC of 2.0°. The TE for the off-centre calculation was 0.51° but increased to 0.73° for measurements taken from the edge of the photograph (see Table 4.8). Both datasets had a SE for this data of 0.15°.

Inspection of Graph 4.13 (page 156) reveals a smooth uniform rise in TE and RC from a well centred photograph (experiment 2 & 3) to measurements of angles situated further from the photographs centre.



Graph 4.9.



Graph 4.10.

Angles Off Centre in the Photograph

Box	Length of Sides (mm)			Cosine Rule (°) Control Data (CD)			Uillinn Method (UM) (°)			Deviation Between CD and UM (°)		
	Side a	Side b	Side c	Ang A	Ang B	Ang C	Ang A	Ang B	Ang C	Angle Dif A	Angle Dif B	Angle Dif C
7	57.2	41.46	77.5	46.133	31.504	102.36	45.03	31.95	103.2	1.103	-0.44	-0.84
8	86.56	49.54	43.84	135.85	23.495	20.659	136.7	22.52	19.81	-0.85	0.975	0.849
9	52.2	74	65.92	43.331	76.605	60.063	44.49	75.41	59.4	-1.16	1.195	0.663
12	48.84	88.2	51.08	27.353	123.93	28.721	27.54	123.8	28.7	-0.18	0.13	0.021
14	59	55.12	52.56	66.399	58.881	54.72	65.71	59.21	54.3	0.689	-0.33	0.42
17	60.2	40.3	68.96	60.28	35.547	84.173	60.3	36.37	83.2	-0.02	-0.82	0.973
18	58.62	82.26	38.6	40.619	114	25.384	40.61	114.5	25.2	0.009	-0.5	0.184
19	86.28	59.08	79.22	75.637	41.555	62.807	75.77	40.31	63.47	-0.13	1.245	-0.663

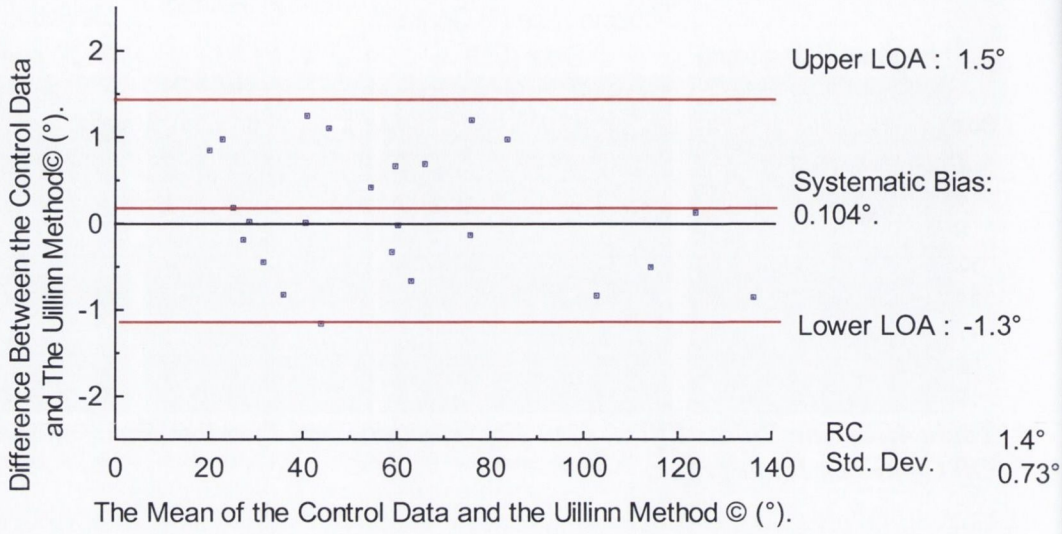
Table 4.7. - The Pillow Effect, Lens Imperfections and Parallax Error on the UM; Off Centre on Photograph.

Angles at the Periphery of the photograph

Box	Length of Sides (mm)			Cosine Rule (°) Control Data (CD)			Uillinn Method (UM) (°)			Deviation Between CD and UM (°)		
	Side a	Side b	Side c	Ang A	Ang B	Ang C	Ang A	Ang B	Ang C	Angle Dif A	Angle Dif B	Angle Dif C
1	35.42	63.98	68.54	30.76	67.49	81.75	32.06	69.1	79.93	-1.3	-1.61	1.821
2	35.68	74.88	86.38	24.24	59.51	96.25	22.87	59.52	97.12	1.373	-0.01	-0.87
3	80.34	59.34	55.92	88.33	47.59	44.09	89.82	46.43	43.76	-1.49	1.156	0.327
4	87.92	49.64	47.48	129.7	25.74	24.55	131.4	24.95	23.7	-1.69	0.795	0.849
5	37.52	66.42	91.68	20.48	38.27	121.3	21.05	39.32	120.4	-0.57	-1.05	0.853
6	64.6	72.88	57.36	58.02	73.12	48.86	57.51	72.72	49.71	0.506	0.4	-0.85
10	87.98	88.68	51.38	72.35	73.84	33.81	71.11	74.23	34.4	1.236	-0.39	-0.59
11	72.96	60.22	41.78	89.44	55.62	34.93	88.69	56.32	34.84	0.754	-0.7	0.093
15	46.48	57.78	63.82	44.6	60.79	74.6	43.09	61.35	74.73	1.511	-0.56	-0.13
16	41.12	52.06	63.52	40.16	54.74	85.09	41.33	56.34	82.94	-1.17	-1.6	2.152
20	83.06	75.7	38.1	87.19	65.55	27.27	87.4	65.1	27.8	-0.21	0.446	-0.53
21	60.36	67.16	69.54	52.37	61.79	65.84	51.66	60.46	66.13	0.71	1.327	-0.29
22	65.72	59.9	47.2	74.65	61.51	43.83	74.65	62.62	42.53	0.001	-1.11	1.305
23	55.02	54.36	50.48	63.18	61.85	54.97	63.5	61.2	54.96	-0.32	0.653	0.005
24	89.98	78.62	36.1	96.19	60.3	23.51	97.56	58.35	23.83	-1.37	1.953	-0.32
25	93.78	78.08	39.16	101	54.82	24.2	99.83	54.57	24.7	1.15	0.25	-0.5

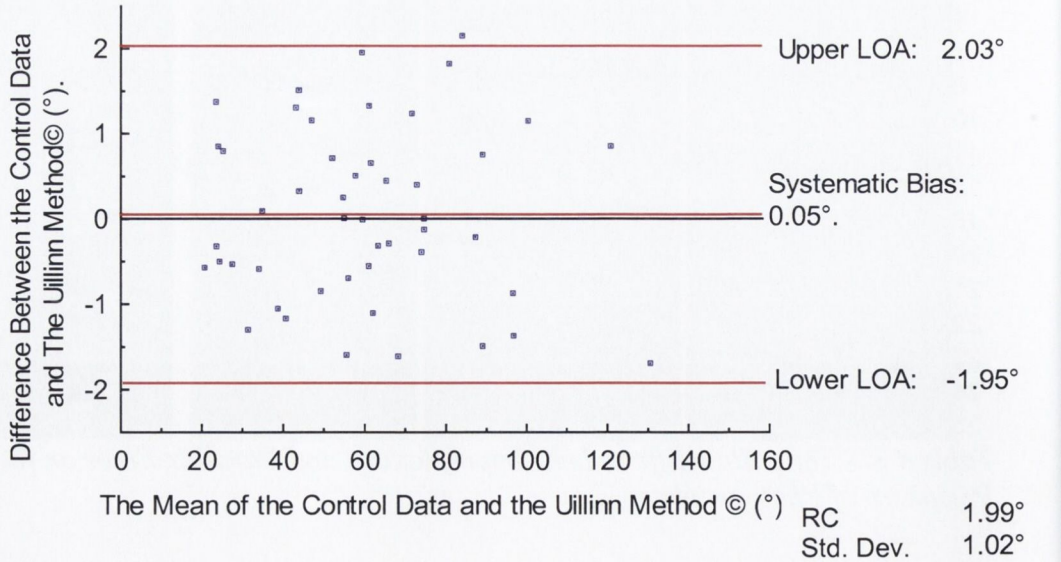
Table 4.8. - The Pillow Effect, Lens Imperfections and Parallax Error on the UM; Periphery of Photograph.

Bland & Altmans 95% Limits of Agreement Data Plot.
The Uillinn Method© Vs. The Control Data (Angles **Off Centre** in Photograph)
(n=16 / r=0.99).



Graph 4.11.

Bland & Altmans 95% Limits of Agreement Data Plot.
The Uillinn Method© Vs. The Control Data (Angles at **Periphery** of Photograph).
(n=48 / *r*=0.99).

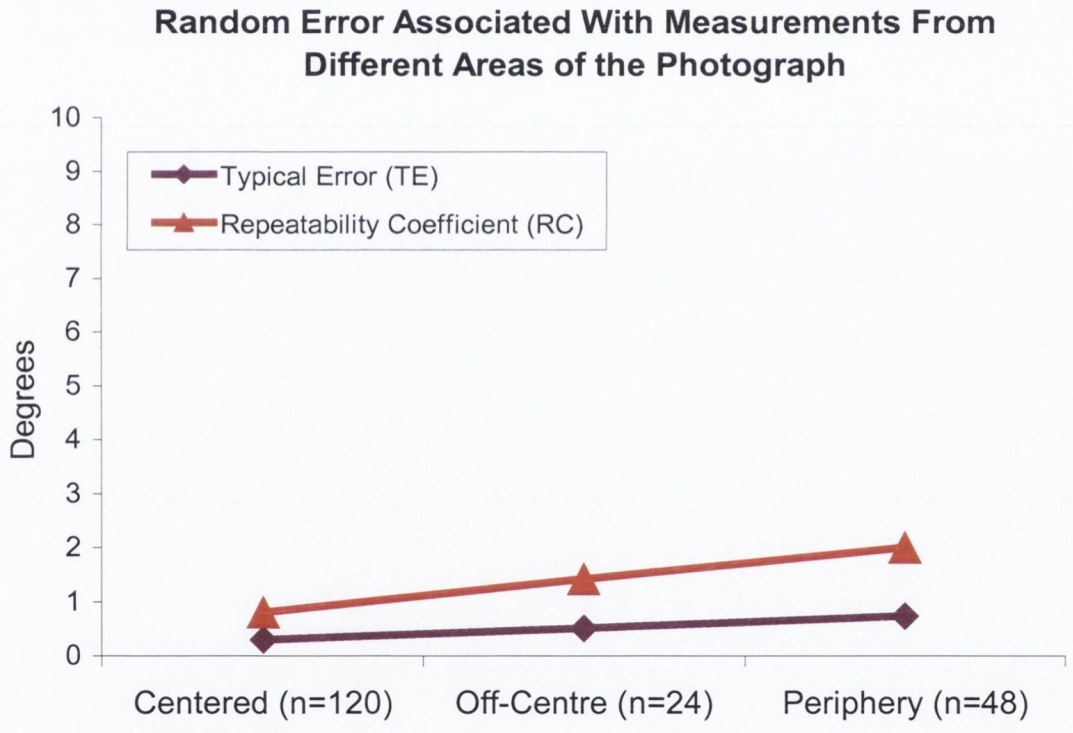


Graph 4.12.

Differences Between CD and UM			
	Centred (n=120)	*Off-Centre (n=24)*	*Periphery (n=48)*
Mean (systematic bias)	0.09°	0.104°	0.05°
S.D. (Differences)	0.41°	0.73°	1.02°
SE of Difference ($\sqrt{SD^2/n}$)	0.038°	0.148°	0.147°
Typical Error (SD/ $\sqrt{2}$)	0.29°	0.51°	0.73°
Repeatability Coefficient	0.81°	1.42°	2°
Range of Difference	-0.81° to 0.95°	-1.16° to 1.24°	-1.69° to 2.15°

Differences Between CD and UM

Table 4.9. Comparison of Variables for Measurements in Different Areas of the Photograph.



Graph 4.13. Uniform Increase in Observed Error as Measured Angles Move away from the Centre of the Photograph.

4.4.8. Experiment 4: To investigate the influence of camera or target rotation on the UM.

In order to determine the effect that a rotational component had on the UM, a trial was designed that subjected a small mounted whiteboard (25 cm x 28 cm) to incremental amounts of rotation during photography. Similar to the previous experiments, the whiteboard had individual non-specific triangles depicted on it, these were used to calculate the CD.

In order to accurately calibrate the amount of rotation that was to be introduced into the photographic procedure the camera and whiteboard position were closely managed. The floor was marked with two 50 cm strips of white tape, they were placed exactly parallel to each other and 1m apart. The parallel relationship was ensured by assessing the 1m distance between them at 10 different points on the lines. One line was called the 'camera line' and the other was labelled the 'principle line'. The camera was mounted directly over the centre of the camera line with its lens parallel to and facing the principle line.

At the centre point of the principle line which was exactly perpendicular to the position of the camera, 5 straight lines were marked. The first of these lines was at 5° rotation to the principle line, the next was at 10° and the subsequent 3 lines were at increments of 10° to the next (see Figure 4.6). The position that each of these lines was calibrated and drawn using the protractor. The mounted whiteboard was placed on the principle line directly over the point where the rotational lines crossed (see Figure 4.6).

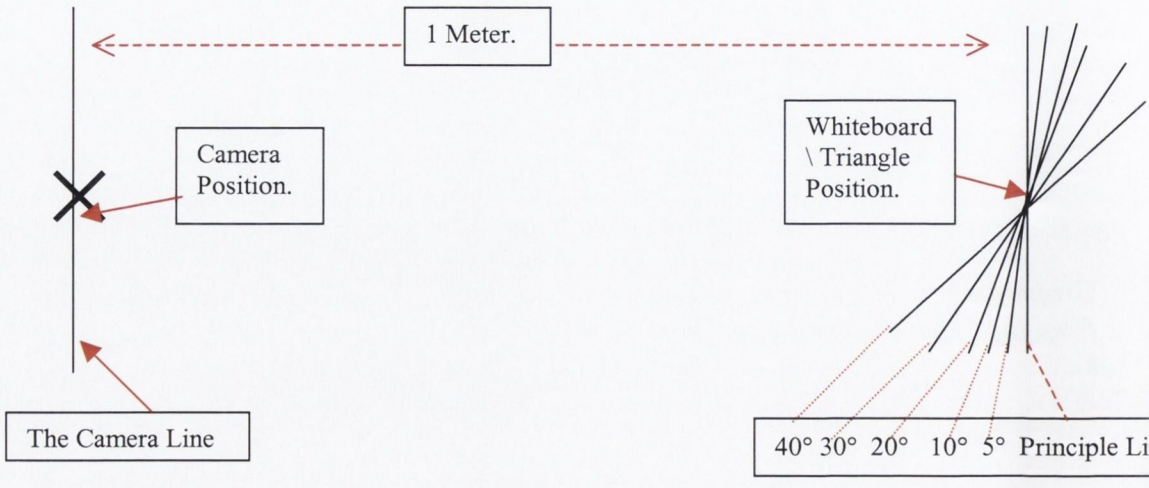


Figure 4.6. Illustration of Experimental Scheme into the effects of Rotation on the Uillinn Method©.

Initially, the whiteboard was positioned so it was exactly aligned to the 5° line on the floor. To ensure that it was exactly aligned, a plum-line was dropped from the whiteboard at several different points (see Figure 4.7). Only when all of the plum lines fell directly onto the 5° line was the board photographed. A total of 5 triangles were drawn and measured in this position. One photograph was taken for each triangle.

The board was then further rotated to the next position (10°). Plum-lines were again used to ensure that the board was in the correct position (see Figure 4.7). Again 5 triangles were drawn, measured, and were then photographed. The same technique was used for the rest of the angled lines, i.e. 20°, 30° and 40° to the principle line.

A total of 5 different random triangles were photographed in each amount of rotation. This produced a total of 15 random angles in each rotational position. This data was labelled according to the position the angles were measured in, i.e. control data 5° (CD5), control data 10° (CD10), etc (see Table 4.10).

The same examiner used the UM to calculate each of the angles in the photographs. The examiner was blind to the CD as those calculations were made after the UM data was gathered.

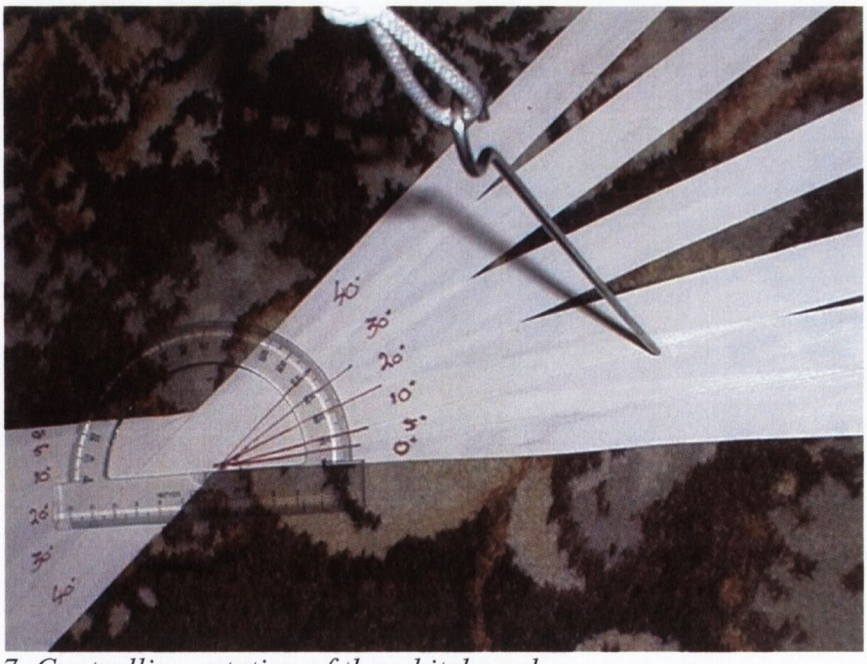


Figure 4.7. Controlling rotation of the whiteboard.

4.4.9. Results: Experiment 4.

This experiment investigated the influence of rotation during the photographic stage of the UM, and estimated how that rotation effected the precision of the angular measurements made by the system. The methodology assessed 15 different angles rotated away from the cameras lens, at 5 increasing amounts of rotation (see Figure 4.4, 4.6, 4.8).

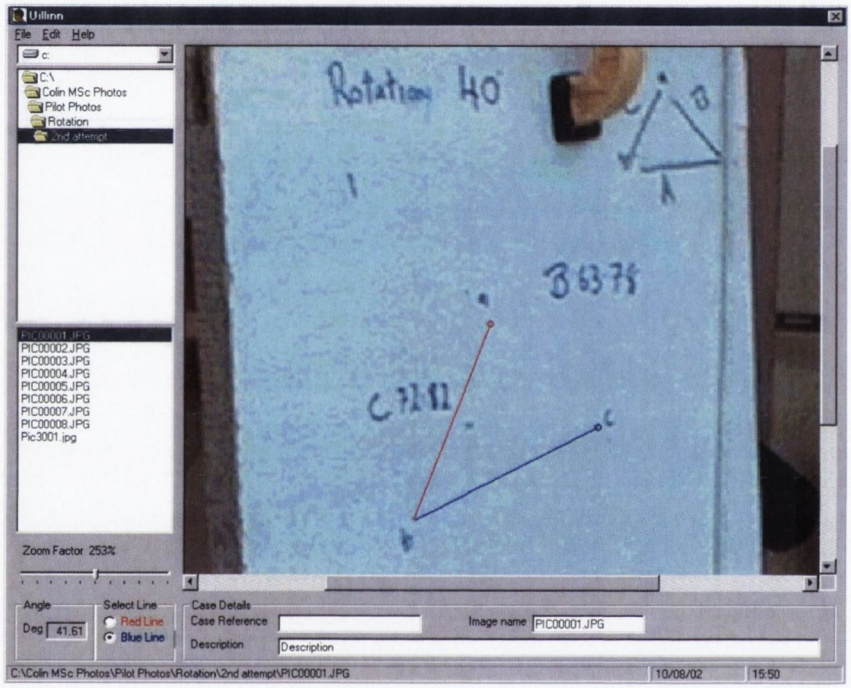


Figure 4.8. The Uillinn Method© of measurement: A photograph with the angle in 40° of rotation.

Standard deviations (SD), absolute means, the PPMCC and range of the differences between the UM and the CD are listed in Table 4.9, they reflect an increasing deterioration of precision with a larger rotational component. However at 5° rotation these measures were similar to those seen with optimal photographic conditions which implied that 5° rotation is not influential on the UM's angular measurements (see Table 4.6, Graph 4.8).

The mean differences between the CD and the UM did not change with increasing rotation and did not approach the 1.5° limit of SB. This implies that a SB of total triangle values did not develop with the addition of a rotational component to the photographic procedure.

Random error as assessed by the RC and TE was compared with the results from optimal photographic conditions (Table 4.6) and there was no difference in these measures up to rotation of 5° (see Table 4.9, Graph 4.14).

However, from 10° on random error increased sequentially. Ninety-five per cent scatter plots for each amount of rotation also showed a steady increase in the spread of the data with the increased rotation (see Graph 4.15, Graph 4.16, Graph 4.17, Graph 4.18, Graph 4.19).

	Length of Sides (mm)			Cosine Rule (°) Control Data (CD)			Uillinn Method (UM) (°)			Difference Between CD and UM (°)		
5°	Side a	Side b	Side c	Ang A	Ang B	Ang C	Ang A	Ang B	Ang C	Dif A	Dif B	Dif C
1	104.3	102.1	87.02	66.39	63.76	49.85	67.08	64.51	49.72	-0.69	-0.75	0.13
2	73.14	105	63.44	43.27	100.2	36.48	43.14	100.5	36.55	0.13	-0.3	-0.07
3	51.32	127.9	124.9	23.39	81.63	74.98	23.7	82	75.3	-0.31	-0.37	-0.32
4	105	92.44	69.74	79.33	59.92	40.75	78.73	59.89	41.06	0.6	0.03	-0.31
5	98.16	98.74	34.38	78.99	80.9	20.11	78.42	80.5	20.24	0.57	0.4	-0.13
10°	Side a	Side b	Side c	Ang A	Ang B	Ang C	Ang A	Ang B	Ang C	Dif A	Dif B	Dif C
1	131	71.6	131	74.14	31.72	74.14	73.29	31.13	75.1	0.85	0.59	-0.96
2	41.38	78.34	57.78	30.96	103.1	45.91	32.61	101.3	47	-1.65	1.8	-1.09
3	107.1	62.32	85.38	91.6	35.57	52.83	90.68	34.64	54.24	0.92	0.93	-1.41
4	60.7	106.5	80.48	34.44	96.98	48.58	32.86	97.62	49.02	1.58	-0.64	-0.44
5	97.96	66.28	41.74	128.7	31.86	19.42	127	31.67	20.29	1.7	0.19	-0.87
20°	Side a	Side b	Side c	Ang A	Ang B	Ang C	Ang A	Ang B	Ang C	Dif A	Dif B	Dif C
1	47.08	104.2	62.36	15.42	144	20.62	15.79	140.7	22.64	-0.37	3.3	-2.02
2	107.1	68.28	43.58	145.4	21.22	13.35	141	22.45	15.42	4.4	-1.23	-2.07
3	65.12	115.8	56.26	18.84	145	16.2	20.45	139.4	19.9	-1.61	5.6	-3.7
4	87.58	85.64	66.44	69	65.91	45.09	63.31	68.69	47.61	5.69	-2.78	-2.52
5	84.1	84.28	23.28	81.61	82.49	15.89	75.22	88.39	16.09	6.4	-5.9	-0.2
30°	Side a	Side b	Side c	Ang A	Ang B	Ang C	Ang A	Ang B	Ang C	Dif A	Dif B	Dif C
1	128.4	86.94	65.16	114.5	38.03	27.5	103.8	39.97	35.5	10.7	-1.94	-8
2	123.4	118.9	40.78	86.56	74.18	19.27	76.42	81.93	21.29	10.14	-7.75	-2.02
3	59.38	113.1	62.46	21.31	136.2	22.47	22.16	130	28.5	-0.85	6.2	-6.03
4	63.62	77.02	41.14	55.64	92.1	32.26	50.24	102	26.98	5.4	-9.9	5.29
5	76.78	71.44	40.18	81.76	67.05	31.19	70.33	76.25	32.41	11.43	-9.2	-1.22
40°	Side a	Side b	Side c	Ang A	Ang B	Ang C	Ang A	Ang B	Ang C	Dif A	Dif B	Dif C
1	93.46	63.78	72.82	86.07	42.91	51.02	65.94	41.61	72	20.13	1.3	-20.9
2	103.4	62.46	102.2	73.36	35.38	71.26	57.53	44.33	77.5	15.83	-8.95	-6.24
3	68.52	116.7	61.86	27.94	127	25.02	17.28	145.9	16.59	10.66	-18.9	8.43
4	97.12	82.84	66.38	80.39	57.24	42.37	61.02	70.69	48.15	19.37	-13.45	-5.78
5	128.1	89.6	60.38	116	38.93	25.05	102	40.06	37.9	14	-1.13	-12.9

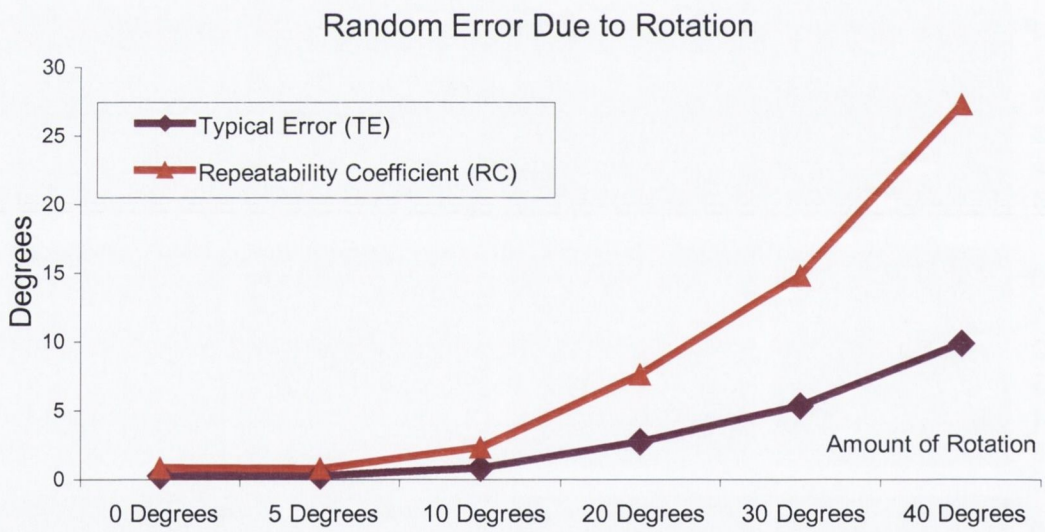
Table 4.10. Control Data (CD) and the Uillinn Method© (UM) data with Rotation during Photography .

Amount of Rotation	**0°** (n=120)*	**5°** (n=15)	**10°** (n=15)	**20°** (n=15)	**30°** (n=15)	**40°** (n=15)
Mean Difference	*0.088°*	*-0.09°*	*0.1°*	*0.2°*	*0.15°*	*0.1°*
SD (Differences)	*0.416°*	*0.41°*	*1.18°*	*3.86°*	*7.5°*	*13.9°*
SE of Difference ($\sqrt{SD^2/n}$)	*0.038°*	*0.1°*	*0.3°*	*1°*	*1.9°*	*3.6°*
Typical Error (SD/ $\sqrt{2}$)	*0.29°*	*0.28°*	*0.83°*	*2.7°*	*5.3°*	*9.8°*
RC	*0.816°*	*0.8°*	*2.3°*	*7.6°*	*14.8°*	*27.27°*
Range of Data	*-0.86° to 0.953°*	*-0.75° to 0.6°*	*-1.65° to 1.8°*	*-5.9° to 6.4°*	*-9.9° to 11.43°*	*-20.98° to 20.13°*
Absolute Mean Difference	*0.34°*	*0.3°*	*1°*	*3.2°*	*6.4°*	*11.8°*
PPMCC	*0.99*	*0.99*	*0.99*	*0.99*	*0.97*	*0.907*

* From Table 4.6 LOA = Limits of Agreement.

PPMCC = Pearsons Product Moment Correlation Coefficient.

Table 4.11. – Calculations of Agreement and Random Error for the UM due to Rotation.

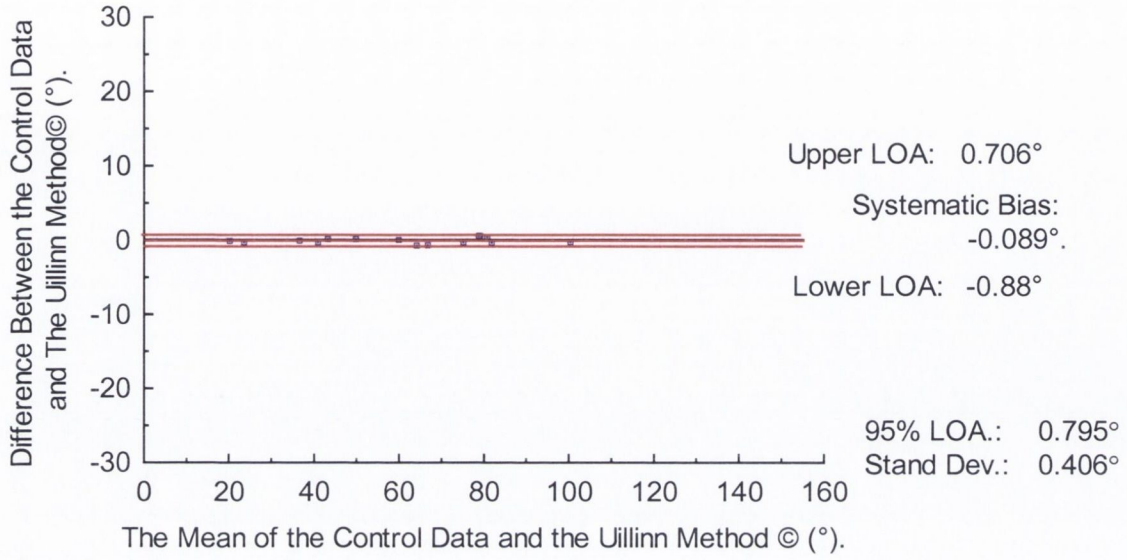


Graph 4.14. Random Error with Increasing Rotation between the Angle and the Camera.

Amount of Rotation	**0°**	**5°**	**10°**	**20°**	**30°**	**40°**
Typical Error (TE)	*0.29°*	*0.28°*	*0.83°*	*2.7°*	*5.3°*	*9.83°*
Lower confidence limit	*0.26°*	*0.22°*	*0.64°*	*2.1°*	*4.1°*	*7.5°*
Upper confidence limit	*0.33°*	*0.42°*	*1.21°*	*4°*	*7.8°*	*14.4°*

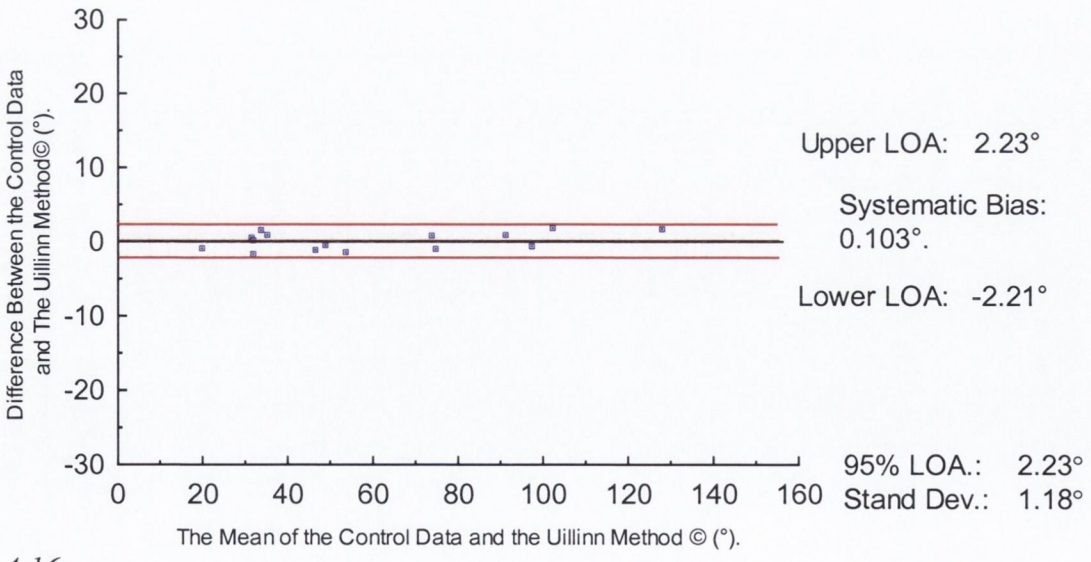
Table 4.12. – Typical Error due to Rotation (with Confidence Limits).

Bland & Altmans 95% Limits of Agreement Data Plot.
The Uillinn Method© Vs. The Control Data (**5°** rotation during Photograph).
(r=0.99 / n=15).



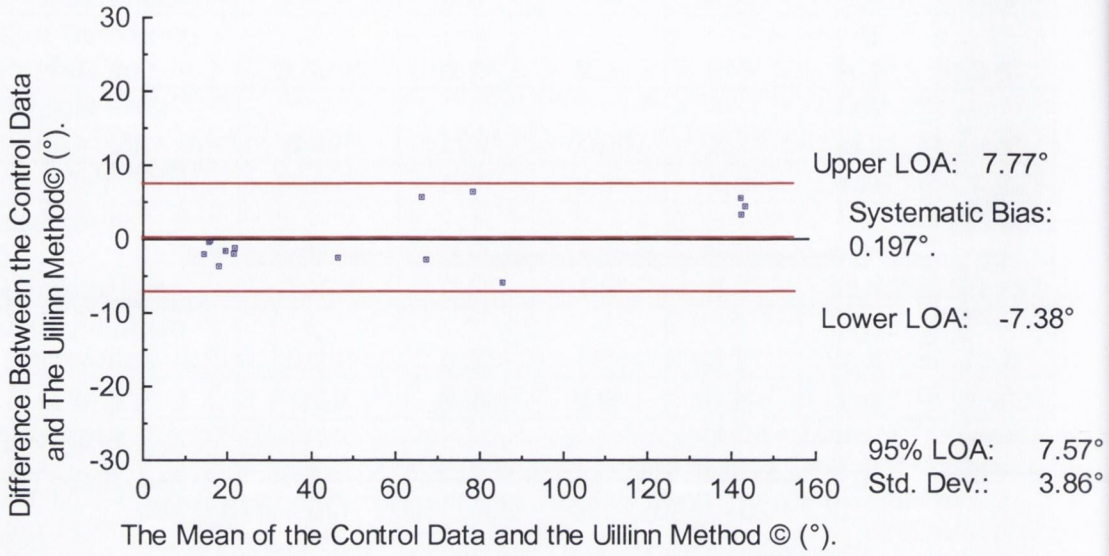
Graph 4.15.

Bland & Altmans 95% Limits of Agreement Data Plot.
The Uillinn Method© Vs. The Control Data (**10°** rotation during Photograph).
(r=0.99 / n=15).



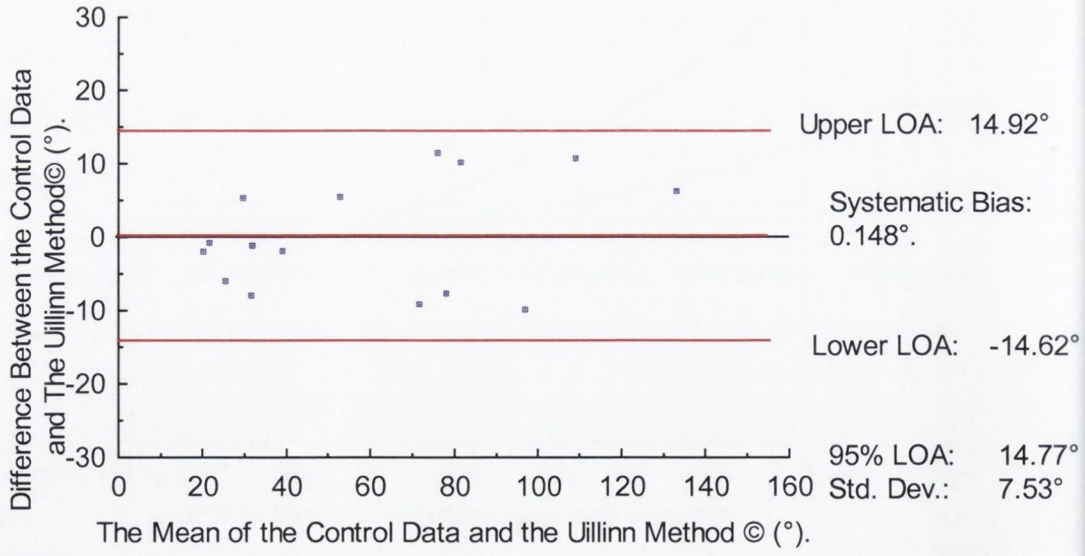
Graph 4.16.

Bland & Altmans 95% Limits of Agreement Data Plot.
The Uillinn Method© Vs. The Control Data (**20**° rotation during photograph).
(r=0.99 / n=15).



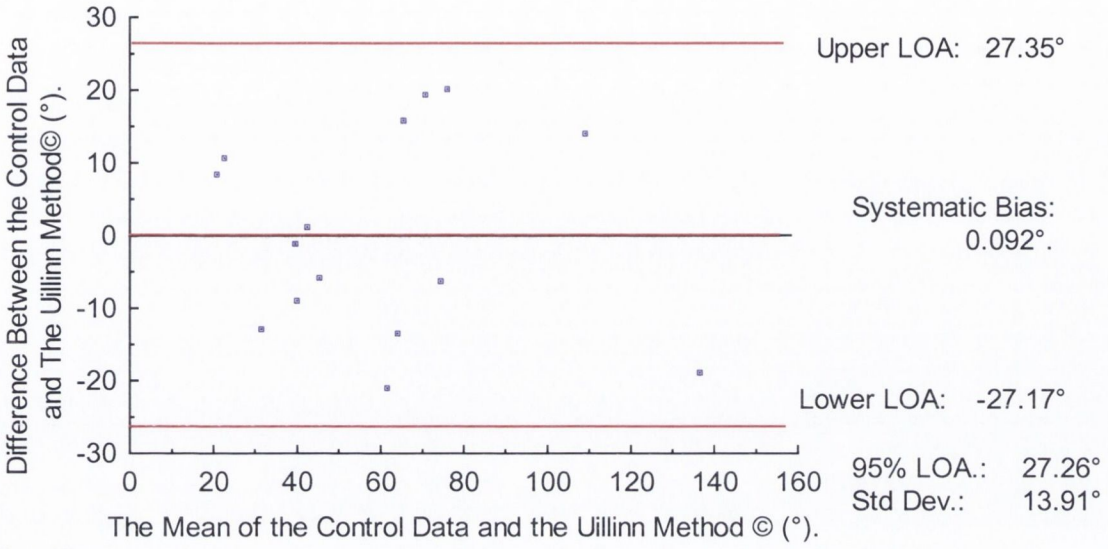
Graph 4.17.

Bland & Altmans 95% Limits of Agreement Data Plot.
The Uillinn Method© Vs. The Control Data (**30**° rotation during Photograph).
(r=0.98 / n=15).



Graph 4.18.

Bland & Altmans 95% Limits of Agreement Data Plot.
The Uillinn Method© Vs. The Control Data (**40°** rotation during Photograph).
(r=0.90 / n=15).



Graph 4.19.

4.4.7 Guidelines for Use of the Uillinn Method©.

The final aim of the study was to compile guidelines for the optimal use of the Uillinn Method© (UM) based on the information gathered from the different experiments. From the findings of experiment 2 and 3 it was clear that the use of a tripod was not required for the photographic component of the UM procedure. However, experiment 3 and 4 clearly illustrated the need for the angle to be located within the centre of the photograph and that there should be less than 5° of rotation between the camera and the angle. Therefore, in order to achieve precise angular measurements using the UM the proposed guidelines for the photographic component are;

1) The camera lens should be positioned in such a way that allows the majority of the angle to appear close to the centre of the photographic image. The viewfinder crosshairs should be aimed at the approximate centre of the angle and the cameras zoom-out feature or photographic distance should be adjusted to avoid the angle appearing at the edges of the photograph. The zoom-in tool in the UM software can be used to maximise the angle size during the measurement process.

2) An effort should be made to keep the perpendicular angle of incident of the lens to the target angle below 5°. To do this the photographer should stand in front of,

and perpendicular to, the angle that is to be photographed. The photographer should check the alignment of the target angle and ensure that the camera is held approximately parallel to it. If the photographer is not certain of the orientation of the target angle with regard to their point of view, the amount of error in the subsequent assessment of the image will be unknown.

3) During the photographic procedure the camera may be held free-hand or mounted on a tripod.

4.5 Discussion.

The UM concept represented a new method of angular calculation based on digital photography. It was unclear how well the system performed in relation to this task. A major concern for the UM, as a photographic based system was the possibility of error being introduced to the measurements arising from misrepresentation of the dimensions of an angle in the photograph. That is, if the photograph does not accurately depicting the exact dimensions of an object the systems virtual measurement arms (VMA) will be placed incorrectly and the calculated measurements will be inaccurate. It was also unknown whether the mathematical calculation system was accurate.

This study was designed to test these issues using inanimate angles. Specifically, its objectives were to establish the amount of error associated with the calculations and to determine sources of error. The study did not aspire to achieve external validity in the measurement of human joint angles.

The study had a simple design that initially aimed to establish the probable minimum level of error associated with its calculations (under optimal conditions). Once this data was known, the photographic procedure was changed to assess the effect of free-hand photography, the effects of rotation between the camera and the angle and image alterations due to the design of the camera or parallax error. While the study design was rudimentary it answered all of the questions posed by its initial aims.

When all the data was gathered it was possible to establish procedural guidelines, based on scientific evidence, to minimise procedural error.

The results show that there is no difference in error if a tripod or freehanded photographic technique is used and that TE for either technique will be 0.3° with a 95% agreement of less than 0.8° (n=120). This data clearly indicate that the measurement error for the system under ideal conditions is minimal.

Dunlevy et al (2000) performed a simpler version of this current experiment. The study assessed 20 random angles using a standard UG as the criterion measurement as oppose to the Cosine rule used in this current study. Their results showed average differences between the UG and the UM as 0.03° (± 0.2°). Their smaller SD of the difference reflects slightly higher expected agreement between

their control data (UG) and the UM. However, this slightly closer agreement may be a reflection of the lack of precision for the UG which is limited to increments of 1° and the relatively small number of angles they assessed. Also, their angles ranged from 17° to 93°, whereas the current study had data in a much larger range from 11° to 149°.

The sensitivity of the data analysis methods used throughout the different experiments was illustrated in the results for experiment 3. Small changes in random error were seen as the angle moved from the centre to the periphery of the photograph.

The TE associated with the optimal position of the angle in the photograph, i.e. the centre, was determined by experiment 1 and 2 (0.29°). However, experiment 3 showed that TE rose to 0.51° for the measurement of angles situated in the area half way between the centre and the edge of the photograph (Off-centre). It rose again at the edge of the image (Periphery), TE = 0.73°. These figures imply that TE may rise by an additional 0.22° for the off-centre area of the photograph and 0.44° for the periphery of the photograph, which is a small but noteworthy rise in error.

If a rotational component above 5° is introduced to the photographic procedure, the TE and the RC will increase much more dramatically (see Table, 4.11, Graph 4.15, Graph 4.16, Graph 4.17, Graph 4.18, Graph 4.19). A rotational component of 10° will push TE up by 0.5° (0.8°) and the RC will be almost 2.5°.

Gray et al (1991) investigated angular measurements made using X-ray films, they reported statistically significant increases in measurement error with rotation over 5°. This finding is in close agreement with the findings from the current study despite different methodologies. Both studies support the view that rotation over 5° will have a negative influence on measurement error.

Rotation of the target angle as oppose to the camera should be the main concern for the examiner. This is because if the camera is rotated it will be impossible to align the viewfinders crosshairs on the target angle and therefore will be obvious to the examiner. It was the experience of this study that rotation of the target angle greater than 5° should be easily discernable by the examiner and

therefore should not require equipment to ensure appropriate alignment, although no evidence is provided to support this view-point.

These studies provided a basis which allowed the completion of the final aim of the study. This aim was to use the data obtained from the experiments to compile a set of simple photographic guidelines for the optimal use of the UM. The study successfully drafted these guidelines using the evidence provided by the experimental results (4.4.7), the guidelines hoped to instruct the examiner as to how to avoid procedures that were found to increase random error to the measurement process and were therefore a guide to best practice.

4.5.1 Limitations of the Study.

A strong control data (CD) was needed for comparisons to be made with the UM data. This was provided by a mathematical model using calculations from the Cosine rule. This system of calculations required the precise measurement of each side of a triangle. A total of 79 triangles measured during all experiments, which corresponded to the measurement of 237 lines all of different lengths. The average length of these lines was 75mm (± 24mm).

All linear measurements were made using a Vernier measuring callipers, which is accurate to 0.02mm. The effect of that magnitude of measurement limitation would have on the CD was unclear. Therefore the amount of the error introduced to the methodology from this source was unknown.

To determine this error a simple calculation of the effect 0.02mm measurement error was theorised using an equilateral triangle model with 75mm as the length of its sides. This figure was chosen because it was the average linear measurement throughout all the experiments. When the theorised error was accounted for the dimensions of the theorised triangle were adjusted to, 74.98mm, 75.02mm and 75.02mm (see Appendix B). An equilateral triangle of exactly 75mm would have angles of 60°, however, the calculations with the theorised triangle produced an angle of 59.96, i.e. an error of 0.04°. It was acknowledged that this error was small but was considered a part of the overall random error in the study.

The dots that were used to depict the triangles on the white board were probably a greater cause for concern for the introduction of random error. These dots

needed to be approximately 2 mm in diameter so that they could be clearly seen on the photographs. It was postulated that substantial error could be introduced to the calculations if the exact centre of the dot was not used to gain the linear measurements. If the edge of the dots were used instead of the centres, an additional of 1mm per dot would be added to linear measure. This could accumulate to 2mm of error for both dots in the line. If the example of the 75mm equilateral triangle was used again, the length of the lines could have been measured as either 77mm or 73mm. A true equilateral triangle measured with this error would have recorded sides of 73mm, 77mm and 73mm. In this instance the Cosine rule would calculate the angles up to 3.6° away from the true 60° angle. Likewise, if the VMAs of the UM were positioned at the edges of the dots instead of their centre it would also have a negative effect on the measured data. An experimental test showed potential error of 3° in this scenario (see Figure 4.9), which is a similar magnitude to that potentially illustrated in the CD. These errors would have a harmful effect on the internal validity of the study. Therefore, it was vital that great care was taken during the measurements to use the centre of each dot when calculating both the CD and the UM data.

Mayerson and Milano (1984) report that there is some error associated with misreading analogue measurement tools, such as the universal goniometer (UG). Therefore, misreading of the callipers used in this study to measure the 237 lines was also a concern as a source of error. It was felt that this was particularly pertinent as the examiner found it quite difficult to read the callipers analogue scale. To overcome this problem each side of the triangles were measured twice.

In summary, the 3 sources of random error for the calculation of the CD were identified as the 0.02mm limitation to measurements made by the Vernier callipers, the possible misplacement of the measurement devices away from the centre of the 2mm dots (which constructed the triangles) and the possibility of misreading the callipers.

However, the low result of TE (0.29°) seen between the CD and the UM (n=120) reflects a low level of random error in the data (Batterham and George, 2000; Hopkins, 2000a), which suggests that the two datasets were in close

agreement and methodological error from these sources were successfully kept to a minimum.

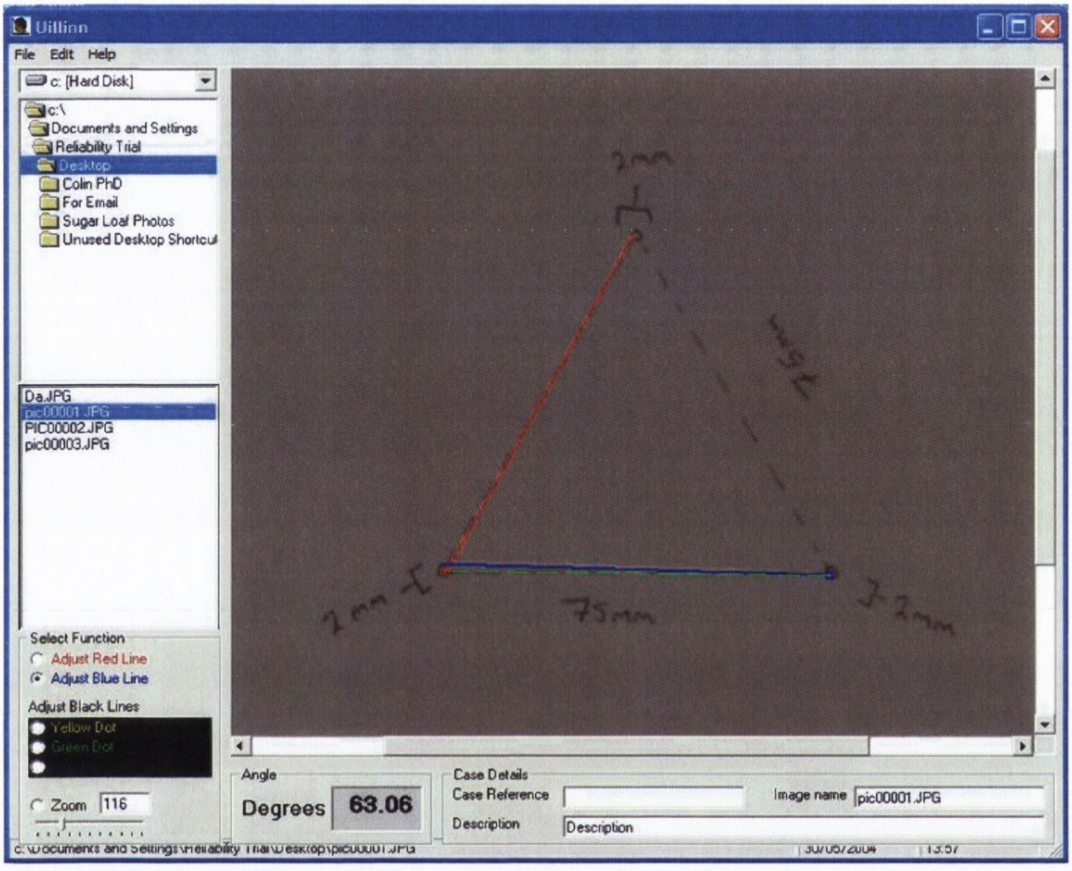


Figure 4.9. Potential Measurement Error of 3° of an Equilateral Triangle due to Inaccurate Placement of the Measurement Arms.

One possibly substantial limitation of using trianges to investigate error at the periphery of the photograph and perspective error due to rotation is the possibility that non-random error seems to have been difficult to detect. This may be because as there was consistenetly a set of 3 angles adding up to 180°, measurement of an obtuse angle would always be associated with the measurement of an acute angle, with may have balance a non-random error. It may have been wiser to separate the acute and obtuse angles and analysis them separately. This may be a consideration for further work.

4.5.3 Limitations of Data Analysis

The main conclusions of this study were drawn from both the RC and TE. Both of these analysis tools have been recommended as the optimal analysis tools

for the testing of precision and comparisons between repeated measurement of the same subjects, this ultimately implies the testing of reliability (Batterham and George, 2000; Hopkins, 2000a). However, due to the use of a criterion measurement (the Cosine rule calculations) that were essentially a gold-standard, this current set of experiments may be more correctly defined as validity testing. Additionally, measurements were only conducted once by each method which is a feature of a validity test. This along with the fact that the chief objective of the study was to provide evidence on issues akin to sensitivity and specificity, which are traits of validity, add further support that the study should have been labelled a validity study.

While the test of TE would remain appropriate in this instance (Hopkins, 2000a, 2000b), no authors appear to have suggested the RC for such a test. However, in defence of the RC for this trial, the design of the study set out to establish levels of agreement between two methods of assessing angles, neither of which, as discussed above were free from random error. Therefore, the trial may more correctly be labelled as an inter-method reliability study, instead of an internal validity study.

This issue regarding differentiation between a validity trial and a form of reliability trial has been discussed previously (1.2.3). Some researchers have controversially classified their work as validity trials when comparing the standard UG against the fluid filled goniometer (FFG) (Rheault et al, 1988; Petherick et al, 1988) or the UG against the internet based goniometer (Russel et al, 2003). However, Norkin and White (2002), state that only radiological studies provide strong enough evidence for a validity trial of this nature, which implies that they too may be best classified as inter-method reliability trials.

Whether this current study should be best classified as a reliability or validity study is largely academic, George et al (2000) have justified the 95% LOA test for validity testing, which is ultimately the same calculation as the RC (when there is no systematic bias (SB) present). In addition, Hopkins (2000a) states that the 95% LOA and the TE tests are appropriate methods of assessing random error between two cohorts. As the quantification of random error (and systematic bias) was the underlying goal of this study, along with the fact that the RC is based on the 95% LOA test, it would appear that the RC and TE were both appropriate analytical tools.

These analysis methods were all based on interpretations of the SD of the differences between the data sets, which allows for simple calculations and easy interpretation of the results (Altman, 1999; George et al, 2000).

Calculations based on the SD of the difference are the only type of analysis suggested in the recent literature (Bland and Altman, 1986; Atkinson and Nevill, 1998; Rankin and Stokes, 1998; Altman, 1999; Hopkins 2000a, 2000b; Batterham and George, 2000; George et al, 2000).

The accompanying 95% LOA scatter plots, descriptive data presentations and correlation graphs were used to complete the comprehensive provision of results.

4.5.4 Recommendations for Future Research

It is acknowledged that further investigation into error originating from the intrinsic camera design issues may be required. One issue that seems obvious is the need for an investigation into the effect of different lenses on the error that was observed in experiment 3. Higher quality lens, or camera systems may better cope with the error generated as the angle moved towards the edge of the image, than those observed in the Olympus, D-340L, 1.3 megapixel digital camera, which was used in this study.

The method of calculation of the CD was novel and appeared to be very effective. Calculations made using the Cosine rule yielded highly specific data. Increased accuracy in the linear measurements of the triangles would imply stronger calculations of the CD. Error may be introduced into the CD calculations by inaccuracy in determining the length of the triangle sides, by misreading the callipers, the limitations of the measurement tool scale and by resolving the Cosine rule equation incorrectly.

To ensure that these issues are limited as sources of error, it is recommended that a computer generated spreadsheet, programmed with the Cosine rule is constructed, this allows for simplicity in calculations of the CD. This was a feature in the current study and saved a lot of time associated with repeating Cosine rule calculations.

However, the relatively small analogue callipers that was used to make the linear measurements was felt to be problematic. They were difficult to read and

limited in the size (131mm) of the measurements that they could make. Larger triangles would have attached less relevance to the size of the dots used to construct the triangles and therefore lessened the size of the possible error associated with the placement of the equipment on the exact centre of the dots. Larger digital callipers are available but could not be sourced at the time that these experiments were carried out. The use of such callipers would make assured measurement precision of larger triangles possible. These devices are accurate to 0.01mm, which is an improvement on the accuracy of the tool used in this study. Additionally, reading the digital read-out of these instruments is less complicated. Armstrong et al (1998) found improvements in reliability when he compared a standard analogue UG with a digital read out goniometer.

The potential accuracy of the UM was ultimately established by combining the data from experiment 1 and 2. This gave a large amount of angles (n=120) upon which to base results. These angles were limited in a range of 11° to 149°. This limitation was due to the natural construction of triangles whose total angles must summate to 180°, this implied that in order to construct an easily discernable random triangle the range of angles were limited. A further investigation may wish to assess the full 360° of possible angles.

There may be some simple solutions towards the design of such a study, which may include assessing the external angles in the triangle instead of the internal angles. It would seem possible to construct a formula to allow the Cosine rule to accurately provide CD for such a methodology and therefore make the study possible.

While the study into the potential accuracy of the system had a large number of angles (n=120), the rotational experiments and the parallax/ rotation experiments had more limited numbers (n=15; n=24, 48 respectively). Hopkins (2000a) states that a degrees of freedom of at least 25 is needed to maintain less than 5% of bias in these type of calculations. Therefore, in order to improve confidence in these results a study with more numbers should be undertaken.

Another procedural consideration that was not investigated might have been the influence of perpendicular distance between the camera and the angle, i.e. how

perpendicular distance between the target angle and the camera effects measurement error. Dunlevy et al (2000) investigated this issue on a simple level and found no evidence of increased error with increased distances. The question itself may be irrelevant as the UM software can zoom in on any part of the photograph to over 300% of its original photographic size.

While the potential to increase size by 300% initially seems to be very advantageous, the zoom feature only allows practical enhancement up to a certain point, after which image quality suffers and the image is blurred making accurate measurements difficult. However, this issue would be likely to change dramatically with the amount of pixels attributed to the specific camera used to take the photograph. The quality of lens that the camera is using and the way in which the camera processes and stores its images would also be a factor (Dunmore, 2003). Therefore, it was felt that there was limited practical benefit to designing an experiment that investigated the distance between the camera and the target object as it would be difficult to infer these findings to other photographic equipment.

As stated above, rotation of the target angle above 5° will present an increase in measurement error to the UM. Also proposed was the opinion that it should be possible for the examiner to detect rotation of such magnitude without the use of specialised equipment. However, no support for this opinion was provided by this study and because this issue may have an enormous effect on measurement error for the UM, it is proposed that a study is need to investigate the ability of users to maintain correct alignment between the camera and the target angle.

The results of this study should not be used to infer to the application of the UM towards the measurement of human joint angles. This area contains far greater challenges when compared to measurement of inanimate objects (Bland and Altman, 1996; Hageman and Sorensen, 1991; Duckworth, 1999; Dunlevy et al, 2000). This current study was undertaken in an effort to assess the potential of the UM to accurately gauge the dimensions of an angle. Nonetheless, the UM has been proposed as a tool that might be used for osteokinematic (joint angle) measurement. Therefore, subsequent to this set of experiments another series of investigations would be required to address these issues. These investigations would deal with

issues such as the reliability and validity of human joint angle measurement made with the UM.

4.6 Conclusion.

The Uillinn Method© (UM) was designed to assess human joint angles. However, its method of angular calculation had not been tested sufficiently. This needed to be done prior to exposure to the human model. This type of investigation was necessary in order to establish the potential accuracy of the system and to determine whether the system required any further development or calibration.

An additional issue related to the photographic procedures employed and their potential contribution to measurement error.

It was found the UM would consistently assess angles to within 1° (95% LOA, 0.9 to –0.73) of a control data (CD) under optimal conditions, and typically this error would be only 0.3°. However, there was potential to impair this accuracy if 2 basic rules were not followed. These rules are that, firstly, the angle being assessed must be approximately in the centre of the cameras lens and, secondly, that there must be minimal rotation between the target angle and the camera. The latter issue has the potential to add substantial error to the measurements. The use of a tripod over holding the camera freehand during the photographic procedure did not have any impact on the results.

Chapter 5.

RELIABILITY OF THE UNIVERSAL GONIOMETER AND THE UILLINN METHOD©.

5.1. Introduction.

The concept of measurement reliability refers to the consistency of a measurement technique over repeated assessments (LoBiondo-Wood and Haber, 2002). Researchers must consider at least two themes when undertaking a reliability investigation; inter-rater and intra-rater reliability. Int*ra*-rater reliability describes the level of concordance between repeated measurements made by the same examiner. Int*er*-rater reliability deals with agreement of measurements made among different examiners.

Reliability is central to the appropriateness of a particular measurement tool for a specified purpose. In fact, it may be justifiable to count reliability as a more fundamental issue than validity (Batterham and George, 2000), as if a measurement tool is unable to reproduce the same data under the same circumstance, the validity of a measurement will be subject to change over repeated measurements.

The reliability of the universal goniometer (UG) is not clearly defined in the literature. The newly developed Uillinn Method© (UM) remains untested for these issues. Therefore, research into the reliability of both these tools was needed.

The series of experiments conducted in chapter 4 explored the accuracy of the UM's method of data calculation and how important the operational procedures were for minimising error. The study established the amount of error expected with the UM due to the fundamental system design. Only inanimate angles were used in the methodology and no predictions were made regarding the error associated with the system when measuring human joint angles.

The UM has 5 different approaches to the placement of it's virtual measurement arms (VMA) about a photographed joints of the knee and elbow (No Marker, Parallel, Mid-point, 3-Marker, 4-Marker) (see 2.3). Which of these provided the strongest reliability was an associated issue for the general reliability question.

One of the objectives of the study was to contrast the levels of reliability between the UG and the 5 UM approaches. Altman (1999) states that it is informative to compare new clinical measurement tools against the old standard methods.

Other issues that arose out of these reliability investigations relate to the reliability of using adhesive skin markers. This is a practice that currently appears to

be used in all measurement techniques based on photography, despite questions being raised about it's reliability and validity (Marks and Karkouti, 1996; Sati et al, 1996; Karkouti and Marks, 1997; Selfe, 1998; Szulc et al, 2000). The subject matter for the study was the joints of the elbow and knee.

5.2 Specific Aims of the Study.

1. To compare the levels of int*ra*-rater reliability for the 5 different approaches of the UM and the UG, for joint angle measurements at the elbow and the knee.
2. To compare the levels of int*er*-rater reliability for the 5 different approaches of the UM and the UG, for joint angle measurements at the elbow and the knee.
3. To report the reliability data with appropriate descriptive and statistical tools, including the Bland and Altman 95% limits of agreement, the repeatability coefficient and typical error (TE).
4. To report on issues regarding the reliability of re-applying adhesive skin markers by the same examiner for the purpose of measurements.

5.2.1 Justification of the Aims.

Aim (1). To compare the levels of intra-rater reliability for the 5 different approaches of the UM and the UG for joint angle measurements at the elbow and the knee.

The issue of intra-rater reliability for the 5 different UM approaches and the UG is important for a number of reasons. Commonly, a patient undergoing physiotherapy or medical intervention will be assigned to one primary professional over the course of their treatment. Therefore, typically only one examiner will assess the patient using any number of appropriate outcome measures. In the case of joint pathology, range of movement (ROM) assessments are frequently used as the main objective outcome measure. The examiner who is assessing the joint must know the likely error associated with their chosen measurement technique. This can help the clinician decide when an observed change is real or when it is due to random error.

Most researchers agree that intra-rater reliability will consistently be stronger than inter-rater reliability (see Table 1.2 to 1.5). Therefore, optimal consistency is assured by using the same examiner and the same type of measurement instrument over repeated examinations. Subsequently, an intra-rater reliability trial would probably quantify the most optimistic results for the reliability of the UM and the UG, and due to the supposed normal practice of health care provision by one

primary clinician the results from this trial may provide the most clinically relevant data.

Aim (2). To compare the levels of inter-rater reliability for the 5 different approaches of the UM and the UG for joint angle measurements at the elbow and the knee.

Similar to aim (1), an investigation into the inter-rater reliability of the UG and the 5 different approaches to the UM would provide clinically relevant data regarding the performance of these devices for measuring human joint angles and can also be justified as an important clinical issue. While it may be common that a course of treatment may be carried out by only one health professional, it is also common that during the initial diagnostic process many different health professionals may by involved. For example, a patient may initially attend their local general practitioner with a joint dysfunction, who may decide to refer them to a medical consultant in a secondary or tertiary setting, who in turn may refer them to a physiotherapist or vice versa. Consequently, three or more health professionals may be involved in the care path for this patient. Each of these professionals may elect to use the same outcome measure to assess the patient. If the measurement techniques used do not compare well or posses substantial inaccuracies there may be confusion between these health professionals regarding the extent of the dysfunction or even the diagnosis. Therefore it is important that there is reasonable agreement between each of the professionals measurement data, to achieve this the same measurement procedures must be used.

From the literature review there was evidence to support the notion that reliability is improved when the same device is used during repeated examinations (Clapper and Wolf, 1988; Greene and Wolf, 1989). There is also evidence to suggest that random error increases when different examiners are involved in the measurement procedures. It is vital that the amount of this error and its precise impact on measurement reliability is known.

Aim (3). To report the reliability data with appropriate descriptive and statistical tools, including the Bland and Altman 95% limits of agreement, the repeatability coefficient and typical error (TE).

Poor statistic analysis has been a major limitation of much of the previous research carried out into the reliability of osteokinematic measurement techniques. Therefore, it was of utmost importance to conduct appropriate and informative statistical analysis on the data gathered during a reliability investigation. While there is some controversy regarding the best analytic tools to use, most recent publications back the use of Bland and Altmans 95% limits of agreement (LOA) or repeatability coefficient (RC) (Bland and Altman, 1986; Atkinson and Nevill, 1998; Rankin and Stokes, 1998; Altman, 1999; Hopkins 2000a, 2000b; Batterham and George, 2000; George et al, 2000). Hopkins (2000a, 2000b) further suggest that an assessment of typical error (TE) is important. That author also supports the use of correlation coefficients, which also would allow some comparisons between this current study and previous studies in which such analysis was commonplace.

It was hoped that at the conclusion of this set of experiments, the data analysis and graph design would allow the reader to easily understand and interpret the findings from this research.

Aim (4). To report on issues regarding the reliability of re-applying adhesive skin markers by the same examiner for the purpose of measurements.

The main objective of this chapter was to establish the level of reliability for the UG and each of the 5 different approaches to the UM. Two of these approaches, the '3-Markers' and '4-Marker', used adhesive skin markers to highlight anatomical landmarks on the limb segment. These markers are used to help the examiner construct the angle and correctly place the VMAs on the 2-dimensional (2D) photographic image. However, while this approach is recommended and commonly used by researchers (Fish and Windgate, 1985), its reliability has been questioned (Marks and Karkouti, 1996; Karkouti and Marks, 1997). Therefore, it was important to examine the error associated with the re-placement of adhesive skin markers on anatomical landmarks over repeated joint assessments, and to comment on the necessity of using skin markers for photographic measurement systems.

5.3 Methodology.

The outlined aims were concerned with testing the precision of the measurement tools themselves and not the ability of the subject to re-gain the same the position over repeated examinations. Therefore, subject position had to be controlled tightly. To achieve this the study design chose to make the necessary repeated assessments without moving the subjects between examinations. This would help to ensure all examinations were conducted under the same conditions, and minimise biological error.

Data for both the intra-rater and inter-rater reliability of different measurement tools would need to be provided by the study design. This meant that repeated assessments of the same joint position, by the same and different examiners, using different measurement techniques would be required. The procedure needed to consider that a single examination would entail assessment using a blinded UG, photography with 4 adhesive markers placed over anatomical landmarks and photography without markers on the limb segment.

This entire examination process would have to be repeated twice, by a number of different examiners. It was envisaged that at least 3 minutes might be needed for each examination, which would be enough time to ensure that the examinations were not rushed and therefore minimise examiner error.

The time involved in the examinations and its effect on the comfort of the subjects, who would be asked to remain perfectly still during the procedures were seen as a possible limiting factor for the study. This issue would primarily limit the number of examiners that could be involved in the assessments and would particularly limit the degrees of freedom for the inter-rater reliability aspect of the study. Consequently, only 3 examiners were asked to participate in the experiments, which meant that the subjects would be asked to remain static for approximately 18 minutes for both their elbow examinations and the knee examinations. The study procedures were therefore deigned to accommodate the participation of 3 examiners, which primarily influenced the number of subjects that could be examined simultaneously.

A rotational system of measurement procedures was designed that allowed each examiner to continuously rotate between subjects during the research sessions. This rotational system prevented the observance of one examiners measurement

technique and results by the other examiners, who would be conducting their own examinations at the same time, and maximised time efficiency.

5.3.1 The Examiners

The 3 examiners used in the trial were qualified physiotherapists who were either 2 or 3 years post graduation. The examiners were initially given a tutorial on the UM which included instruction about the systems best operational and photographic procedures (see 4.4.7; Appendix C, D) and the importance of these procedures towards minimising the presence of random error in the results.

The 5 different approaches to the application of the UM were verbally explained to the examiners and they were given a short period of time to practice the operation of the UM using these different procedures. When the examiners were content that they could operate the digital camera and the UM software, the investigative procedures were commenced. No instruction was given regarding the use of the standard UG, except to explain the method of blinding of this tool.

5.3.2 The Subjects

A total of healthy 52 subjects (14 Male, 38 female) were used in the study. The subjects were drawn from volunteers who worked in the Adelaide and Meath incorporating the National Children's Hospital (AMNCH), Dublin 24, Ireland, which was the location for the study. None of whom had a history of elbow or knee pathology or were currently in any musculoskeletal pain. The subjects were approached informally as a sample of convenience. The nature of the study and their potential involvement was explained and consent was gained verbally.

While 52 subjects were involved, 51 subjects had their knees assessed and only 49 had their elbows measured. This disparity in numbers was due either to the subjects being called away by work commitments during the measurement sessions or one episode of technical difficulties when one of the digital cameras batteries failed during the measurement sessions and caused the session to over-run and ultimately to be abandoned before completion.

5.3.3 Equipment

A large room in the physiotherapy department of the AMNCH was used to conduct all of the examinations. Three subjects attended together for each of the multiple test sessions that were held. Upon arrival to the test area they were provided with a loose pair of shorts and a sleeveless vest and asked to remove all other unnecessary clothing, such as shoes and socks.

Three adjustable examination plinths and non-adjustable chairs with supportive arms and a fixed upright back were used to position the subjects during the assessments.

When knee examinations were being conducted, the back rest of the plinth was placed in the upright position so that the subject could be positioned in long sitting. When the elbows were being assessed the subject's chair was placed directly beside the middle of the plinth, which was raised to the level of the axilla of each patient. With this set-up the subject's arm could rest at 90° abduction on the plinth when they sat on the chair.

The study sought to assess random angles at both joints, yet to be able to report on a wide range of different angles. To facilitate this it was decided that at each of the 3 plinths the subjects would be asked to position themselves in either an inner range, a mid-range or outer-range joint position. Depending on which plinth the subject randomly decided to go to, they were asked to comply with the appropriate joint position.

If the subjects were not familiar with the concept of inner, mid or outer joint range, they were asked to bend their knee or elbow 'a lot', 'in the middle' or the bend their joint only 'a little or keep it straight'. Initially, each subject was also asked to choose whether they would like to have their elbow or knee measured first. If the chose their knees, they were asked to long sit on the bed, if they chose their elbows they were asked to sit on the chair with their arm resting on the plinth. A mix of left and right was achieved by the orientation of the plinths in the room.

When the subject was being positioned they were asked to get themselves into a comfortable position, which could be maintained for approximately 15 to 20 minutes. This was done to minimise movement during the examinations each of the subjects limbs were supported using a number of sand bags or other types of weight bags (approximately 2 kg, each) (see Figure 5.1, 5.2). In addition, when the outer range knees were being positioned, a small wooden wedge was used to prevent

discomfort from prolonged plantarflexion of the ankle and to improve surface area contact with the plinth to maximise friction between the surfaces and inhibit the foot from slipping on the plinth surface (see Figure 5.3). The wedge was also used to support the shank during measurement of knee hyperextension (see Figure 5.4).



Figure 5.1. The 2 kg weight bags used to stabilise the foot (no markers, first rotation).

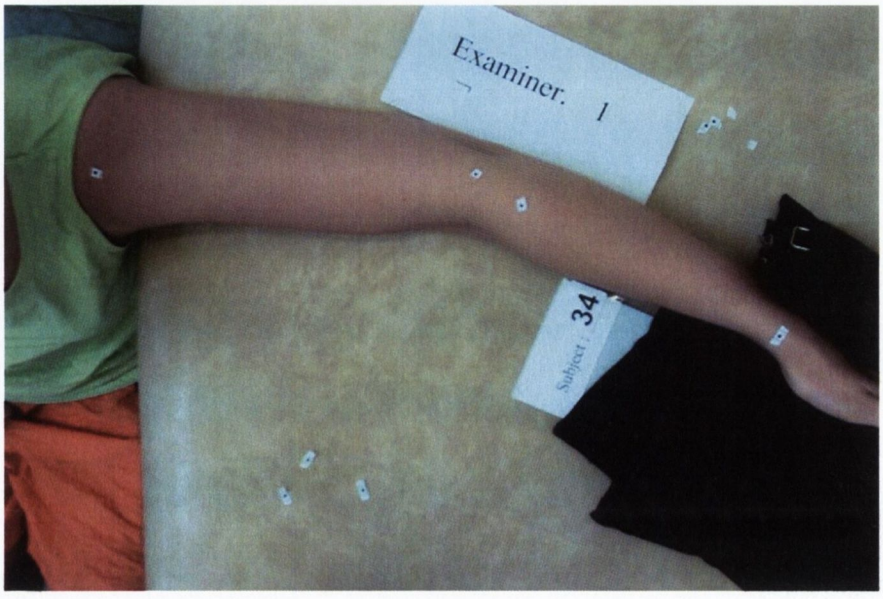


Figure 5.2. The 2 kg weight bags used to stabilise the arm (with markers, first rotation).



Figure 5.3. The 2 kg weight bags used to stabilise the foot with the wooden wedge.

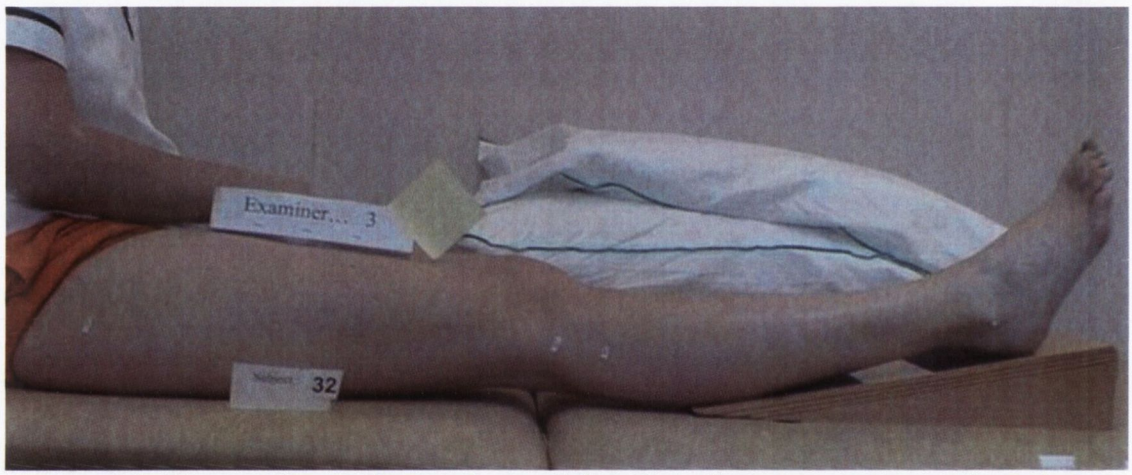


Figure 5.4. The wooden wedge used to stabilise Knee hyperextension (with markers, 2nd rotation).

Adhesive skin markers were need for 2 of the 5 approaches to the application of the UM. The markers were custom made on a Microsoft® Excel 97 SR-1 spreadsheet [Redmond, Washington, USA], by constructing a page of 5 mm by 5 mm grid squares, each with a small black dot (2 mm in diameter) in it. Only every second square had a dot in it. The spreadsheet of small squares were then printed on A4 sheets of adhesive label paper using a standard ink-jet style printer. The individual squares were then cut into strips using a pair of scissor without removing them from the labels. When the examiner wanted to use an adhesive dot during the measurement procedure they simply ripped one off the strip as they went.

The photographs were taken using either an Olympus D-340L 1.3 Megapixel digital camera [Olympus Optical Co., London, UK] or a Polaroid PDC 2070, 2.1 Megapixel digital camera [Polaroid co. Waltham, MA, USA]. There were only two

cameras used in the study which were shared between the examiners during the procedures. Both cameras produced pictures of a similar quality and size.

The pictures were transferred from these cameras to a 1.3 GHz PC [Intel Corporation, Hillsboro, OR, USA] via a SmartMedia card drive [SanDisk, Sunnyvale, CA, USA.] (see Figure 4.3, page 139). When the photographs were on the PC they were renamed, sorted and grouped according to the examiner and subject number. The PC used in the study had the UM software (version 2.4) previously installed on it.

There were 3 UGs used in the study, one placed at each measurement station. Each of these devices had a circular piece of white card attached to its measurement arm (see Figure 5.5). This card prevented the examiner reading the angle while placing the UG about the joint, but allowed free movement between the stationary arm and measurement arm. Only when the UG was removed from the joint and turned over could the examiner read their calculated measurement. This practice successfully blinded the UG data from the examiner and allowed unbiased joint angle measurements.

The data was recoded and stored appropriately on a series of Microsoft® Excel 97 SR-1 spreadsheets [Redmond, Washington, USA]. The data analysis and graph creation was performed on this package and KyPlot 2000© (Koichi Yoshioka).

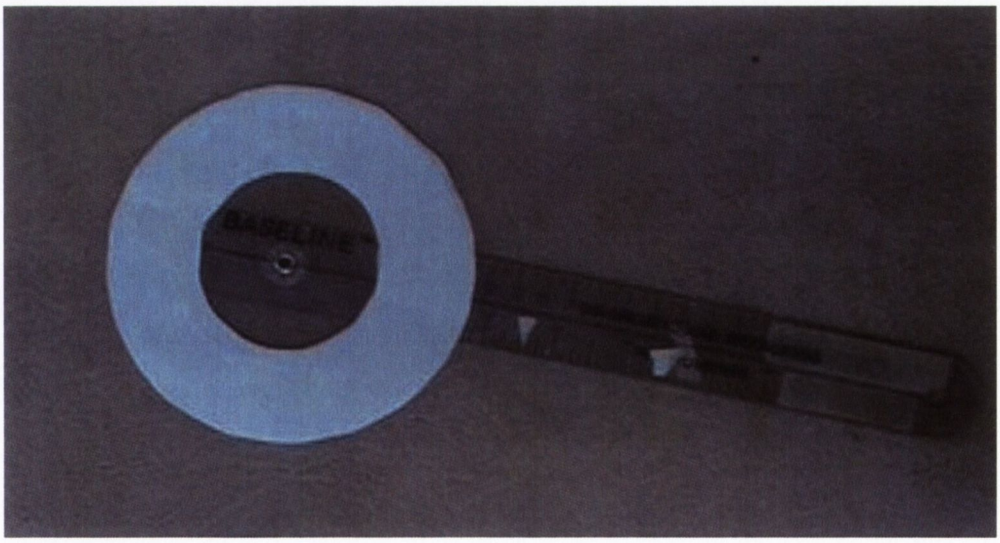


Figure 5.5. One of the Blind Universal Goniometers (UG).

5.3.4 Order of Measurements

The procedures aimed to allow the simultaneous assessment of 3 subjects by 3 examiners. To achieve this in the most efficient time a rotational system of examination was constructed, Examiner 1 started their measurements at the station where Subject 1 was situated, Examiner 2 started at Subject 2 and Examiner 3 commenced at Subject 3. The measurements procedures followed were; a blind UG measurement, a sagittal photograph, followed by application of 4 adhesive markers and further photography.

When all examiners had completed their assessments they rotated to the next subject and repeated the assessments again. This circuit of three-way alternations was completed twice. At its completion each examiner had visited and measured each subject on two occasions. When this was done, the subjects changed position to allow assessment of their other joint (either elbow or knee). After the switch and re-set-up of their positions, two rotations of measurements were repeated again.

Only three subjects were assessed at anyone session, it was usual that only one session was carried out per day, normally between 12:50pm to 1:30pm. On two occasions an evening session was held (at approximately 5pm) in addition to those carried out during the day. Due to logistical considerations measurement sessions were not carried out everyday, but were completed over a 7-week period in May and June 2003.

5.3.5 Examination Procedures

Just prior to commencing the measurements all of the subjects who were to have their knees assessed had their shorts rolled up to expose the greater trochanter, those who were to have their elbows assessed were checked to ensure that no clothing or jewellery covered their arm from their acromion to the styloid process of their radius. They were than asked if they were comfortable or if the needed to adjust themselves and they were instructed that it was vital that they remained in that particular position until all examinations were complete.

A small sign was placed beside the subject to indicate the particular subject number. Each examiner also had an identifying sign that they placed beside the subject that was being assessed. The second round of measurements were distinguished by a yellow square of adhesive paper attached to the subject number

(see Figure 5.4). Therefore, it was possible to determine the subject number, the examiner and the assessment number by the information in each individual photograph.

The examiners were reminded to follow the guide to optimal photographic procedures that was provided for them during their training (see Appendix C), with particular attention to avoiding parallax and rotation between the joint and the camera.

When measurements were commenced each examiner recorded their UG data on a pre-lined sheet of paper with subject numbers printed on it. Each examiner was blind to the other examiners results.

Photography of the lateral aspect of the knee was taken at a distance of about 1 meter. The examiner had to squat down in order to get the correct alignment between the knee joint and the cameras lens. In the case of elbow photography, each of the subjects' arms were resting in abduction on the plinth with the subject sitting on a chair beside it. Consequently, the lateral aspects of these joints were facing upwards. The examiner had to stand on the plinth and point the camera downward in order to correctly align the camera to the joint (see Figure 5.6). This examiner position was designed to control error in the methodology and is not recommended as a standard procedure for the UM.

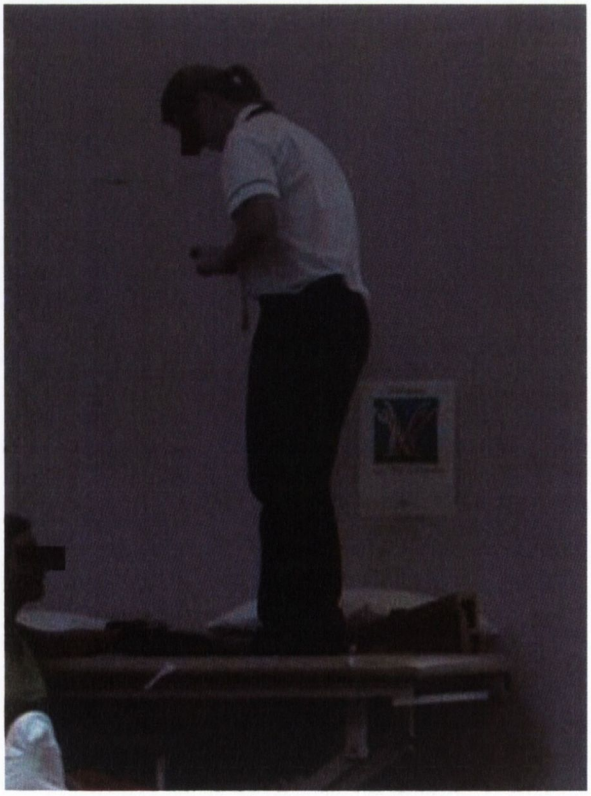


Figure 5.6. The Photographic procedure at the Elbow.

When this initial photograph was taken, the examiner placed adhesive markers to the skin at 4 specified anatomical landmarks (*Knee:* the prominence of the greater trochanter, the inferior tip of the lateral epicondyle, the superior tip of the head of the fibula and the inferior tip of the lateral malleolus (see Figure 2.9); *Elbow:* the antro-lateral tip of the acromion process, the inferior tip of the lateral humeral epicondyle, the prominence on the radial head and the distal aspect of the radial styloid (see Figure 2.10)). The examiners were instructed to place the markers as precisely as possible. After the photograph with the markers in-situ was taken, the examiner removed all of the markers prior to moving on to the next subject in rotation. All examinations were considered complete for a particular session when all of the examiners had assessed both the elbow and knee of each of the subjects.

5.3.6 Measurements by the Uillinn Method©

All of the photographs from both cameras were transferred to the PC on the same day as they were taken. They were then sorted appropriately according to examiner number, subject number, joint and whether it was the first or second photographic assessment. Ten weeks after the last session was completed, examiner number 2 was asked to commence their joint angle calculations using the UM. It was felt that this period of time would diminish to ability of the examiner to accurately recall the previously palpated anatomical landmarks.

The period of time between final subject assessment and commencement of calculations with the UM was even longer for examiner 1 (4 months) and examiner 3 (who was the author) (6 to 8 months).

Prior to conducting the assessments the examiners were again given a trial of using the system to measure elbow and knee joint angles using photographs unrelated to the study. They were also made aware of the rationale behind the different approaches. Only when the examiners were content that they could correctly carry out the operational procedures were they permitted to commence examinations.

The examiners used the UM software version 2.4 to complete the measurements. The examiners were instructed to assess the subjects consecutively

using only one method per photograph, i.e. the entire subject selection were assessed using one of the 5 approaches before moving on to the next approach. The order in which the approaches were applied to the photographs was predefined. The first three approaches used the photographs that did not have adhesive skin markers visible, the order was the No Marker, Parallel and Mid-point approaches (see Figure 2.5, 2.6, 2.11, 2.12, 2.13, 2.14, 2.15, page 99 to 110). Then the examiners moved on to the photographs that had skin markers visible in them and used the 3-Marker and finally the 4-Marker approach (see Figure 2.7, 2.8, 2.9, 2.10, page 103 to 105). This order of measurements was followed to minimise examiner bias that may have been present if the examiners had been exposed to the photographs containing the skin markers, which would have given them knowledge regarding the location of the bony points.

At the end of the UM assessments the examiners had twice assessed 49 elbows and 51 knees using 5 different approaches to the placement of the VMAs. This added up to 10 different assessments (Session 1 (No Marker, Parallel, Mid-point, 3-Marker, 4-Marker); Session 2 (No Marker, Parallel, Mid-point, 3-Marker, 4-Marker)), per joint per subject. This presented 510 knee measurements and 490 elbow assessments, i.e. a total of 1000 calculations per examiner. Due to this large number of assessments the examiners were asked to spread out their measurements times and not to conduct a lot of UM calculations during any one sitting. This was recommended in an effort to prevent fatigue which may have increased examiner error.

5.3.7 Data Analysis

The data analyses methods that were used and the rationale behind their selection were similar to those discussed in chapters 1 (see 1.4.4). In brief, TE and either LOA or RC would be used to calculate inferential data regarding both random and non-random error in the data. All of the LOA and RC comparisons would be accompanied by appropriate graphs. There was a limited use of significance testing and correlation coefficient analysis (PPMCC and ICC (3,1)).

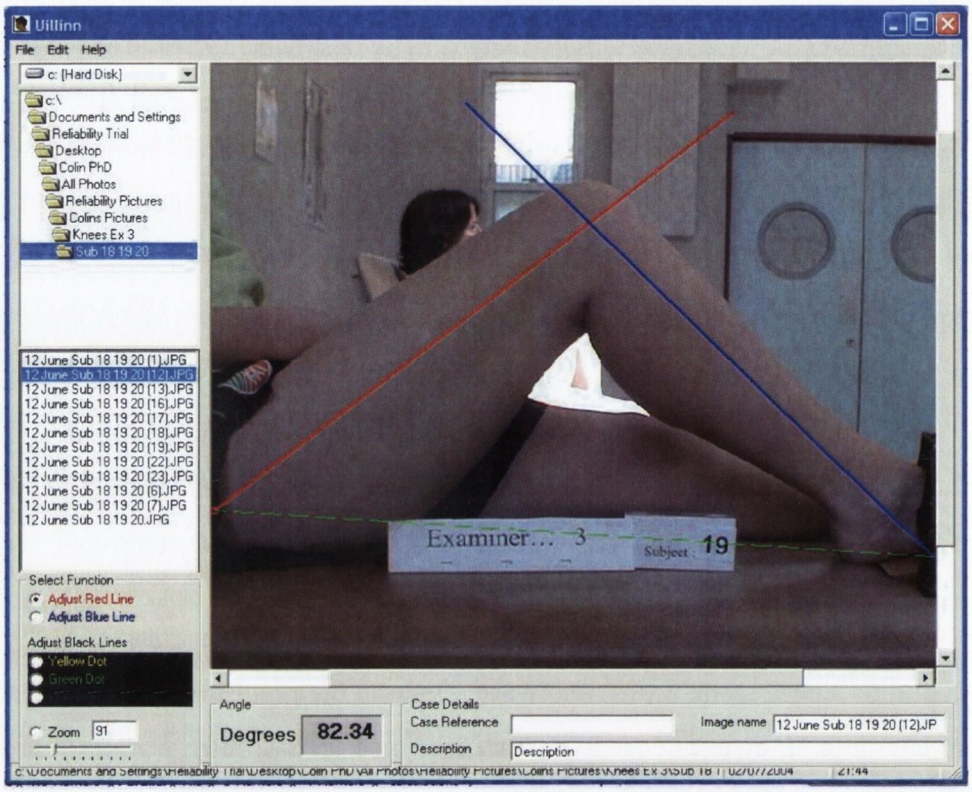


Figure 5.7. The Uillinn Method Software. Analysis of the knee flexion angle (Parallel method).

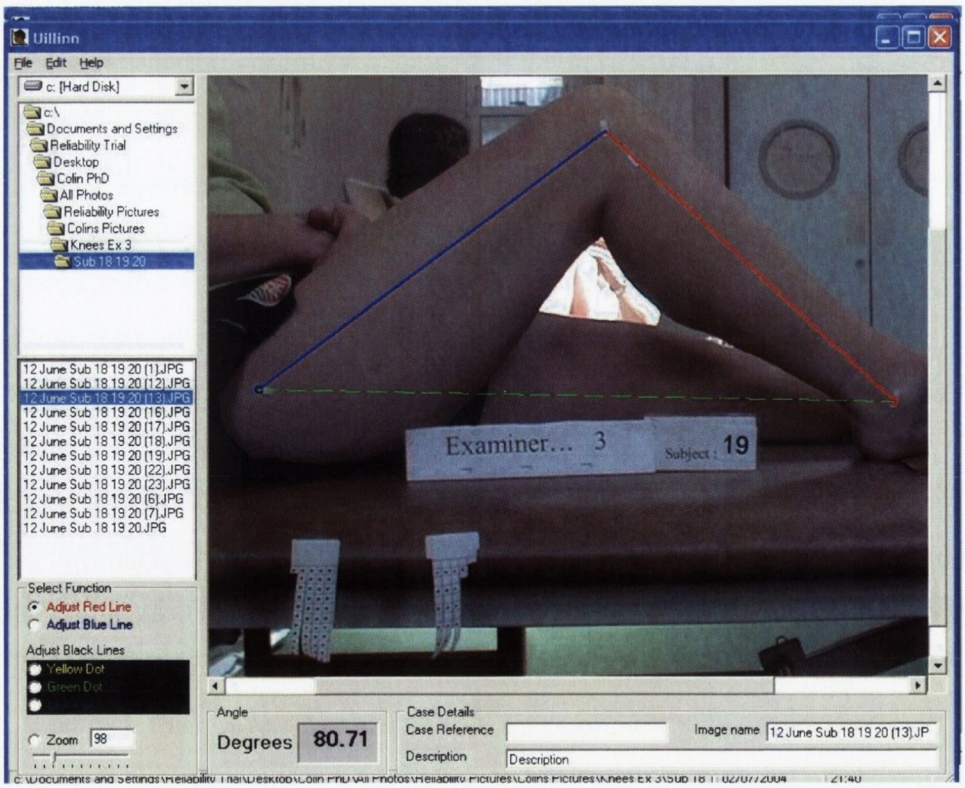


Figure 5.8. The Uillinn Method Software. Analysis of the knee flexion angle (3-Marker method).

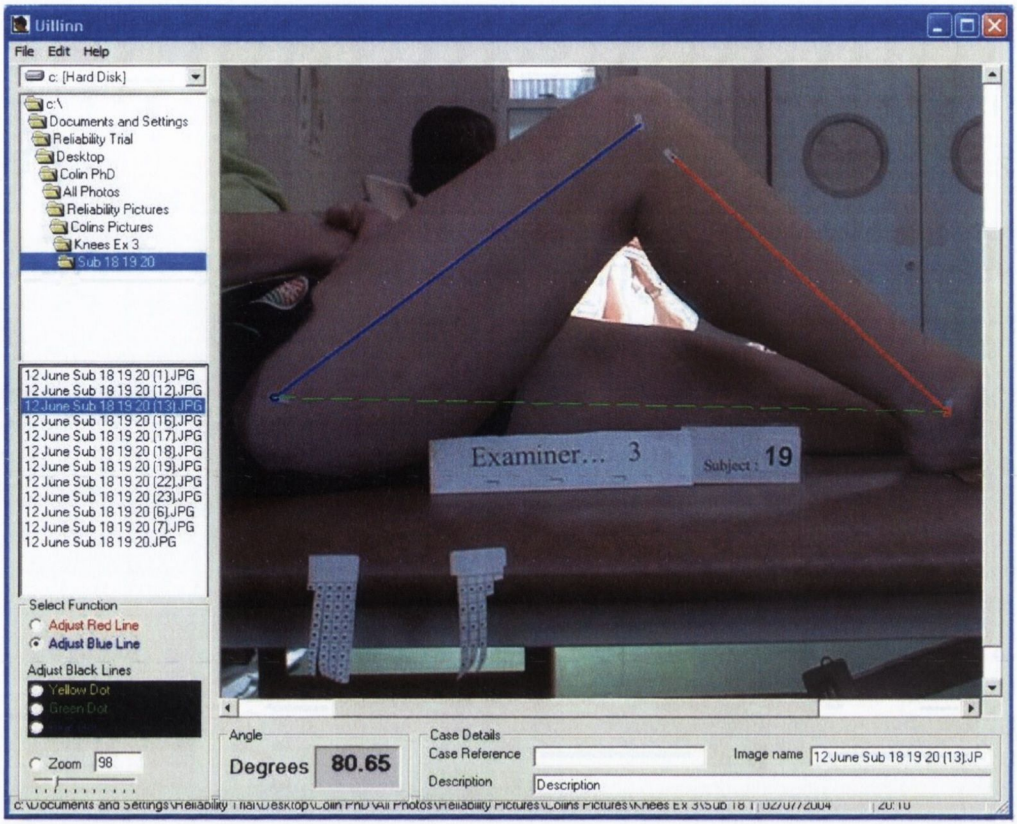


Figure 5.9. The Uillinn Method Software. Analysis of the knee flexion angle (4-Marker method).

5.4 Results.

Three examiners were involved in the trials each repeatedly assessed a total of 51 knees and 49 elbows. Inspection of the UG data revealed that during the investigation the subjects were measured in a range of –2° to 150° (mean 81° ± 46°) at the knees and a range of 17° to 138° (mean 74° ± 37°) at the elbow (see Appendix E).

Four of the elbow measurements were found to contain movement artefact (see Figures 5.10, 5.11, 5.12, 5.13, see page 250 - 252). This added relatively large changes to the repeated measurement data. For this reason 4 subjects were excluded from the study. This implied that the elbow cohort only contained 45 subjects. No knee subjects were excluded on the grounds of unwanted movement.

5.4.1 Intra-Rater Reliability

Each of the 3 examiners carried out repeated measurements of both the elbow and knee joints. This section will outline the results for the intra-rater reliability for each examiner and each of the measurement techniques. The results for the knee and the elbow will be dealt with separately.

Intra-Rater Reliability at the Knee

5.4.2 Universal Goniometer at the Knee

The average measurements for each of the examiners was approximately 81°, this indicated that the limits of SB should be set at 2°. The average difference between the assessments for each of the examiners were well below this limit (see Table 5.1). As SB was not evident in the measurements the data analysis used the repeatability coefficient (RC) rather than the 95% limits of agreement (LOA) test.

The RC showed that for 95% of repeated assessments the UG would agree to within; between 4.3° (examiner 1) to 6.6° (examiner 2). These were the best and worst results seen across the 3 examiners. While, the assessment of typical error (TE) revealed that repeated UG measurements would normally vary by between 1.56° (examiner 1) to 2.41° (examiner 2). The intraclass correlation coefficient

(ICC) (3,1) was calculated as 0.99 between each of the examiners repeated measurements.

Knees	Examiner 1	Examiner 2	Examiner 3
Average Difference	**-0.29°**	**0.55°**	**-0.04°**
Standard Deviations of Differences (SD)	2.21°	3.39°	3.34°
Standard Error of Differences (SE)	0.31°	0.47°	0.46°
Significant Difference (P=0.05)	No	No	No
Repeatability Coefficient (RC)	**4.33°**	**6.64°**	**6.54°**
Typical Error (TE)	**1.56°**	**2.41°**	**2.35°**
Lower Confidence Limits of TE	*1.34°*	*2.07°*	*2.02°*
Upper Confidence Limits of TE	*1.88°*	*2.89°*	*2.83°*

Table 5.1. Intra-rater Reliability of the Universal Goniometer (the Knee)

5.4.3 Uillinn Method© (No Marker Approach) at the Knee.

Measuring the knee flexion angle with the No Marker approach to the UM showed no SB between repeated assessments. This finding was based on the highest average measurement seen in the data for examiner 2, which was 84.5° and set the limits of SB 2.1°. This was below the average difference of 0.32° (examiner 3). The RC ranged from between 2.34° (examiner 1) to 4.86° (examiner 3) and the TE ranged between 0.85° to 1.77°. The ICC (3,1) between each of the examiners repeated measurements was no lower than 0.99.

Knees	Examiner 1	Examiner 2	Examiner 3
Average Difference	**0.31°**	**-0.25°**	**-0.32°**
Standard Deviations of Differences (SD)	1.19°	1.99°	2.48°
Standard Error of Differences (SE)	0.16°	0.28°	0.34°
Significant Difference (p=0.05)	No	No	No
Repeatability Coefficient (RC)	**2.34°**	**3.91°**	**4.86°**
Typical Error (TE)	**0.85°**	**1.42°**	**1.77°**
Lower Confidence Limits of TE	*0.73°*	*1.22°*	*1.52°*
Upper Confidence Limits of TE	*1.02°*	*1.70°*	*2.13°*

Table 5.2. Intra-rater Reliability of the No Marker Approach to the Uillinn Method© (the Knee)

5.4.4 Uillinn Method© (Parallel Approach) at the Knee.

The Parallel approach to knee measurements using the UM had an average value of 82° which set the limits of SB at 2° (examiner 3). The average difference was well below this figure for all examiners. The RC ranged from 2.5° (examiner 1)

to 3.3°(examiner 3), while TE was seen at between 0.77° (examiner 1) to 1.18° (examiner 3). The ICC (3,1) was 0.99.

Knees	Examiner 1	Examiner 2	Examiner 3
Average Difference	**0.06°**	**-0.28°**	**0.07°**
Standard Deviations of Differences (SD)	1.26°	1.54°	1.67°
Standard Error of Differences (SE)	0.17°	0.21°	0.23°
Significant Difference (p=0.05)	No	No	No
Repeatability Coefficient (RC)	**2.48°**	**3.03°**	**3.27°**
Typical Error (TE)	**0.90°**	**1.10°**	**1.18°**
Lower Confidence Limits of TE	*0.77°*	*0.95°*	*1.01°*
Upper Confidence Limits of TE	*1.08°*	*1.33°*	*1.42°*

Table 5.3. Intra-rater Reliability of the Parallel Approach to the Uillinn Method© (the Knee)

5.4.5 Uillinn Method© (Mid-Point Approach) at the Knee.

The limits of SB was set at 2° (average measured value = 82°). The average difference was again well below this figure at 0.27° (examiner 1). The RC ranged from 2.08° (examiner 3) to 3.5° (examiner 2) while the TE ranged between 0.76° (examiner 3) to 1.27° (examiner 2). The ICC (3,1) was no lower the 0.99 for each of the repeated measurements.

Knee	Examiner 1	Examiner 2	Examiner 3
Average Difference	**0.27°**	**0.11°**	**0.07°**
Standard Deviations of Differences (SD)	1.4°	1.78°	1.06°
Standard Error of Differences (SE)	0.19°	0.25°	0.14°
Significant Difference (p=0.05)	No	No	No
Repeatability Coefficient (RC)	**2.74°**	**3.5°**	**2.08°**
Typical Error (TE)	**0.99°**	**1.27°**	**0.76°**
Lower Confidence Limits of TE	*0.85°*	*1.09°*	*0.65°*
Upper Confidence Limits of TE	*1.19°*	*1.53°*	*0.91°*

Table 5.4. Intra-rater Reliability of the Mid-Point Approach to the Uillinn Method© (the Knee)

5.4.6 Uillinn Method© (3-Marker Approach) at the Knees.

The average measured knee value was 83°, which set the limits of SB at 2°. The average difference between the repeated measurements was not greater than 0.27° (examiner 1) which signified that SB was not present. The RC results for the examiners ranged from 2.08° (examiner 3) to 3.7° (examiner 2). The TE calculations

ranged from between 1.06° (examiner 3) to 1.35° (examiner 2). The ICC (3,1) was no lower the 0.99 for each of the repeated measurements.

Knee	Examiner 1	Examiner 2	Examiner 3
Average Difference	**-0.27°**	**0.2°**	**0.07°**
Standard Deviations of Differences (SD)	1.51°	1.88°	1.06°
Standard Error of Differences (SE)	0.21°	0.26°	0.14°
Significant Difference (p=0.05)	No	No	No
Repeatability Coefficient (RC)	**2.97°**	**3.7°**	**2.08°**
Typical Error (TE)	**1.07°**	**1.35°**	**0.76°**
Lower Confidence Limits of TE	0.92°	1.16°	0.65°
Upper Confidence Limits of TE	1.29°	1.62°	0.91°

Table 5.5. Intra-rater Reliability: Uillinn Method©, 3-Marker Approach (the Knee)

5.4.7 Uillinn Method© (4-Marker Approach) at the Knee.

The limits of SB for this investigation was set at 2° (mean measured data 82°), and comparisons with the mean difference between the repeated measurements indicated no SB in the data (see Table 5.6). The RC results for the examiners ranged from 2.95° (examiner 1) to 4.43° (examiner 2), while TE ranged from 1.06° (examiner 1) to 1.62° (examiner 3). The ICC (3,1) was 0.99 for each of the examiners.

Knee	Examiner 1	Examiner 2	Examiner 3
Average Difference	**-0.28°**	**-0.06°**	**-0.06°**
Standard Deviations of Differences (SD)	1.5°	2.13°	2.26°
Standard Error of Differences (SE)	0.21°	0.29°	0.31°
Significant Difference (p=0.05)	No	No	No
Repeatability Coefficient (RC)	**2.95°**	**4.18°**	**4.43°**
Typical Error (TE)	**1.06°**	**1.52°**	**1.62°**
Lower Confidence Limits of TE	0.91°	1.31°	1.39°
Upper Confidence Limits of TE	1.27°	1.82°	1.94°

Table 5.6. Intra-rater Reliability of the 4-Marker Approach to the Uillinn Method© (the Knee)

Intra-rater Reliability at the Elbow

5.4.8 Universal Goniometer at the Elbow

At the elbow the mean difference between the examiners repeated assessments was approximately 78° which set the limits of SB at 1.9°. The average difference between the repeated measurements for all the examiners were well below this limit (see Table 5.7) which implied that there was no SB associated with repeated assessments of the elbow. The RC ranged from 5° (examiner 1) to 8.6° (examiner 3), while TE for this measurement ranged between 1.8° (examiner 1) to 3.14° (examiner 3). The ICC (3,1) between each of the repeated data sets was 0.99.

Elbows	Examiner 1	Examiner 2	Examiner 3
Average Difference	**-0.18°**	**-0.20°**	**-0.28°**
Standard Deviations of Differences (SD)	2.57°	3.55°	4.4°
Standard Error of Differences (SE)	0.36°	0.50°	0.62°
Significant Difference (p=0.05)	No	No	No
Repeatability Coefficient (RC)	**5.04°**	**6.96°**	**8.62°**
Typical Error (TE)	**1.83°**	**2.46°**	**3.14°**
Lower Confidence Limits of TE	*1.57°*	*2.10°*	*2.69°*
Upper Confidence Limits of TE	*2.21°*	*2.96°*	*3.79°*

Table 5.7. Intra-rater Reliability of the Universal Goniometer (the Elbow)

5.4.9 Uillinn Method© (No Marker Approach) at the Elbow.

The average elbow measurement of the No Marker approach to the UM was 84° (examiner 1) which set the limits of SB at 2.1°. The average difference was less than this figure at approximately 0.5° (examiner 1). The RC ranged between 4° (examiner 1) to 5.78° (examiner 2), while the TE between assessments of the elbow using this approach ranged between 1.64° (examiner 1) to 2.26° (examiner 2). The ICC (3,1) for each of the examiners was 0.99.

Elbow	Examiner 1	Examiner 2	Examiner 3
Average Difference	**-0.49°**	**0.31°**	**0.13°**
Standard Deviations of Differences (SD)	2.05°	2.95°	2.35°
Standard Error of Differences (SE)	0.29°	0.42°	0.33°
Significant Difference (p=0.05)	No	No	No
Repeatability Coefficient (RC)	**4.02°**	**5.78°**	**4.61°**
Typical Error (TE)	**1.64°**	**2.26°**	**1.88°**
Lower Confidence Limits of TE	*1.40°*	*1.94°*	*1.61°*
Upper Confidence Limits of TE	*1.98°*	*2.73°*	*2.27°*

Table 5.8. Intra-rater Reliability of the No Marker Approach to the Uillinn Method© (the Elbow)

5.4.10 Uillinn Method© (Parallel Approach) at the Elbow.

Based on the average measured value of 80° there was no SB present in the data (SB limit = 2°). The RC ranged from 3.7° (examiner 1) to 4.5° (examiner 2) while TE ranged between 1.62° (examiner 1) to 1.9° (examiner 3). The ICC (3,1) was 0.99.

Elbow	Examiner 1	Examiner 2	Examiner 3
Average Difference	**-0.39°**	**0.03°**	**-0.13°**
Standard Deviations of Differences (SD)	1.88°	2.25°	2.1°
Standard Error of Differences (SE)	0.27°	0.32°	0.3°
Significant Difference (p=0.05)	No	No	No
Repeatability Coefficient (RC)	**3.7°**	**4.41°**	**4.12°**
Typical Error (TE)	**1.62°**	**1.83°**	**1.89°**
Lower Confidence Limits of TE	*1.38°*	*1.57°*	*1.62°*
Upper Confidence Limits of TE	*1.95°*	*2.21°*	*2.28°*

Table 5.9. Intra-rater Reliability of the Parallel Approach to the Uillinn Method© (the Elbow)

5.4.11 Uillinn Method© (Mid-Point Approach) at the Elbow.

The average measured value for the Mid-point approach at the elbow was 71° which set the limits of SB at 1.7°. This implied that no SB was present in the repeated measurements. The results for the RC analysis ranged from 3.27° (examiner 1) to 3.88° (examiner 2). The TE calculations ranged from between 1.34° (examiner 1) to 1.98° (examiner 3). The ICC (3,1) was no lower the 0.99 for each of the repeated measurements.

Elbow	Examiner 1	Examiner 2	Examiner 3
Average Difference	**-0.47°**	**-0.04°**	**0.31°**
Standard Deviations of Differences (SD)	1.67°	1.98°	1.73°
Standard Error of Differences (SE)	0.24°	0.28°	0.24°
Significant Difference (p=0.05)	No	No	No
Repeatability Coefficient (RC)	**3.27°**	**3.88°**	**3.4°**
Typical Error (TE)	**1.34°**	**1.63°**	**1.98°**
Lower Confidence Limits of TE	*1.15°*	*1.40°*	*1.70°*
Upper Confidence Limits of TE	*1.61°*	*1.97°*	*2.39°*

Table 5.10. Intra-rater Reliability of the Mid-Point Approach to the Uillinn Method© (the Elbow)

5.4.12 Uillinn Method© (3-Marker Approach) at the Elbow.

The Batterham and George (2000) limits of SB of 2.5% of the average measured value indicated no SB in the data (average measured value = 78°; SB limit = 1.9°). The RC results for the examiners ranged from 3.4° (examiner 3) to 4.62° (examiner 1), the TE ranged from 1.63° (examiner 2) to 1.98° (examiner 3). The ICC (3,1) was no lower the 0.99 for each of the repeated measurements.

Elbow	Examiner 1	Examiner 2	Examiner 3
Average Difference	**-0.92°**	**-0.04°**	**0.31°**
Standard Deviations of Differences (SD)	2.35°	1.98°	1.73°
Standard Error of Differences (SE)	0.33°	0.28°	0.24°
Significant Difference (p=0.05)	No	No	No
Repeatability Coefficient (RC)	**4.62°**	**3.88°**	**3.4°**
Typical Error (TE)	**1.97°**	**1.63°**	**1.98°**
Lower Confidence Limits of TE	1.69°	1.40°	1.70°
Upper Confidence Limits of TE	2.38°	1.97°	2.39°

Table 5.11. Intra-rater Reliability of the 3-Marker Approach to the Uillinn Method© (the Elbow)

5.4.13 Uillinn Method© (4-Marker Approach) at the Elbow.

The limits of SB was set at 2° (mean measured data 83°) which indicated no SB was evident in the data (see Table 5.12). The RC results for the examiners ranged from 4.51° (examiner 1) to 5.69° (examiner 2) and the value of TE ranged from 1.95° (examiner 1) to 2.42° (examiner 2). The ICC (3,1) was 0.99 for each of the pair measurements.

Elbow	Examiner 1	Examiner 2	Examiner 3
Average Difference	**-0.93°**	**-0.18°**	**0.07°**
Standard Deviations of Differences (SD)	2.3°	2.88°	2.58°
Standard Error of Differences (SE)	0.33°	0.41°	0.37°
Significant Difference (p=0.05)	No	No	No
Repeatability Coefficient (RC)	**4.51°**	**5.69°**	**5.06°**
Typical Error (TE)	**1.95°**	**2.42°**	**2.10°**
Lower Confidence Limits of TE	1.67°	2.07°	1.80°
Upper Confidence Limits of TE	2.35°	2.92°	2.53°

Table 5.12. Intra-rater Reliability of the 4-Marker Approach to the Uillinn Method© (the Elbow)

5.4.14 Inter-Rater Reliability

The experimental procedure yielded two sets of measurement data, one from the first round and one from the second round of measurements. During the procedures it was recognised that some subjects had moved to some degree, especially during the elbow measurements. It was felt that more movement had occurred during the first round of measurements rather than the second round. For this reason analysis for the inter-rater reliability study were based on comparisons between the second round of joint angle measurements. The results are discussed by comparisons between each of the 3 examiners.

The limits of SB for each of the joint measurement techniques would be identical to those already worked out in intra-rater reliability section above. Therefore, differentiation between random and non-random error would be based on the limits discussed above. Supplementary significance testing is presented in the accompanying tables related to each question. These tests were calculated using a paired t-test ($p=0.05$).

Based on the calculations non-random error was certainly present in at least some of the data. Therefore, it was felt it would be safer to use the 95% limits of agreement (LOA) test to represent the data rather than the RC which can not account for the SB. Even when there was no SB in a particular dataset the LOA test was kept in order to allow comparisons between the difference techniques.

This section will present the results for each joint separately, starting with the knee and lastly reporting the elbow.

Inter-rater Reliability at the Knee

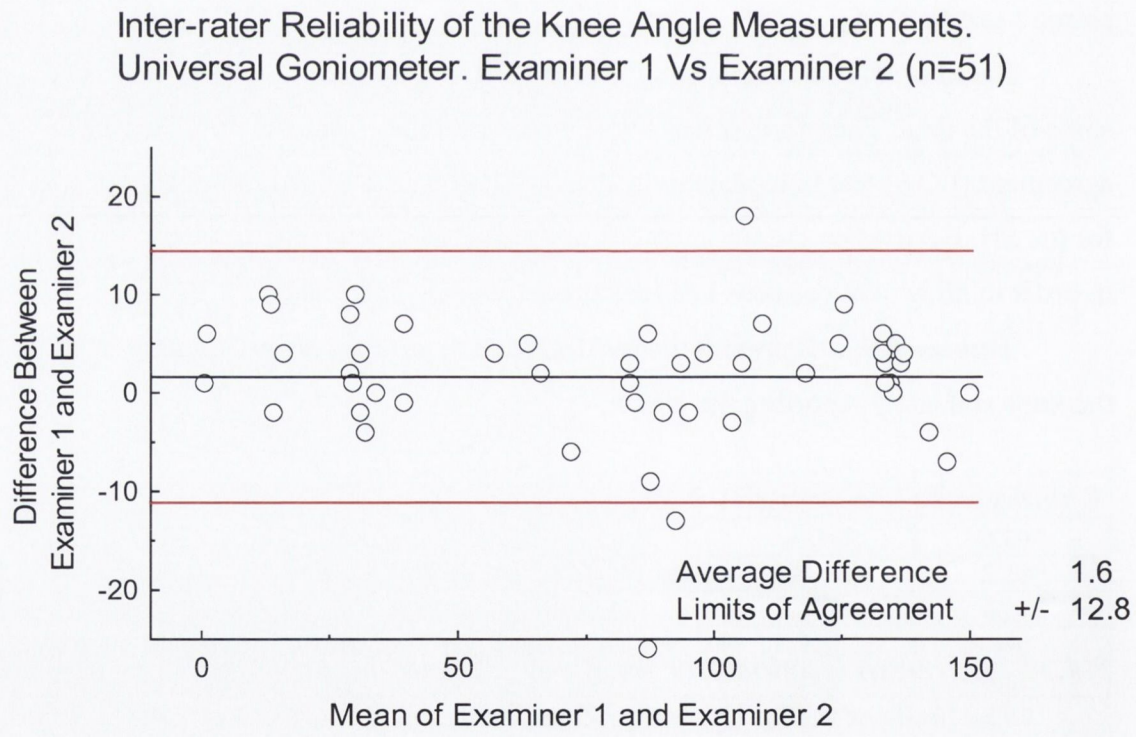
5.4.15 Universal Goniometer for Knee Measurements

The limits of SB for UG measurement at the knee was 2°. While the average difference between any two examiners did not quite reach this designated limit, it was very close to it (Examiner 2 Vs Examiner 3) (see Table 5.13). The LOA ranged from at best +7°, -7.55° (Examiner 1 Vs Examiner 3) to at worst +14.3°, -11° (Examiner 1 Vs Examiner 2) (see Graphs 5.1, 5.2, 5.3). Typical Error (TE) between the examiners ranged from 2.59° (Examiner 1 Vs Examiner 3) to 4.58° (Examiner 1

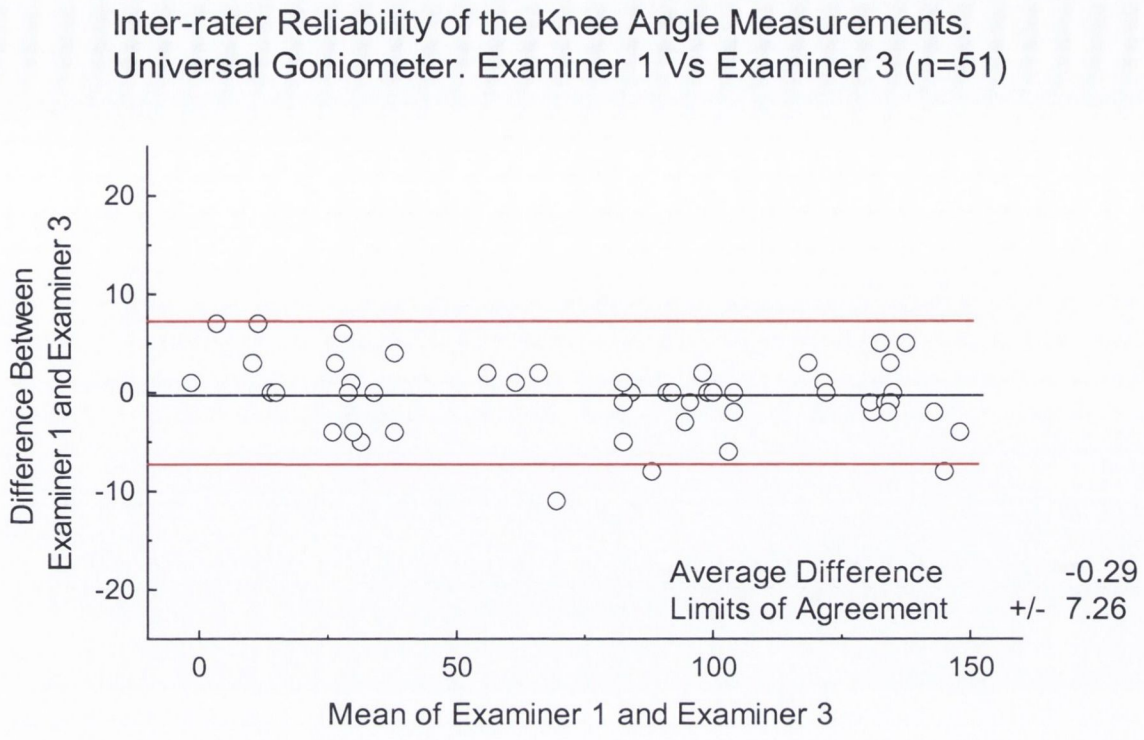
Vs Examiner 2) (see Table 5.13). The total TE (multi-variant) was calculated as 4° and the ICC (3,1) for this measurement was calculated as 0.89.

Knee	Ex 1 Vs Ex 2	Ex 2 Vs Ex 3	Ex 1 Vs Ex 3
Average Difference	**1.6°**	**-1.9°**	**-0.29°**
Above Systematic Bias (SB) Limit (2°)	No	No	No
Standard Deviations of Differences (SD)	6.53°	5.88°	3.7°
Standard Error of Differences (SE)	0.91°	0.82°	0.51°
Significant Difference (p=0.05)	No	No	No
SD * 1.96	12.81°	11.52°	7.26°
95% Limits of Agreement (LOA) - Upper	**14.3°**	**9.62°**	**6.96°**
95% Limits of Agreement (LOA) - Lower	**-11°**	**-13.4°**	**-7.55°**
Typical Error (TE)	**4.58°**	**4.12°**	**2.59°**
Lower Confidence Limits of TE	*3.95°*	*3.55°*	*2.24°*
Upper Confidence Limits of TE	*5.48°*	*4.93°*	*3.11°*

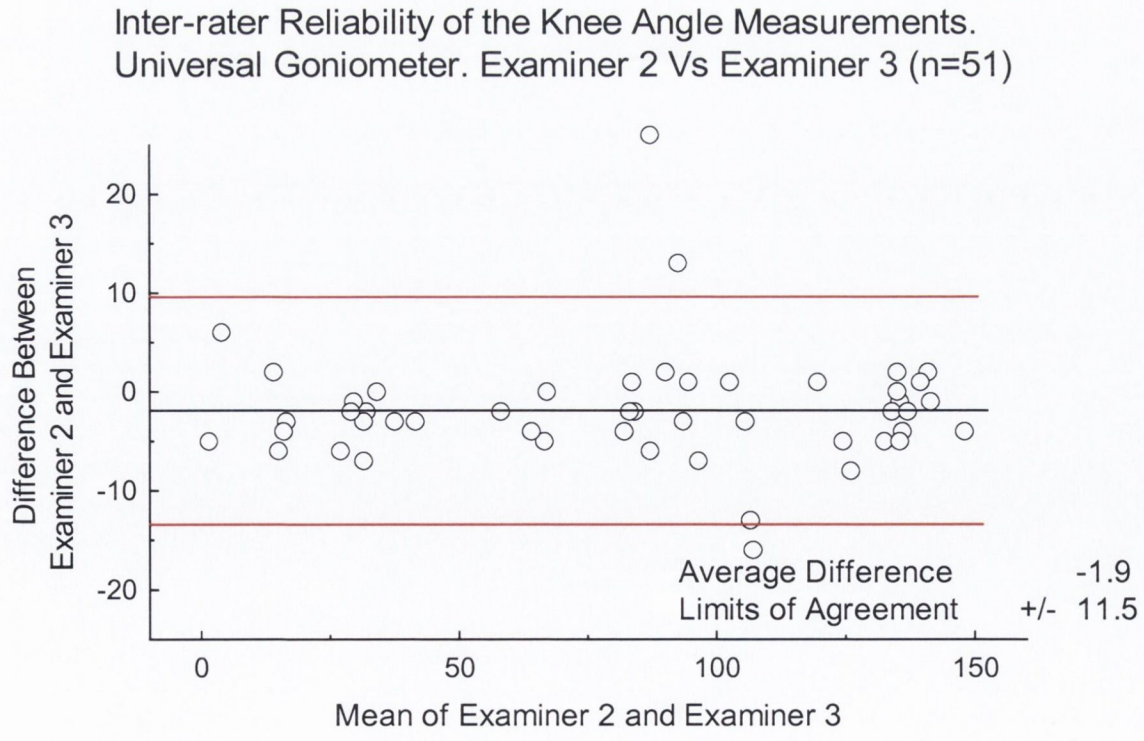
Table 5.13. Inter-rater Reliability of the Universal Goniometer (the Knee)



Graph 5.1. 95% Limits of Agreement Scatter Plot.



Graph 5.2. 95% Limits of Agreement Scatter Plot.



Graph 5.3. 95% Limits of Agreement Scatter Plot.

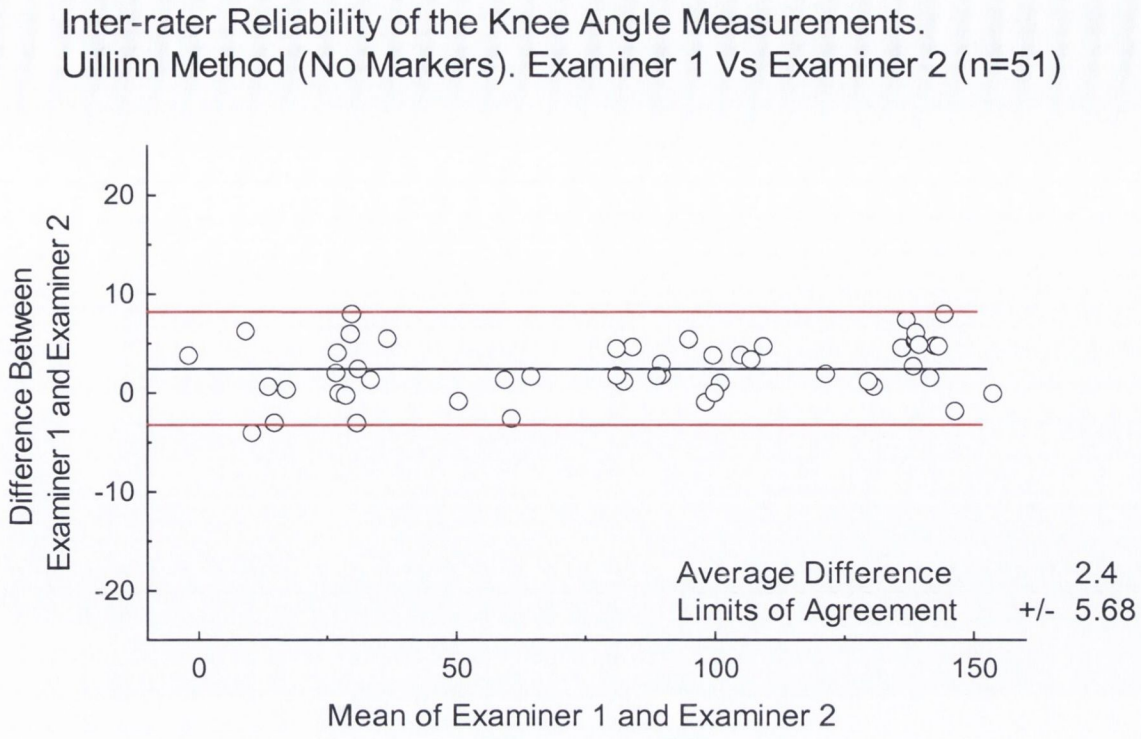
5.4.16 Uillinn Method (No Markers) for Knee Measurements

The No Marker approach to the application of the UM at the knee had a limit of SB of 2.1°. This limit was exceeded by Examiner 2's comparison to both of the other Examiners. To interpret the results presented in Table 5.14, Examiner 2 had consistently higher results than Examiner 1, and Examiner 3 had consistently lower results than Examiner 2. Reading the negative and positive values in the 'Average Difference' figures in the first 2 columns (see Table 5.14) would place Examiner 1 and Examiner 3 as having little difference between them, this is indeed reflected in the average difference in column 3.

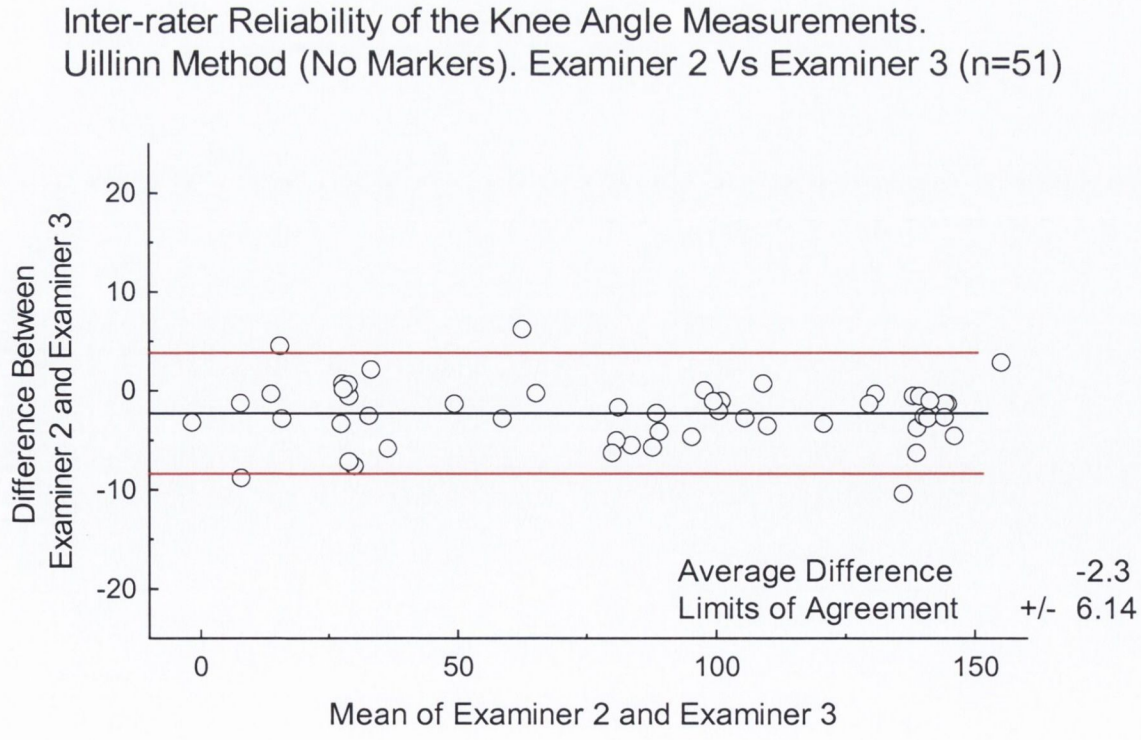
Random error as measured by TE was comparable between all the examiners ranged from 1.8° to 2.2°. The TE (multi-variant) across all examiners was 2.44°. The ICC (3,1) between all of the 3 datasets was calculated as 0.97.

Knee	Ex 1 Vs Ex 2	Ex 2 Vs Ex 3	Ex 1 Vs Ex 3
Average Difference	**2.48°**	**-2.31°**	**0.16°**
Above Systematic Bias (SB) Limit (2.1°)	Yes	Yes	No
Standard Deviations of Differences (SD)	2.9°	3.13°	2.56°
Standard Error of Differences (SE)	0.4°	0.44°	0.35°
Significant Difference (p=0.05)	Yes	No	Yes
SD * 1.96	5.69°	6.14°	5°
95% Limits of Agreement (LOA) - Upper	**8.16°**	**3.82°**	**5.19°**
95% Limits of Agreement (LOA) - Lower	**-3.2°**	**-8.46°**	**-4.85°**
Typical Error (TE)	**2.05°**	**2.21°**	**1.79°**
Lower Confidence Limits of TE	*1.76°*	*1.90°*	*1.54°*
Upper Confidence Limits of TE	*2.45°*	*2.64°*	*2.15°*

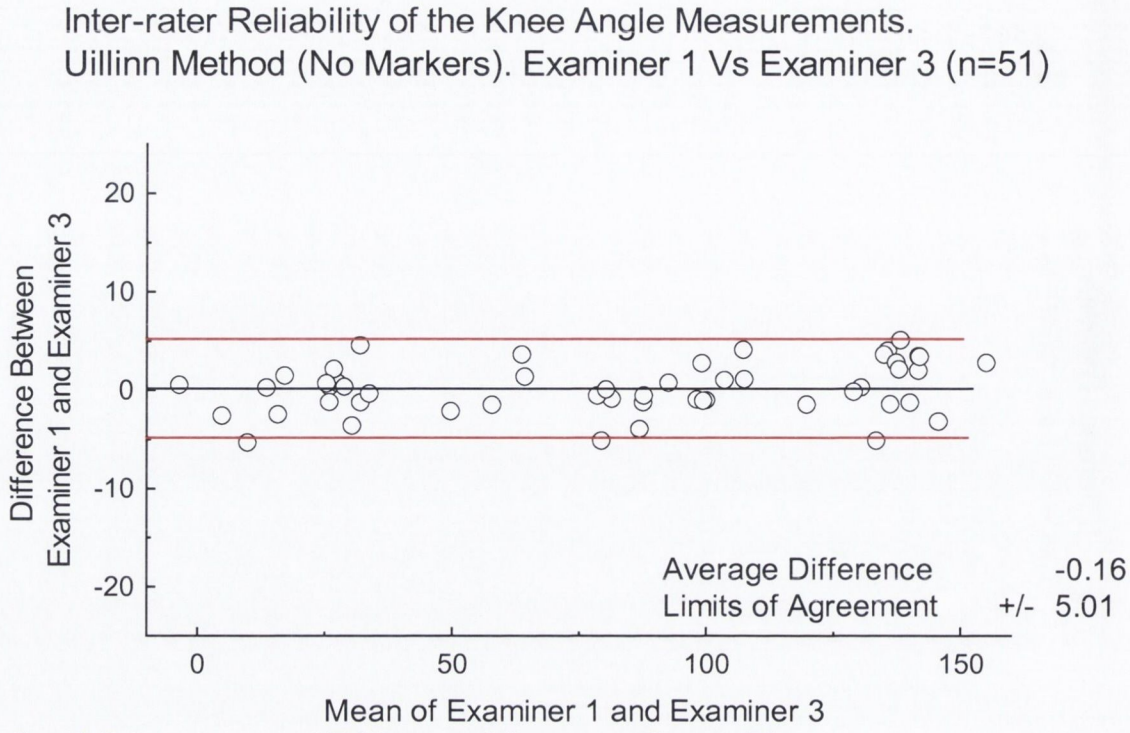
Table 5.14. Inter-rater Reliability of the Uillinn Method (No Marker) (the Knee)



Graph 5.4. *95% Limits of Agreement Scatter Plot.*



Graph 5.5. *95% Limits of Agreement Scatter Plot.*



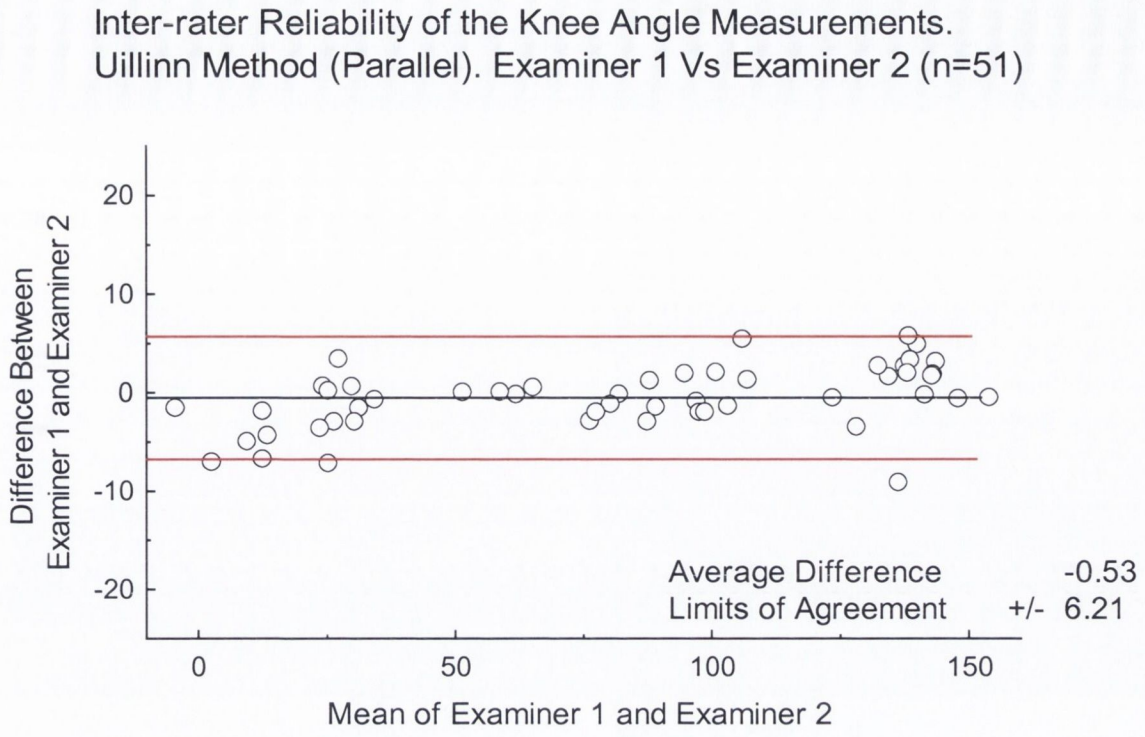
Graph 5.6. *95% Limits of Agreement Scatter Plot.*

5.4.17 Uillinn Method (Parallel) for Knee Measurements

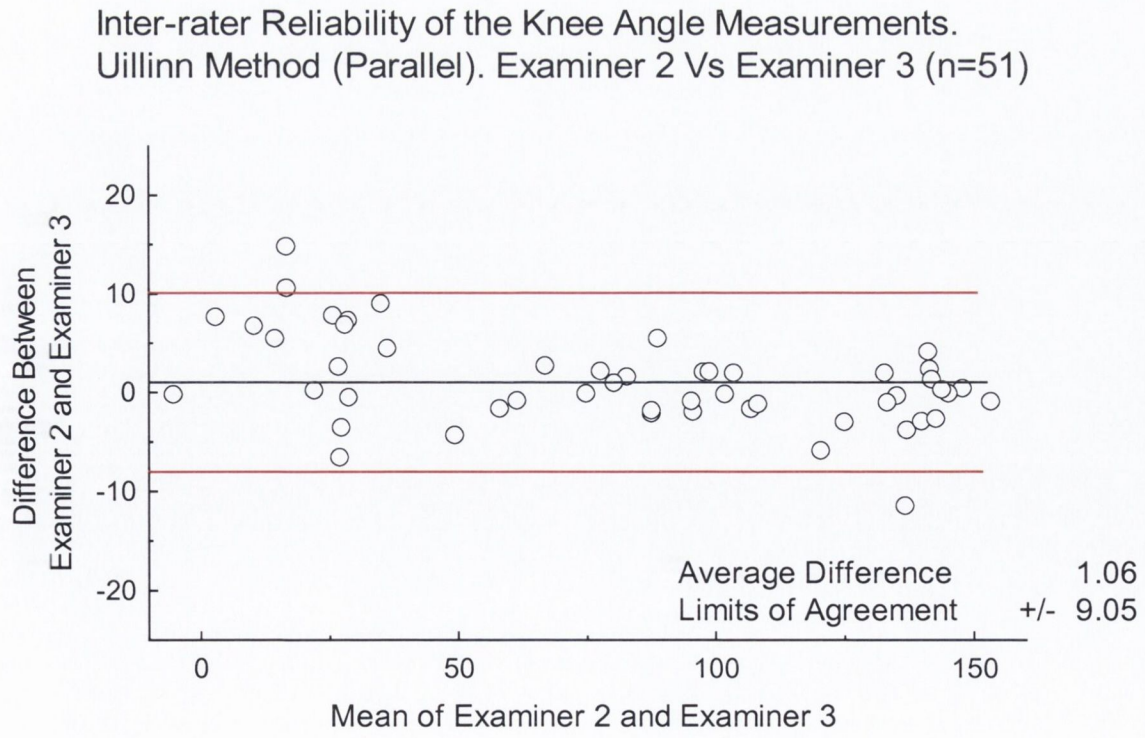
There were relatively low bias' seen between the examiner for the Parallel approach to the measurement of the knee. Comparisons between the different examiners were all below the 2° limit (see Table 5.15). TE ranged between 2.2° (Examiner 1 Vs Examiner 2) to 3.2° (Examiner 2 Vs Examiner 3) and was 2.8° between all examiners (TE (multi-variant)). The LOA graphs (see Graph 5.7, 5.8, 5.9) show best agreement between Examiner 1 and Examiner 2, and worst agreement between Examiner 2 and Examiner 3. The ICC (3,1) was 0.95.

Knee	Ex 1 Vs Ex 2	Ex 2 Vs Ex 3	Ex 1 Vs Ex 3
Average Difference	**-0.53°**	**1.06°**	**0.53°**
Above Systematic Bias (SB) Limit (2°)	No	No	No
Standard Deviations of Differences (SD)	3.17°	4.61°	3.86°
Standard Error of Differences (SE)	0.44°	0.64°	0.54°
Significant Difference (p=0.05)	No	No	No
SD * 1.96	6.21°	9°	7.5°
95% Limits of Agreement (LOA) - Upper	**5.68°**	**10.1°**	**8.1°**
95% Limits of Agreement (LOA) - Lower	**-6.7°**	**-7.9°**	**-7°**
Typical Error (TE)	**2.22°**	**3.24°**	**2.71°**
Lower Confidence Limits of TE	*1.91°*	*2.79°*	*2.33°*
Upper Confidence Limits of TE	*2.66°*	*3.87°*	*3.24°*

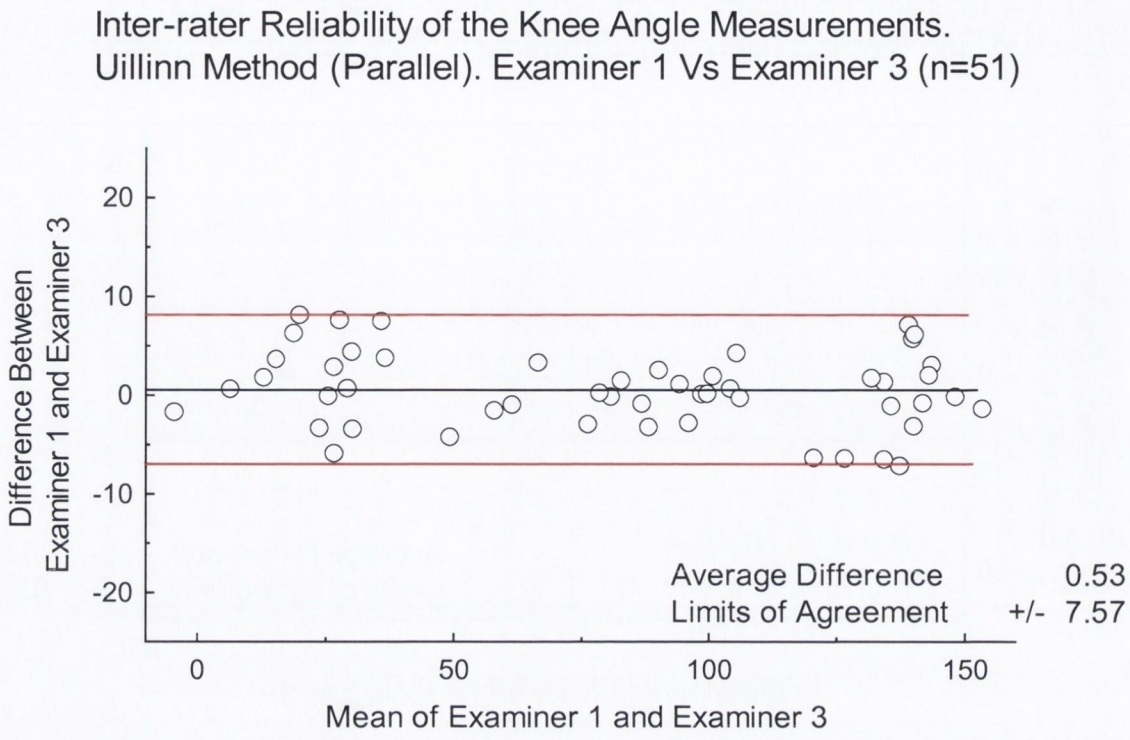
Table 5.15. Inter-rater Reliability of the Uillinn Method (Parallel) (the Knee)



Graph 5.7. *95% Limits of Agreement Scatter Plot.*



Graph 5.8. *95% Limits of Agreement Scatter Plot.*



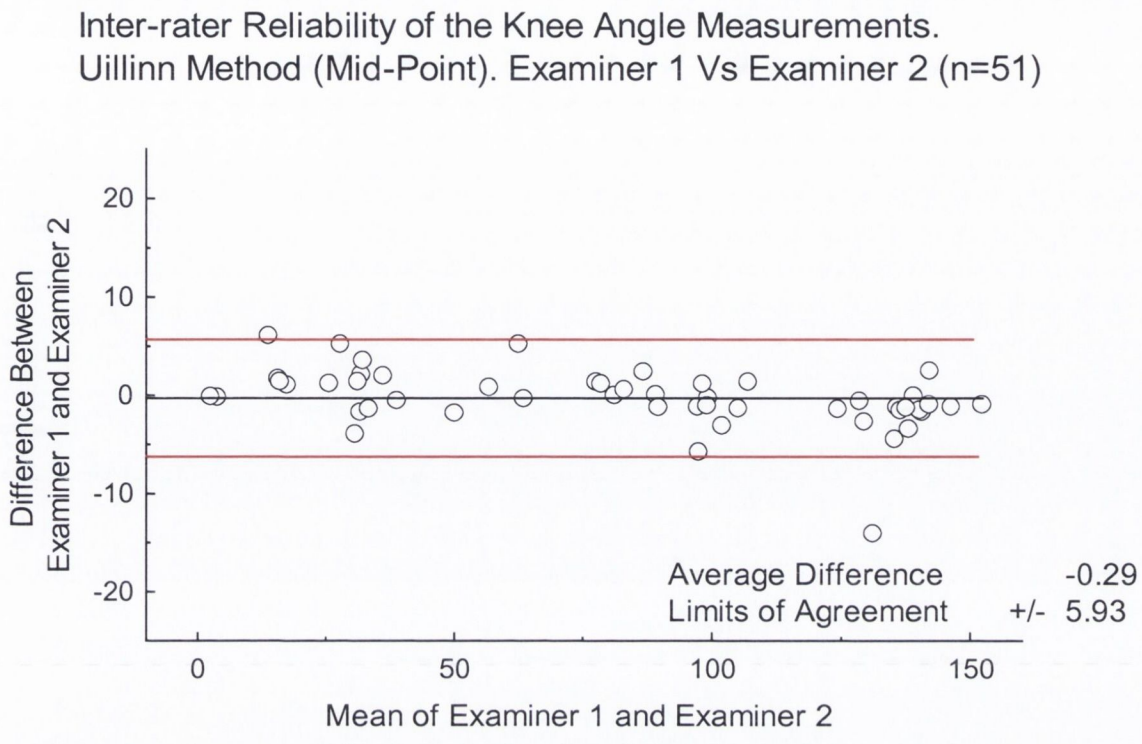
Graph 5.9. *95% Limits of Agreement Scatter Plot.*

5.4.18 Uillinn Method (Mid-Point) for Knee Measurements

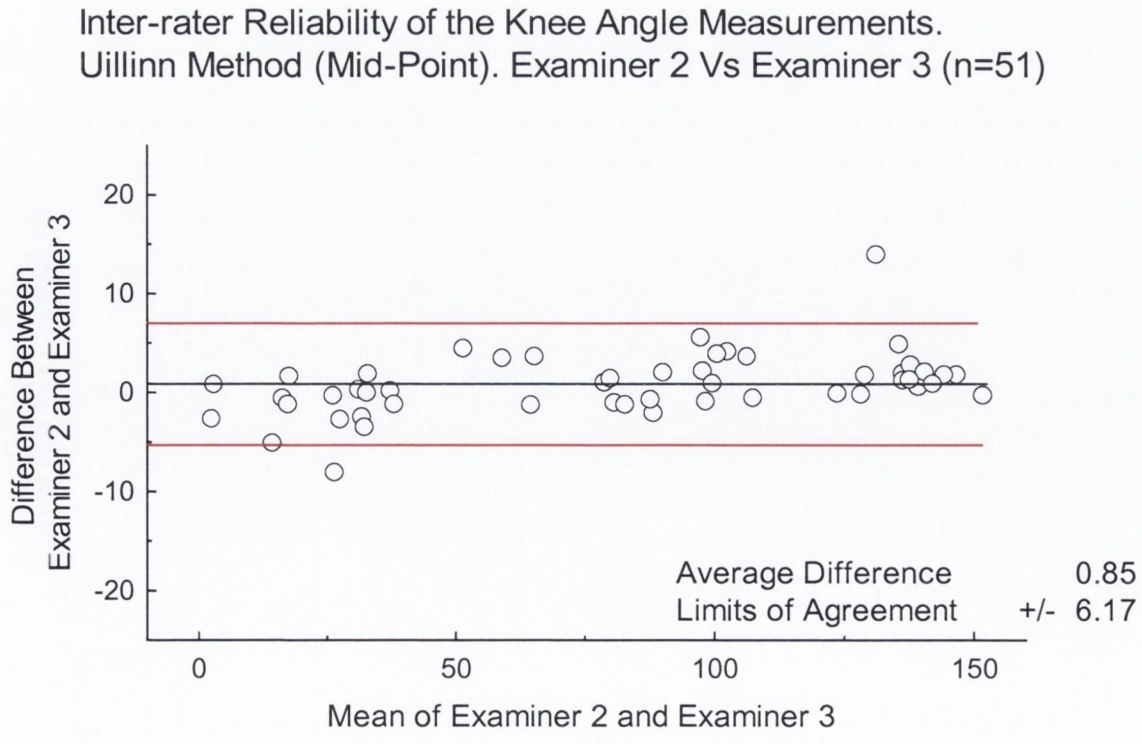
All assessments of non-random error between the examiners for the measurement of the Knee with the Mid-Point approach to the UM were below the SB limit of 2°. The highest TE was 2.21° (Examiner 2 Vs Examiner 3). These examiners had the worst agreement when assessed using the LOA test (see Graph 5.10, 5.11, 5.12), however all graphs looked reasonably comparable. The TE (multi-variant) across all the examiners was 1.99°. ICC (3,1) was calculated as 0.97.

Knee	Ex 1 Vs Ex 2	Ex 2 Vs Ex 3	Ex 1 Vs Ex 3
Average Difference	**-0.29°**	**0.85°**	**0.55°**
Above Systematic Bias (SB) Limit (2°)	No	No	No
Standard Deviations of Differences (SD)	3.03°	3.15°	2°
Standard Error of Differences (SE)	0.42°	0.44°	0.28°
Significant Difference (p=0.05)	No	No	No
SD * 1.96	5.93°	6.17°	3.93°
95% Limits of Agreement (LOA) - Upper	**5.64°**	**7.02°**	**4.48°**
95% Limits of Agreement (LOA) - Lower	**-6.23°**	**-5.34°**	**-3.38°**
Typical Error (TE)	**2.12°**	**2.21°**	**1.41°**
Lower Confidence Limits of TE	*1.83°*	*1.90°*	*1.21°*
Upper Confidence Limits of TE	*2.54°*	*2.64°*	*1.68°*

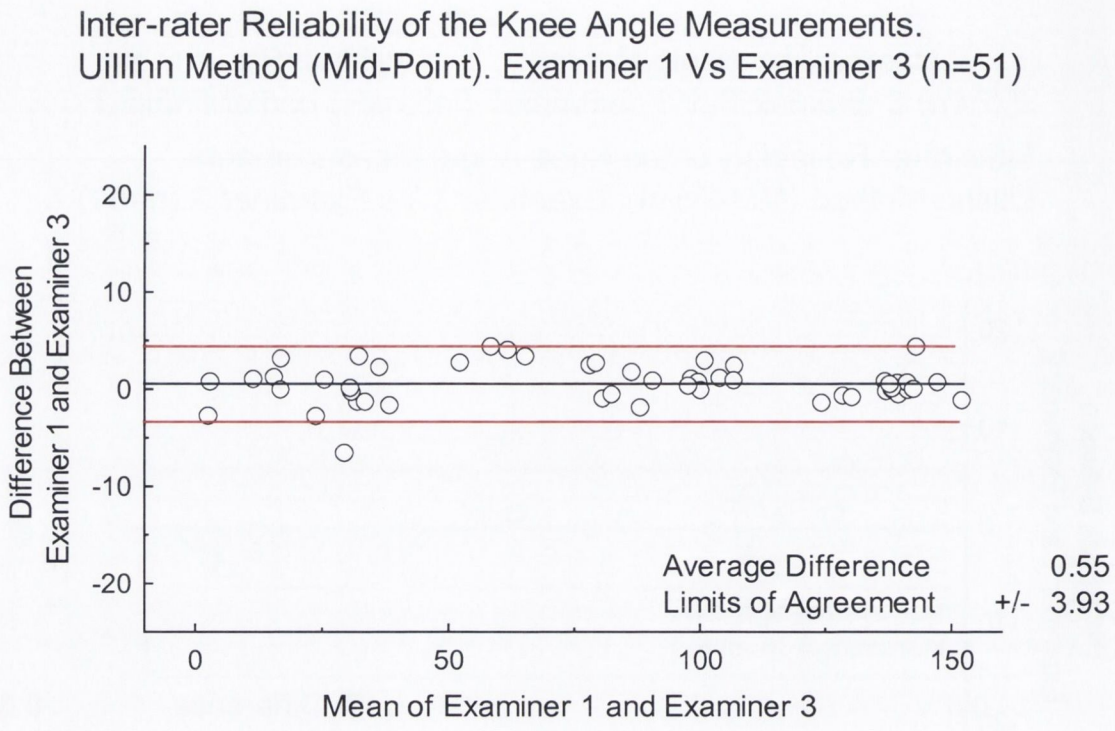
Table 5.16. Inter-rater Reliability of the Uillinn Method (Mid-Point) (the Knee)



Graph 5.10. *95% Limits of Agreement Scatter Plot.*



Graph 5.11. *95% Limits of Agreement Scatter Plot.*



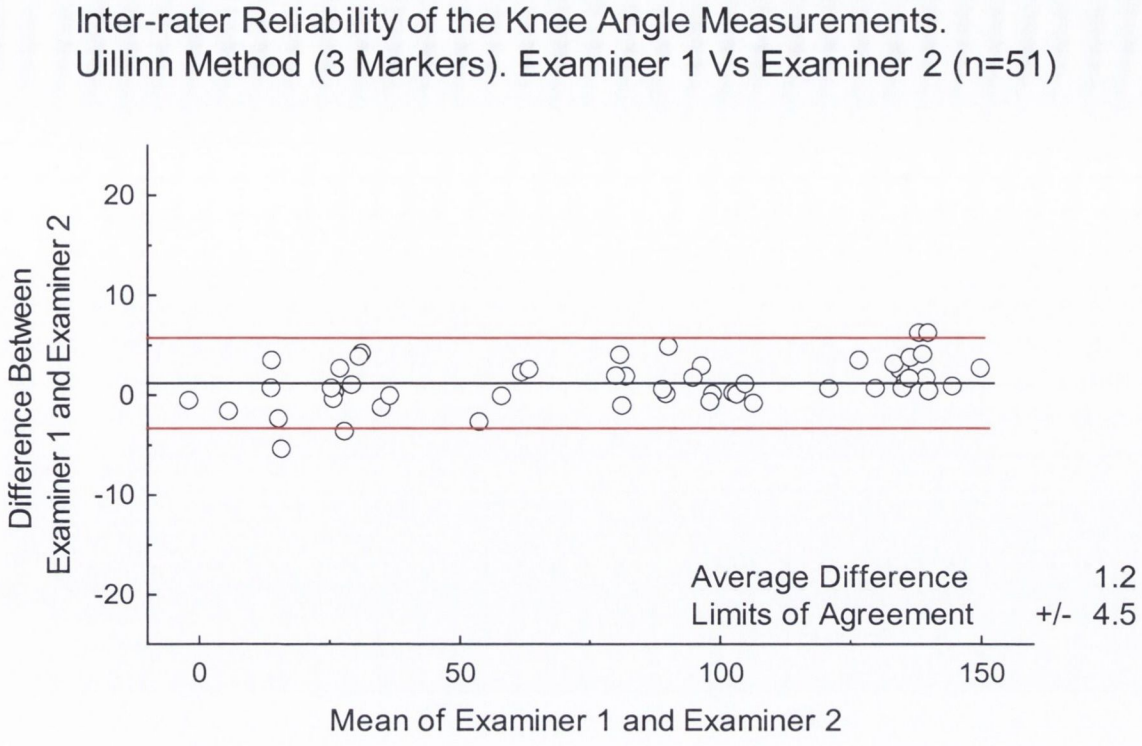
Graph 5.12. *95% Limits of Agreement Scatter Plot.*

5.4.19 Uillinn Method (3-Markers) for Knee Measurements

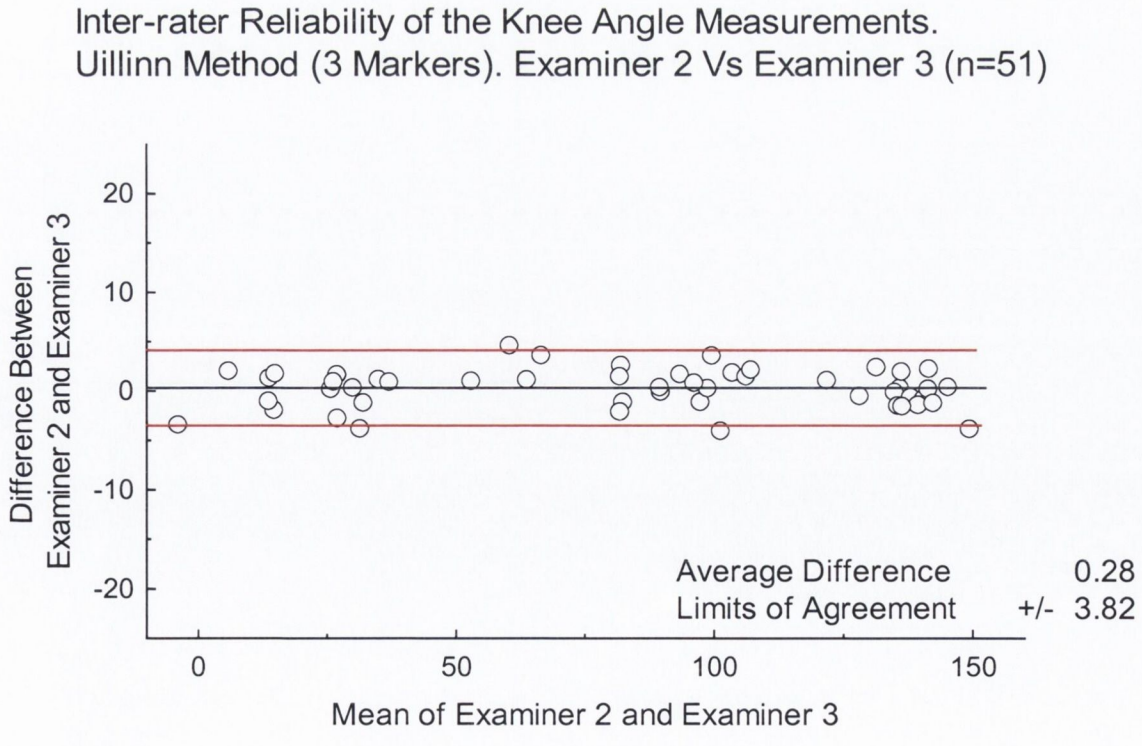
There was least bias in the technique of marker placement between Examiner 2 and Examiner 3 (0.28°), however, comparisons for these examiners with Examiner 1 showed a trend toward SB, without actually reaching the 2° limit. TE (multi-variant) between all examiners was 1.75°, the maximum paired TE was low 1.67° (Examiner 1 Vs Examiner 2) and the LOA scatter plots showed relatively good agreement between the examiners for this approach to the UM of measuring the Knee. The ICC (3,1) was calculated as 0.98 between the examiners measurements.

Knee	Ex 1 Vs Ex 2	Ex 2 Vs Ex 3	Ex 1 Vs Ex 3
Average Difference	**1.22°**	**0.28°**	**1.5°**
Above Systematic Bias (SB) Limit (2°)	No	No	No
Standard Deviations of Differences (SD)	2.32°	1.95°	2.37°
Standard Error of Differences (SE)	0.32°	0.27°	0.33°
Significant Difference (p=0.05)	Yes	Yes	No
SD * 1.96	4.54°	3.82°	4.65°
95% Limits of Agreement (LOA) - Upper	**5.77°**	**4.11°**	**6.16°**
95% Limits of Agreement (LOA) - Lower	**-3.32°**	**-3.53°**	**-3.14°**
Typical Error (TE)	**1.63°**	**1.37°**	**1.67°**
Lower Confidence Limits of TE	*1.40°*	*1.18°*	*1.44°*
Upper Confidence Limits of TE	*1.95°*	*1.64°*	*2.00°*

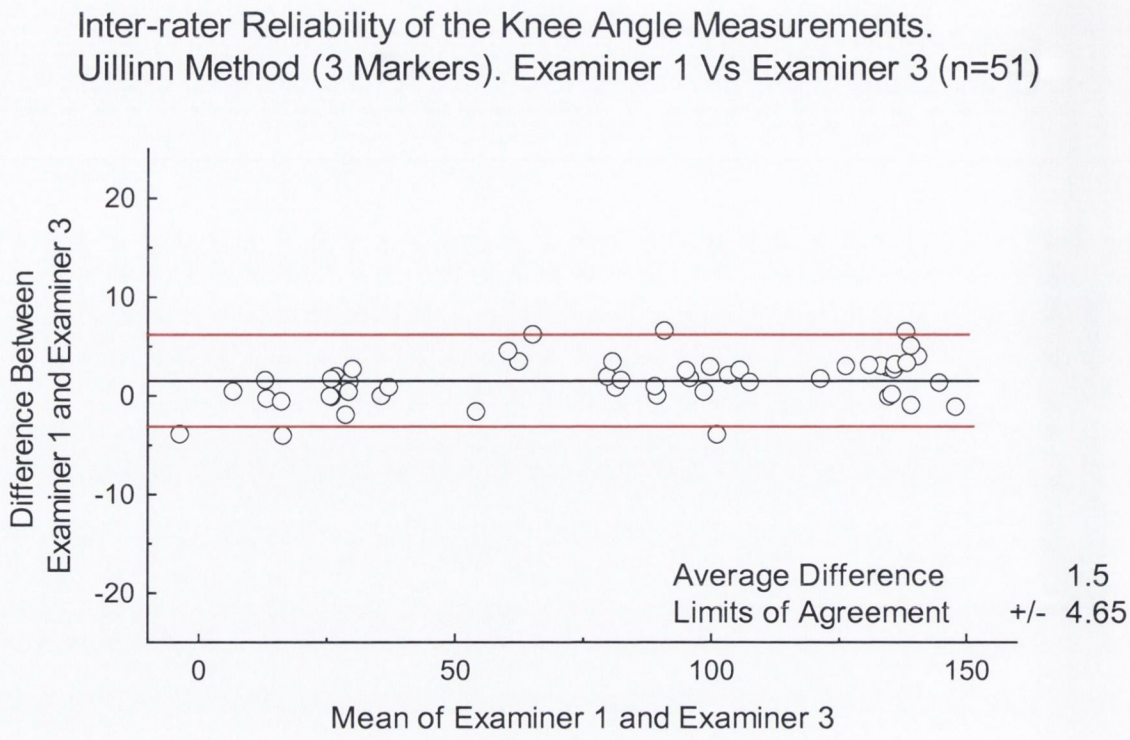
Table 5.17. Inter-rater Reliability of the Uillinn Method (3-Markers) (the Knee)



Graph 5.13. *95% Limits of Agreement Scatter Plot.*



Graph 5.14. *95% Limits of Agreement Scatter Plot.*



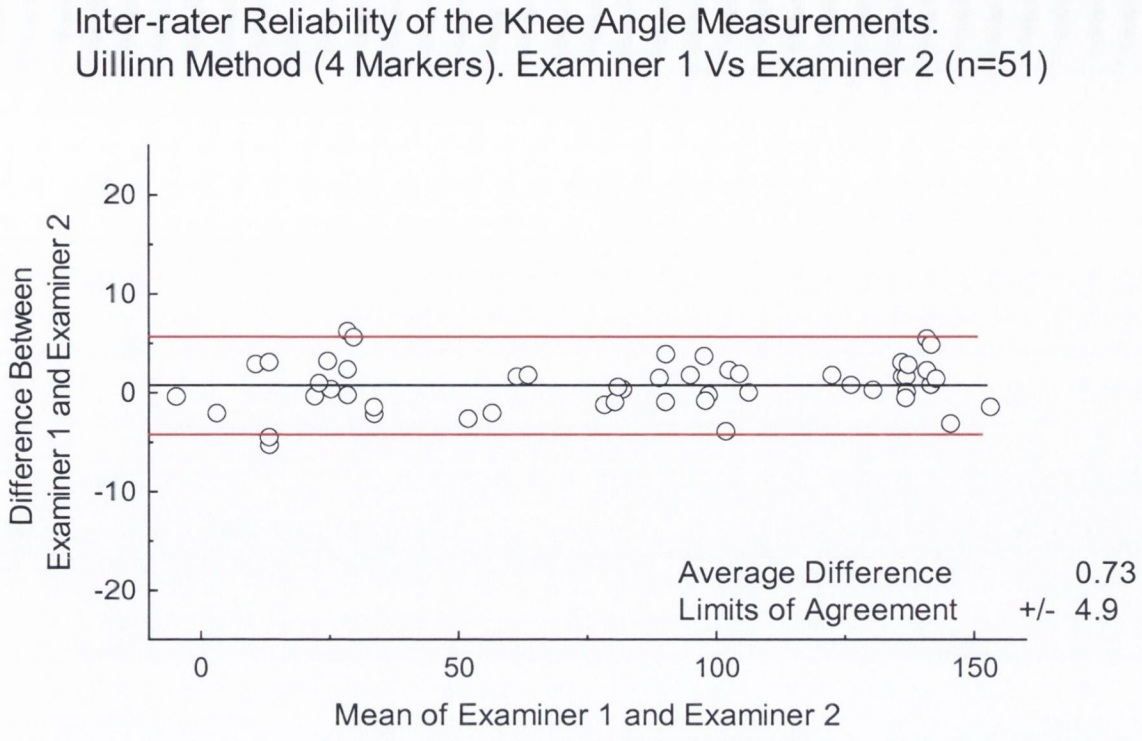
Graph 5.15. *95% Limits of Agreement Scatter Plot.*

5.4.20 Uillinn Method (4-Markers) for Knee Measurements

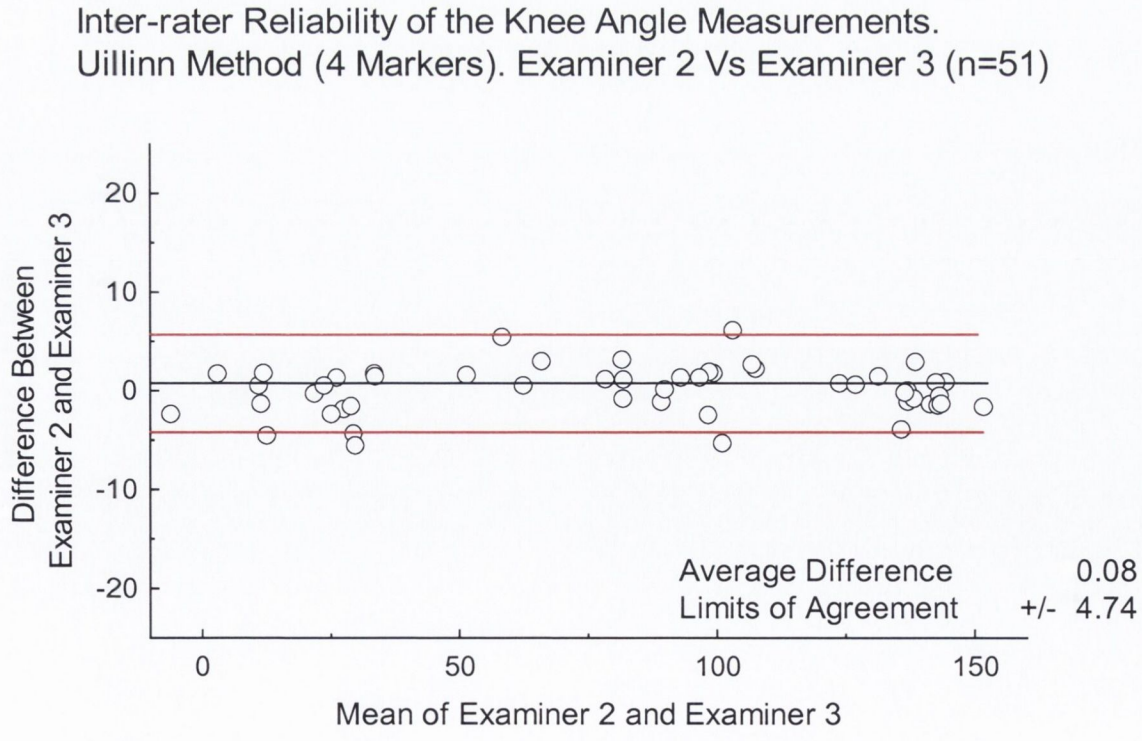
Similar to the 3-Marker approach the 4-Marker approach to the UM's measurement of the knee, the data did not support the presents of SB between the examiners. The highest average difference was 0.81° (Examiner 1 Vs Examiner 3) which was well below the limit of SB of 2°. The TE for pair examiners had a small range from 1.56° to 1.77° and was 1.73° (TE(multi-variant)) for measures between all examiners. All of the LOA scatter plots showed a similar spread (see Graph 5.16, 5.17, 5.19). The ICC (3,1) was 0.98.

Knee	Ex 1 Vs Ex 2	Ex 2 Vs Ex 3	Ex 1 Vs Ex 3
Average Difference	**0.73°**	**0.08°**	**0.81°**
Above Systematic Bias (SB) Limit (2°)	No	No	No
Standard Deviations of Differences (SD)	2.52°	2.41°	2.22°
Standard Error of Differences (SE)	0.35°	0.33°	0.31°
Significant Difference (p=0.05)	Yes	Yes	No
SD * 1.96	4.94°	4.73°	4.35°
95% Limits of Agreement (LOA) - Upper	**5.67°**	**4.82°**	**5.16°**
95% Limits of Agreement (LOA) - Lower	**-4.2°**	**-4.65°**	**-3.53°**
Typical Error (TE)	**1.77°**	**1.69°**	**1.56°**
Lower Confidence Limits of TE	*1.52°*	*1.46°*	*1.34°*
Upper Confidence Limits of TE	*2.11°*	*2.03°*	*1.86°*

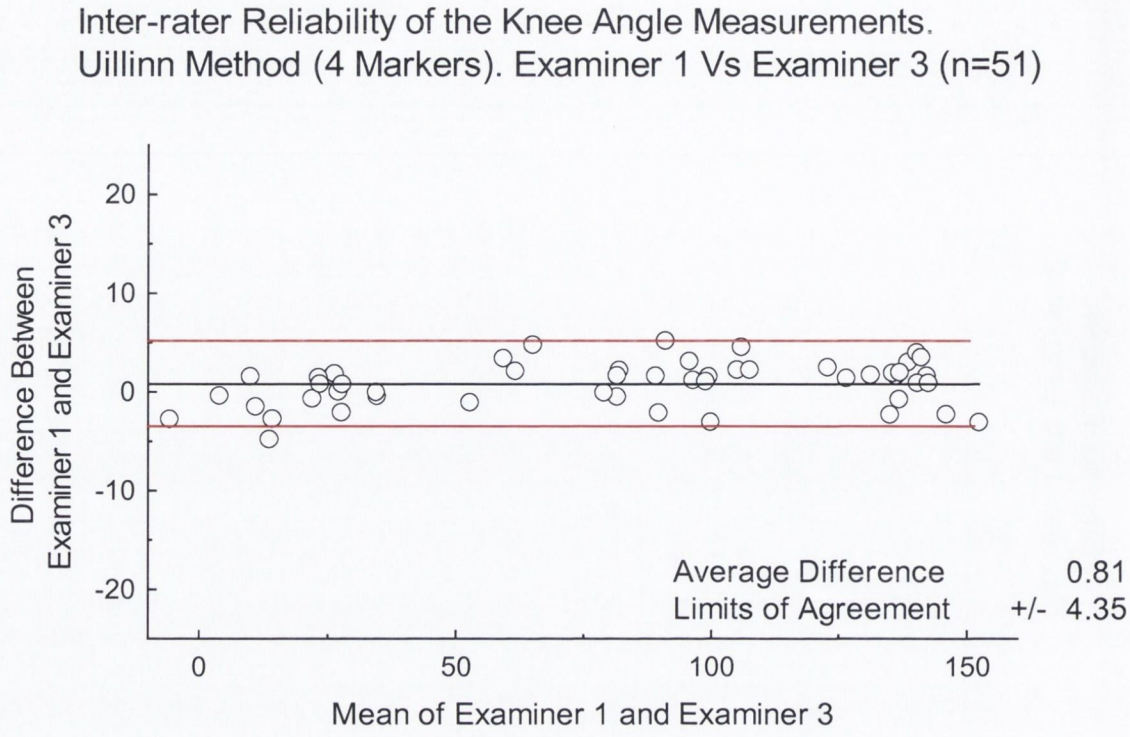
Table 5.19. Inter-rater Reliability of the Uillinn Method (4-Markers) (the Knee)



Graph 5.16. *95% Limits of Agreement Scatter Plot.*



Graph 5.17. *95% Limits of Agreement Scatter Plot.*



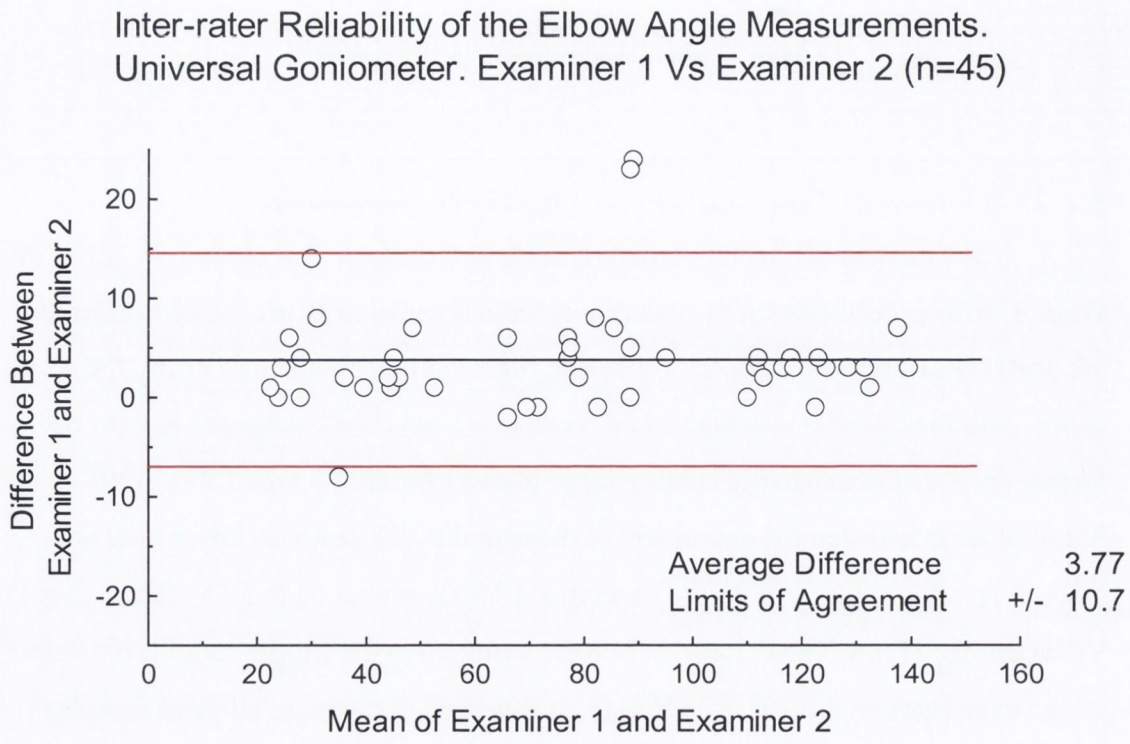
Graph 5.18. 95% Limits of Agreement Scatter Plot.

5.4.21 Universal Goniometer for Elbow Measurements

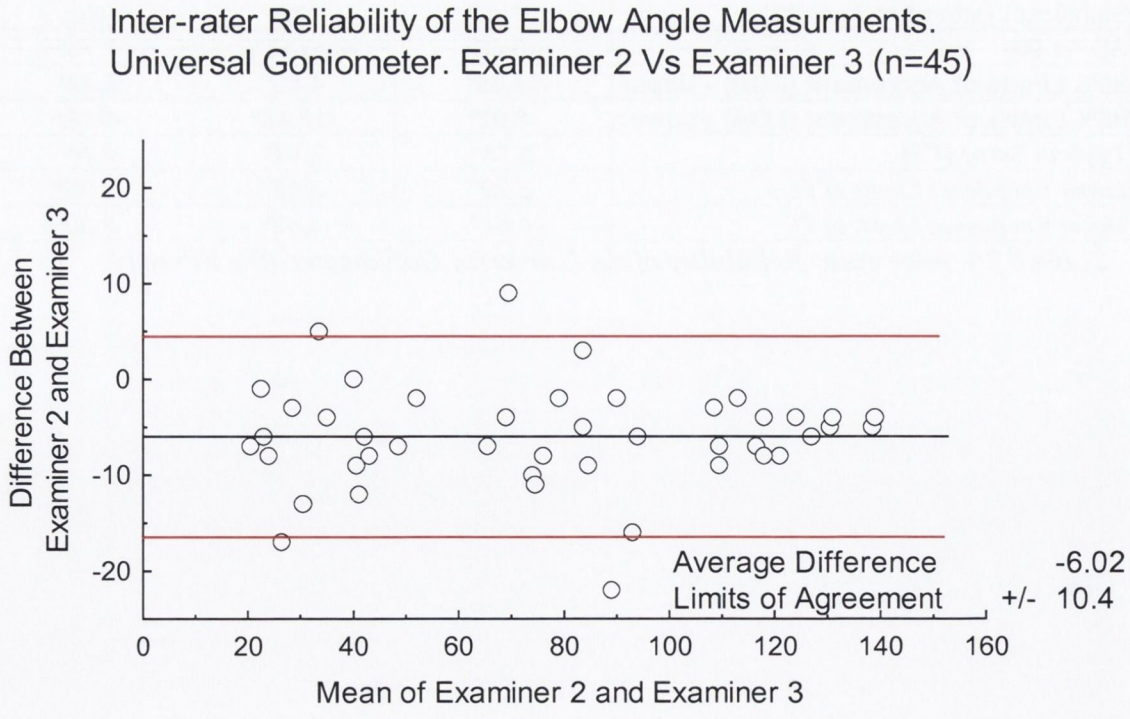
The limits of SB for UG measurements at the elbow was set at 1.9°, all of the average differences between the examiners were above this figure which confirmed the presences of non-random error between the examiners when measuring the elbow with the UG. SB was highest between examiner 2 and examiner 3, with the former examiner consistently measuring 6° above examiner 3 (see Table 5.20). Random error between the examiners as measured by TE test was lower than non-random error ranging from 2.74° (Examiner 1 Vs Examiner 3) to 3.75° (Examiner 1 Vs Examiner 2) (see Table 5.20). While the cumulative TE (multi-variant) between all the examiners was 7.19°. The ICC (3,1) was used to compare all three samples and was calculated to be 0.90.

Elbow	Ex 1 Vs Ex 2	Ex 2 Vs Ex 3	Ex 1 Vs Ex 3
Average Difference	**3.73°**	**-6.02°**	**-2.24°**
Above Systematic Bias (SB) Limit (1.9°)	Yes	Yes	Yes
Standard Deviations of Differences (SD)	5.43°	5.34°	3.94°
Standard Error of Differences (SE)	0.8°	0.79°	0.58°
Significant Difference (p=0.05)	No	Yes	No
SD * 1.96	10.66°	10.45°	7.74°
95% Limits of Agreement (LOA) - Upper	**14.39°**	**4.43°**	**5.49°**
95% Limits of Agreement (LOA) - Lower	**-6.92°**	**-16.48°**	**-9.98°**
Typical Error (TE)	**3.75°**	**3.67°**	**2.74°**
Lower Confidence Limits of TE	*3.22°*	*3.15°*	*2.35°*
Upper Confidence Limits of TE	*4.52°*	*4.42°*	*3.30°*

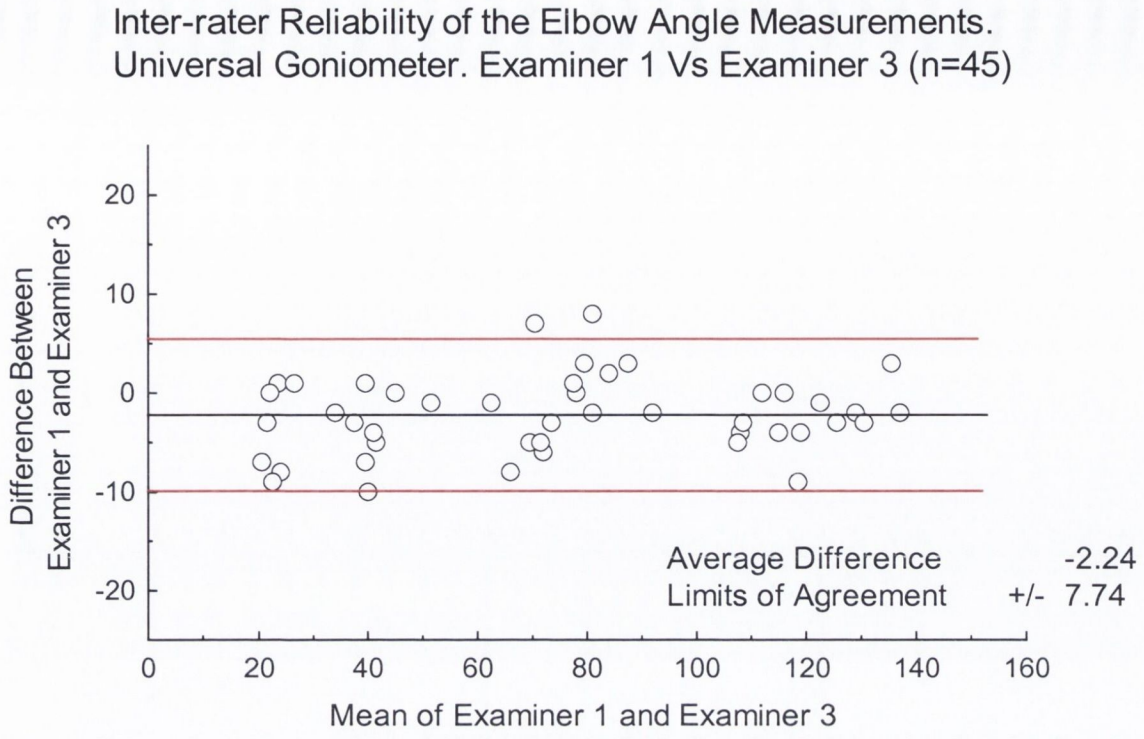
Table 5.20. Inter-rater Reliability of the Universal Goniometer (the Elbow)



Graph 5.19. 95% Limits of Agreement Scatter Plot.



Graph 5.20. 95% Limits of Agreement Scatter Plot.



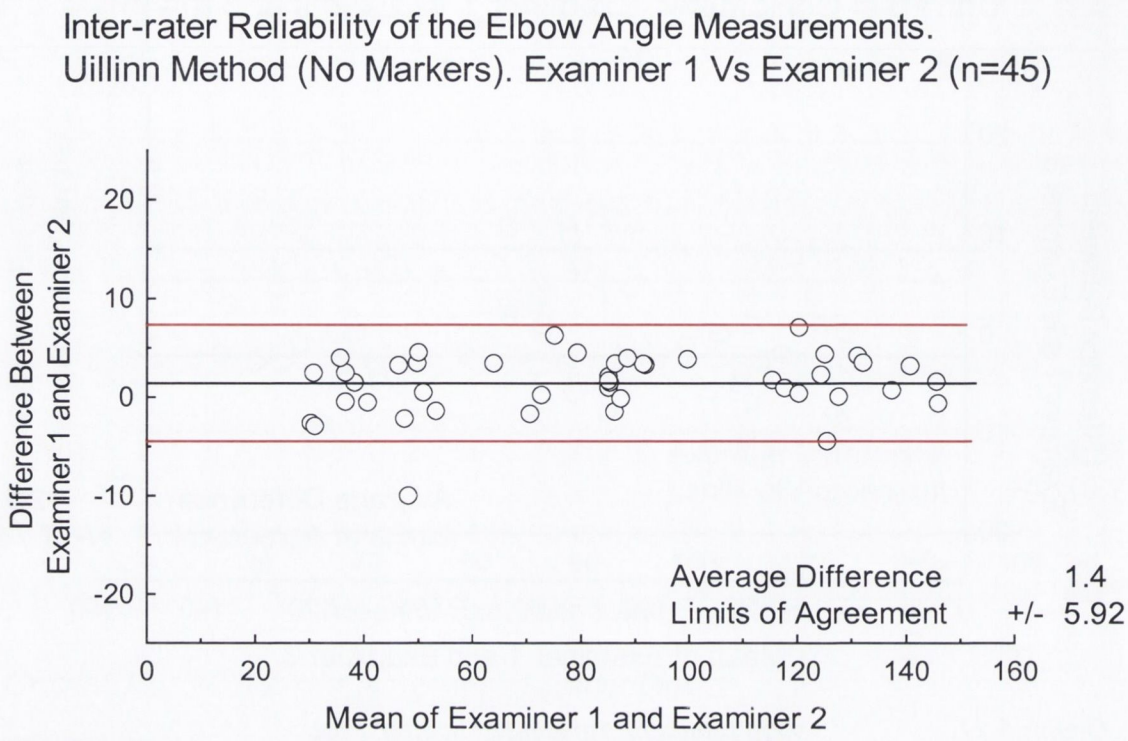
Graph 5.21. *95% Limits of Agreement Scatter Plot.*

5.4.22 Uillinn Method (No Markers) for Elbow Measurements

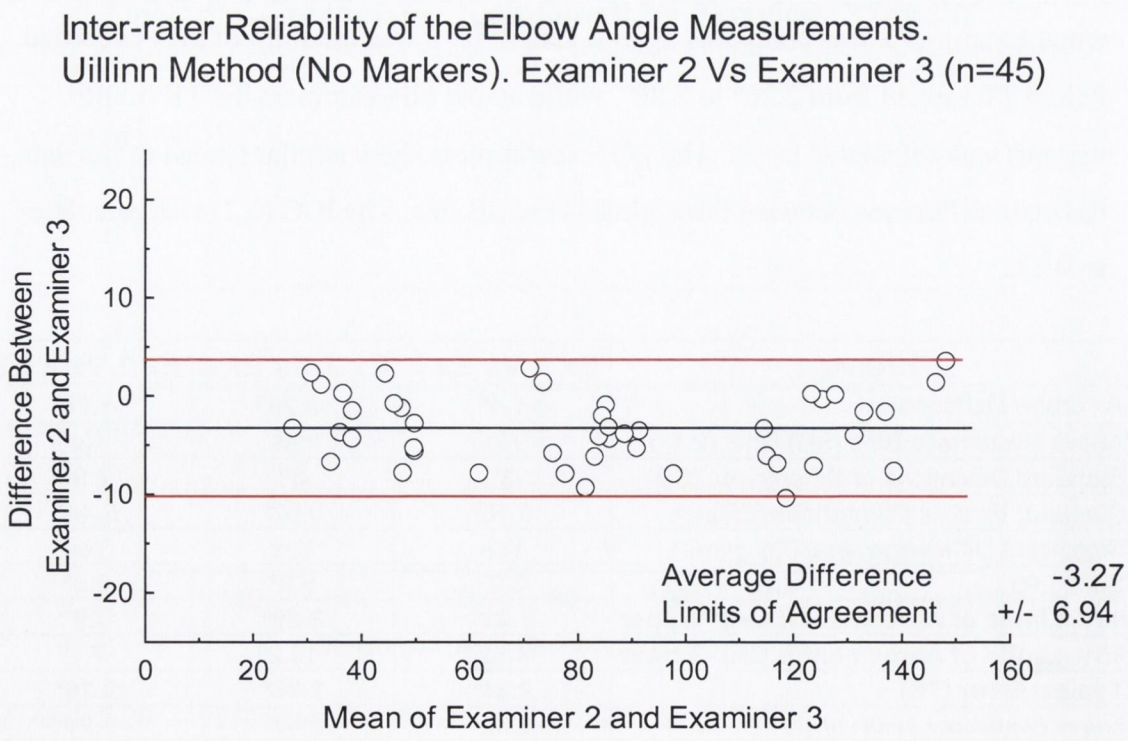
There was a trend toward SB between all of the examiners although only when Examiner 2 was compared against Examiner 3 was the limit of 2.1° exceeded. Paired TE ranged from 2.26° to 2.86°, while across all examiners the TE (multi-variant) was calculated as 3°. The LOA scatterplots show similar spread of the data, the main difference between these plots is the SB line. The ICC (3,1) was calculated as 0.93.

Elbow	Ex 1 Vs Ex 2	Ex 2 Vs Ex 3	Ex 1 Vs Ex 3
Average Difference	**1.4°**	**-3.28°**	**-1.88°**
Above Systematic Bias (SB) Limit (2.1°)	No	Yes	No
Standard Deviations of Differences (SD)	3°	3°	2.9°
Standard Error of Differences (SE)	0.45°	0.52°	0.44°
Significant Difference (p=0.05)	Yes	Yes	Yes
SD * 1.96	5.9°	6.9°	5.8°
95% Limits of Agreement (LOA) - Upper	**7.32°**	**3.66°**	**3.9°**
95% Limits of Agreement (LOA) - Lower	**-4.52°**	**-10.2°**	**-7.7°**
Typical Error (TE)	**2.31°**	**2.86°**	**2.26°**
Lower Confidence Limits of TE	*1.99°*	*2.46°*	*1.94°*
Upper Confidence Limits of TE	*2.79°*	*3.45°*	*2.72°*

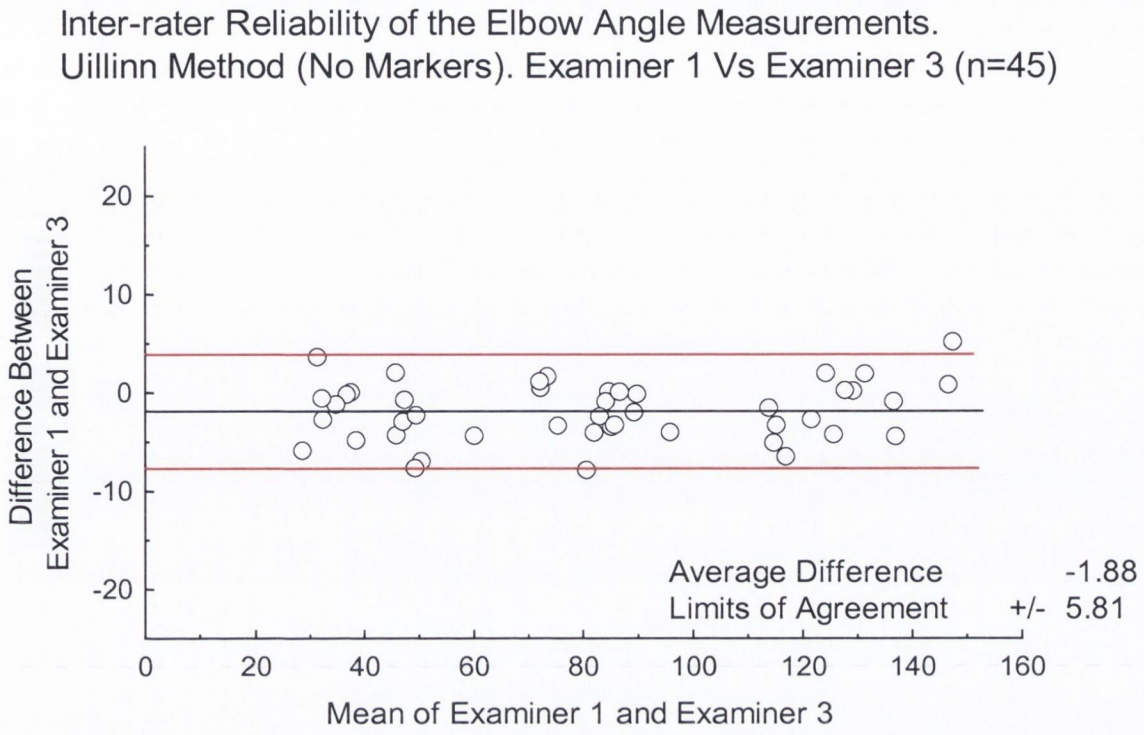
Table 5.21. Inter-rater Reliability of the Uillinn Method (No Marker) (the Elbow)



Graph 5.22. *95% Limits of Agreement Scatter Plot.*



Graph 5.23. *95% Limits of Agreement Scatter Plot.*



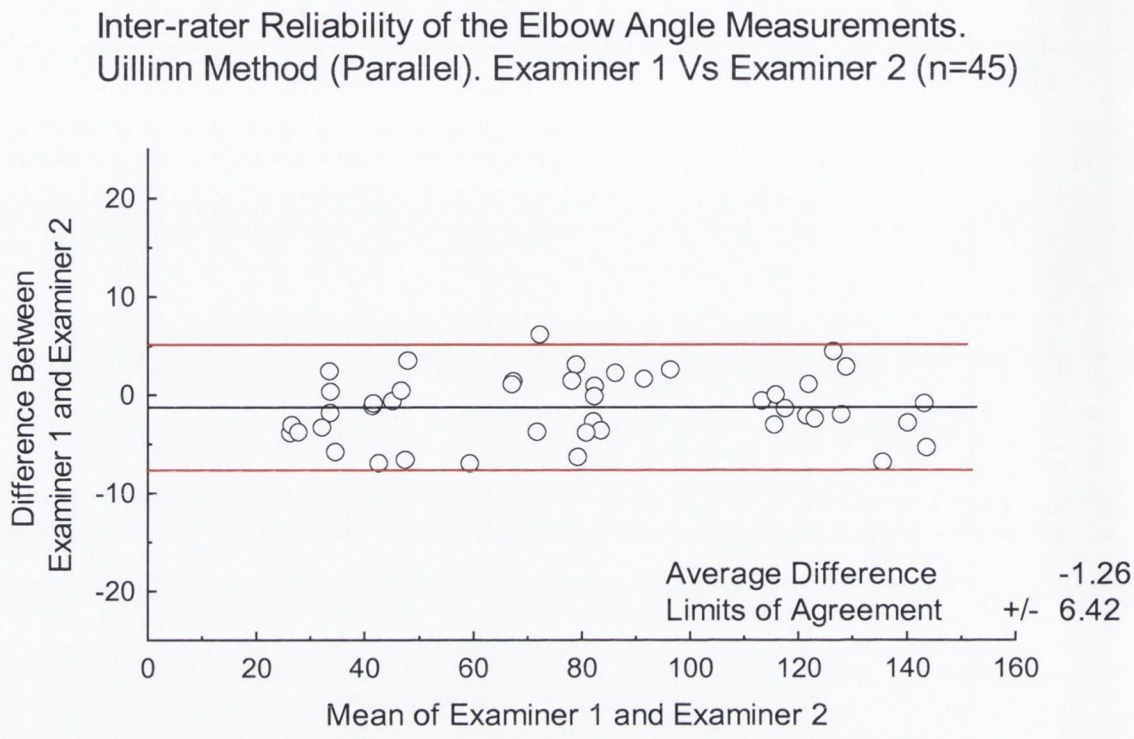
Graph 5.24. *95% Limits of Agreement Scatter Plot.*

5.4.23 Uillinn Method (Parallel) for Elbow Measurements

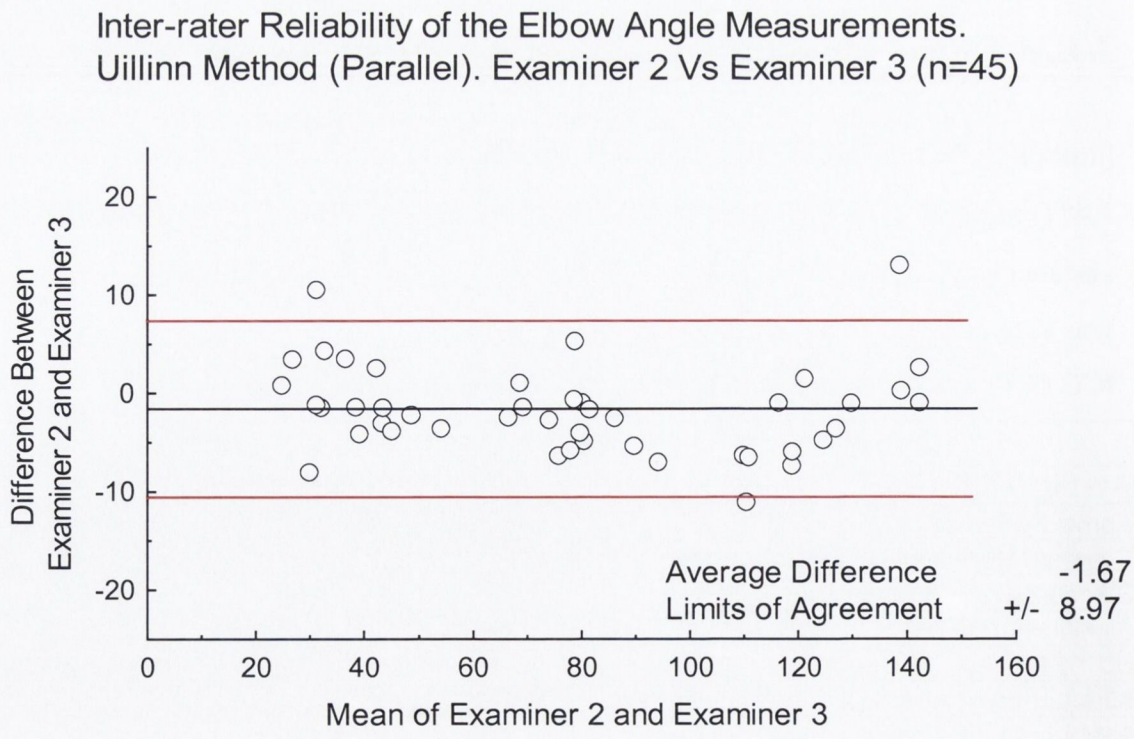
Only comparisons between Examiner 1 and Examiner 3 were over the SB limit of 2° for the Parallel approach to measuring the Elbow. TE ranged between 2.4° (Examiner 1 Vs Examiner 2) to 3.38° (Examiner 2 Vs Examiner 3). The multi-variant calculation of TE was 3.4°. The largest spread of data in the scatter plots was seen between Examiner 2 and Examiner 3 (see Graphs 5.25 5.26, 5.27). The ICC (3,1) was calculated between the 3 examiners as 0.90.

Elbow	Ex 1 Vs Ex 2	Ex 2 Vs Ex 3	Ex 1 Vs Ex 3
Average Difference	**-1.26°**	**-1.67°**	**-2.94°**
Above Systematic Bias (SB) Limit (2°)	No	No	Yes
Standard Deviations of Differences (SD)	3.27°	4.57°	4.46°
Standard Error of Differences (SE)	0.48°	0.68°	0.66°
Significant Difference (p=0.05)	Yes	Yes	Yes
SD * 1.96	6.42°	8.97°	8.76°
95% Limits of Agreement (LOA) - Upper	**5.16°**	**7.29°**	**5.82°**
95% Limits of Agreement (LOA) - Lower	**-7.68°**	**-10.65°**	**-11.7°**
Typical Error (TE)	**2.40°**	**3.38°**	**3.20°**
Lower Confidence Limits of TE	*2.06°*	*2.90°*	*2.75°*
Upper Confidence Limits of TE	*2.90°*	*4.07°*	*3.86°*

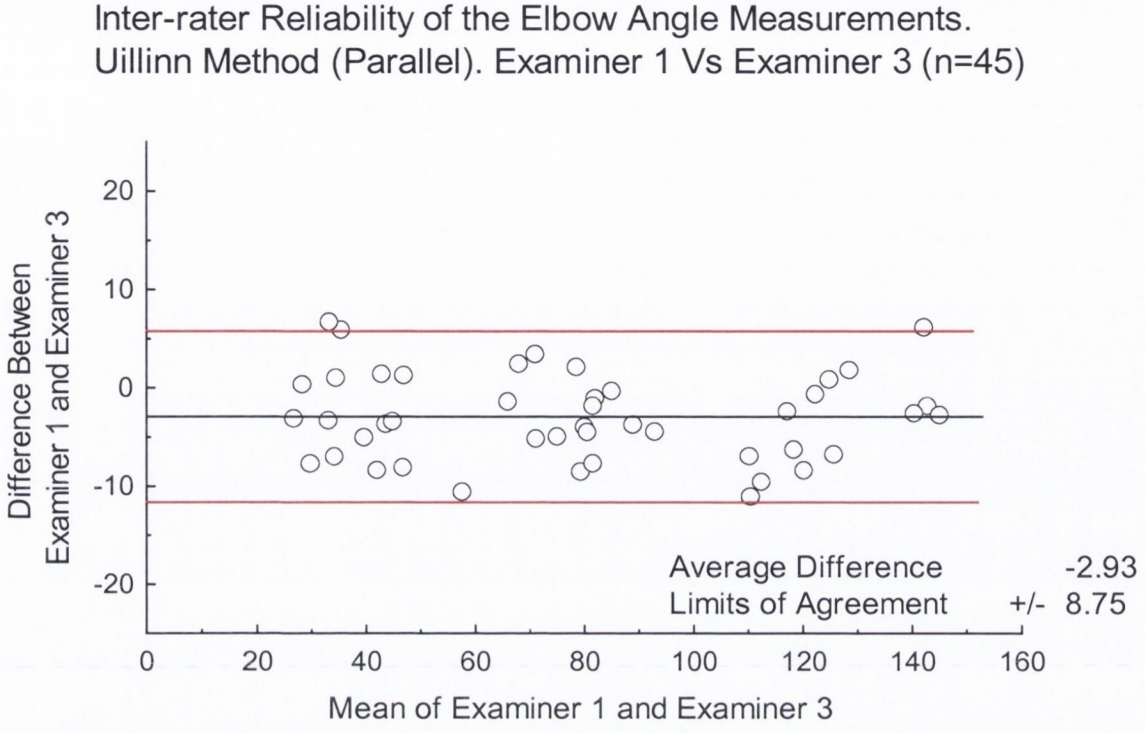
Table 5.22. Inter-rater Reliability of the Uillinn Method (Parallel) (the Elbow)



Graph 5.25. *95% Limits of Agreement Scatter Plot.*



Graph 5.26. *95% Limits of Agreement Scatter Plot.*



Graph 5.27. *95% Limits of Agreement Scatter Plot.*

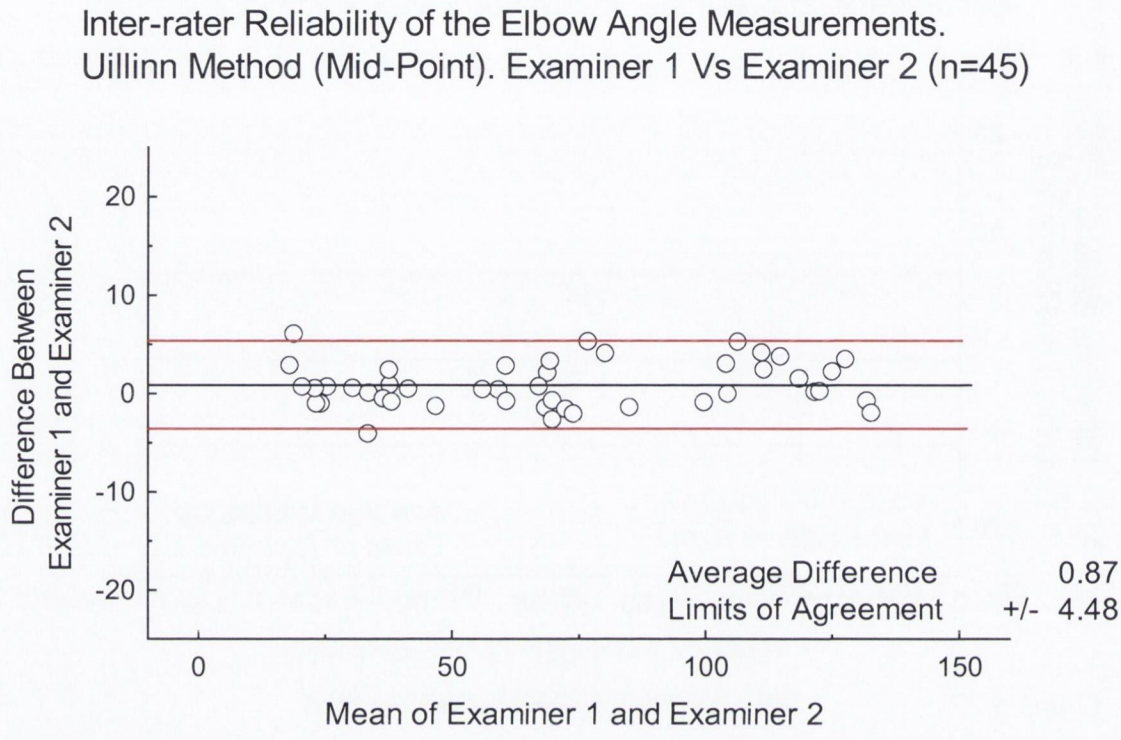
5.4.24 Uillinn Method (Mid-Point) for Elbow Measurements

Similar to the results seen at the knee, all comparisons for the Mid-point approach to the UM at the elbow showed no evidence of SB between the examiners. The highest TE was 2.33° (Examiner 2 Vs Examiner 3) which was also similar to that seen at the knee and the TE (multi-variant) between all examiners was 2.1°. Scatter plots and LOA tests show close agreement between examiners (see Graph 5.28, 5.29, 5.30). The ICC (3,1) was 0.95 between all of the examiners.

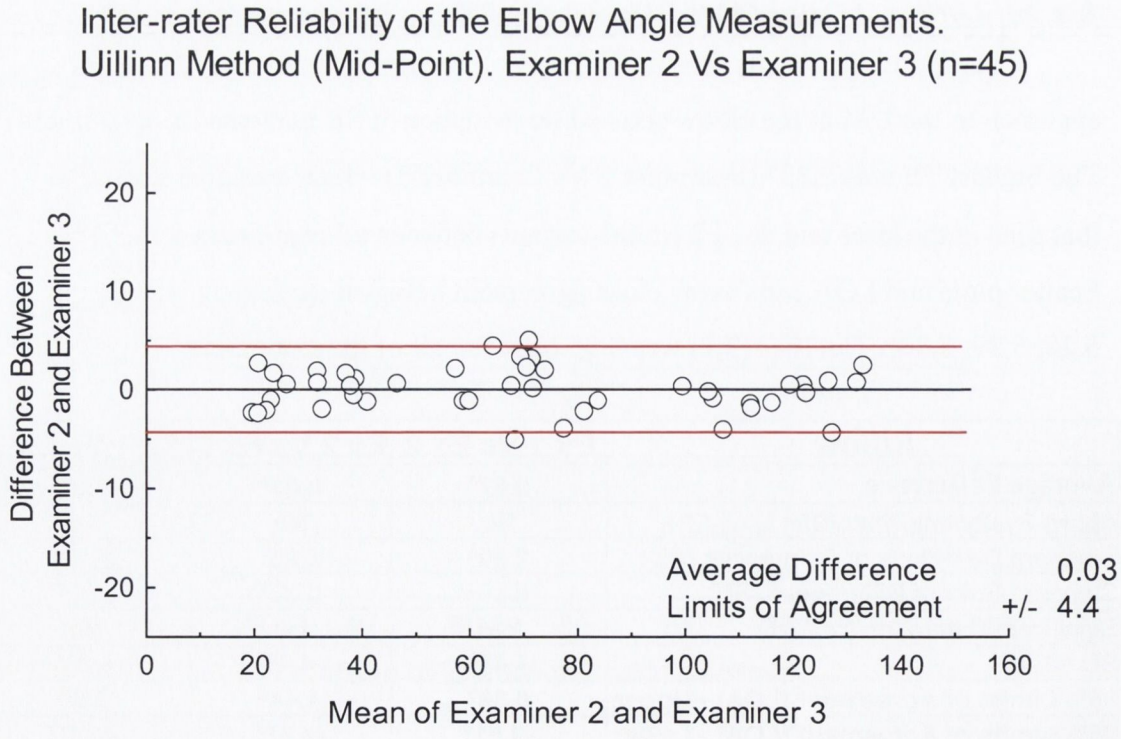
Elbow	Ex 1 Vs Ex 2	Ex 2 Vs Ex 3	Ex 1 Vs Ex 3
Average Difference	**0.87°**	**0.03°**	**0.9°**
Above Systematic Bias (SB) Limit (2°)	No	No	No
Standard Deviations of Differences (SD)	2.29°	2.25°	2.19°
Standard Error of Differences (SE)	0.34°	0.33°	0.32°
Significant Difference (p=0.05)	Yes	No	No
SD * 1.96	4.48°	4.4°	4.29°
95% Limits of Agreement (LOA) - Upper	**5.36°**	**4.44°**	**5.2°**
95% Limits of Agreement (LOA) - Lower	**-3.61°**	**-4.37°**	**-3.39°**
Typical Error (TE)	**1.55°**	**2.33°**	**2.22°**
Lower Confidence Limits of TE	*1.33°*	*2.00°*	*1.91°*

Upper Confidence Limits of TE	*1.87°*	*2.81°*	*2.68°*

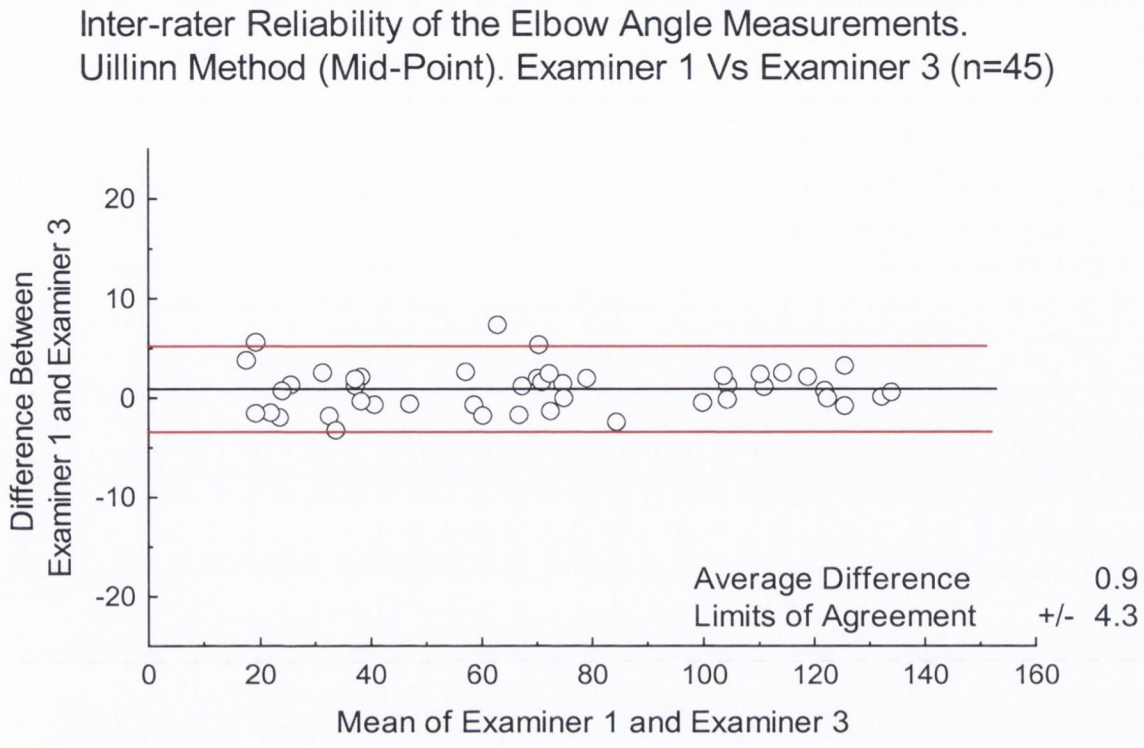
Table 5.23. Inter-rater Reliability of the Uillinn Method (Mid-Point) (the Elbow)



Graph 5.28. *95% Limits of Agreement Scatter Plot.*



Graph 5.29. *95% Limits of Agreement Scatter Plot.*



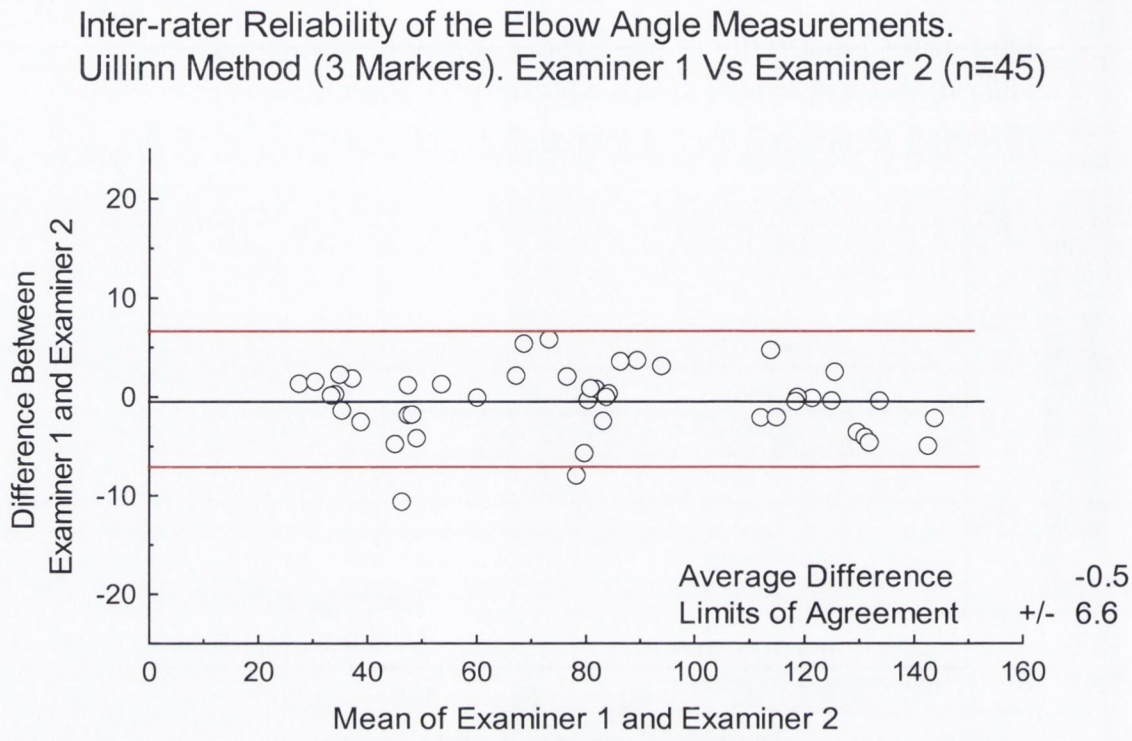
Graph 5.30. *95% Limits of Agreement Scatter Plot.*

5.4.25 Uillinn Method (3-Markers) for Elbow Measurements

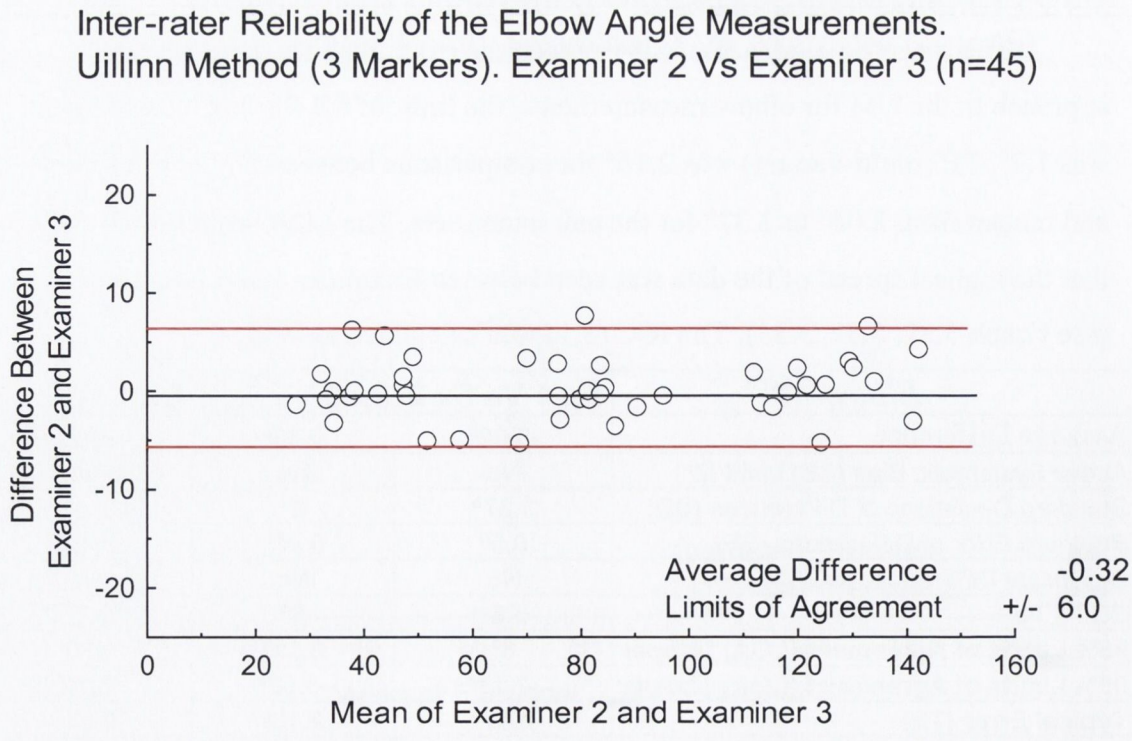
There was no evidence of SB between the examiners for the 3-Marker approach to the UM for elbow measurements, the limit of SB for this measurement was 1.7°. TE (multi-variant) was 2.16° for comparisons between all the examiners and ranged from 2.04° to 2.37° for the pair examiners. The LOA scatter plots show that the highest spread of the data was seen between Examiner 1 and Examiner 2 (see Graph 5.31, 5.32, 5.33). The ICC (3,1) was calculated as 0.94.

Elbow	**Ex 1 Vs Ex 2**	**Ex 2 Vs Ex 3**	**Ex 1 Vs Ex 3**
Average Difference	**-0.52°**	**0.32°**	**-0.2°**
Above Systematic Bias (SB) Limit (2°)	No	No	No
Standard Deviations of Differences (SD)	3.37°	3°	2.8°
Standard Error of Differences (SE)	0.5°	0.45°	0.42°
Significant Difference (p=0.05)	No	No	No
SD * 1.96	6.6°	6°	5.5°
95% Limits of Agreement (LOA) - Upper	**6°**	**6.33°**	**5.33°**
95% Limits of Agreement (LOA) - Lower	**-7.13°**	**-5.68°**	**-5.72°**
Typical Error (TE)	**2.37°**	**2.13°**	**2.04°**
Lower Confidence Limits of TE	*2.03°*	*1.82°*	*1.75°*
Upper Confidence Limits of TE	*2.85°*	*2.56°*	*2.45°*

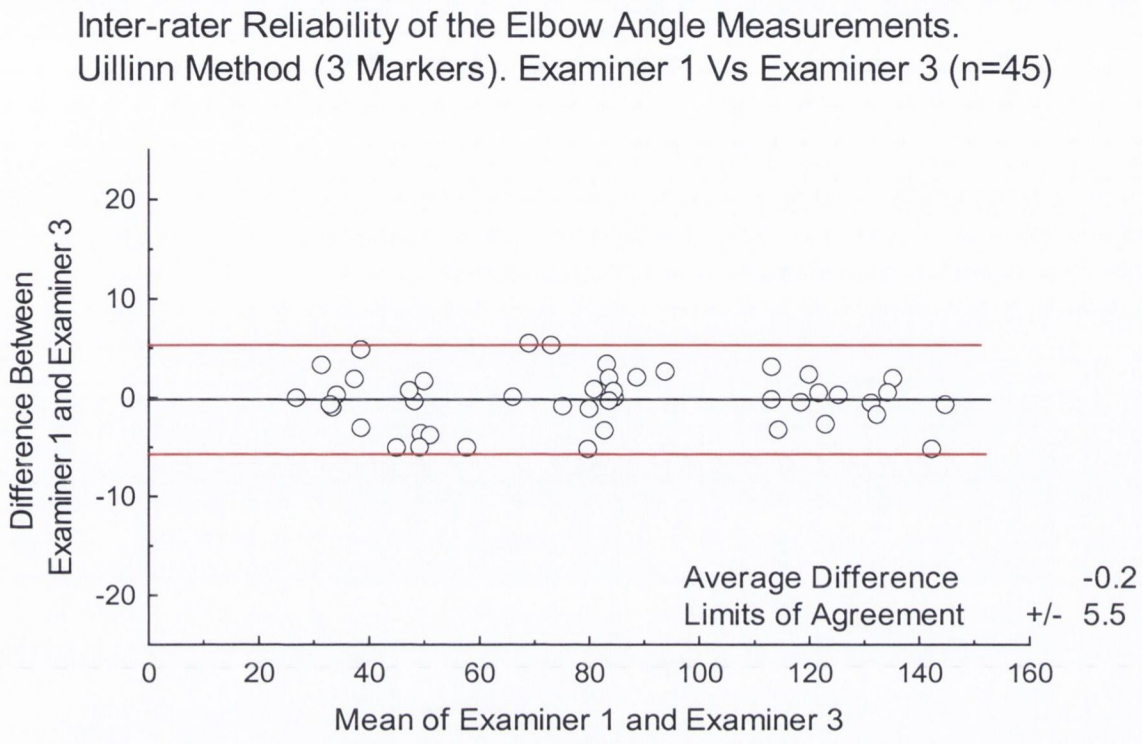
Table 5.24. Inter-rater Reliability of the Uillinn Method (3-Markers) (the Elbow)



Graph 5.31. *95% Limits of Agreement Scatter Plot.*



Graph 5.32. *95% Limits of Agreement Scatter Plot.*



Graph 5.33. *95% Limits of Agreement Scatter Plot.*

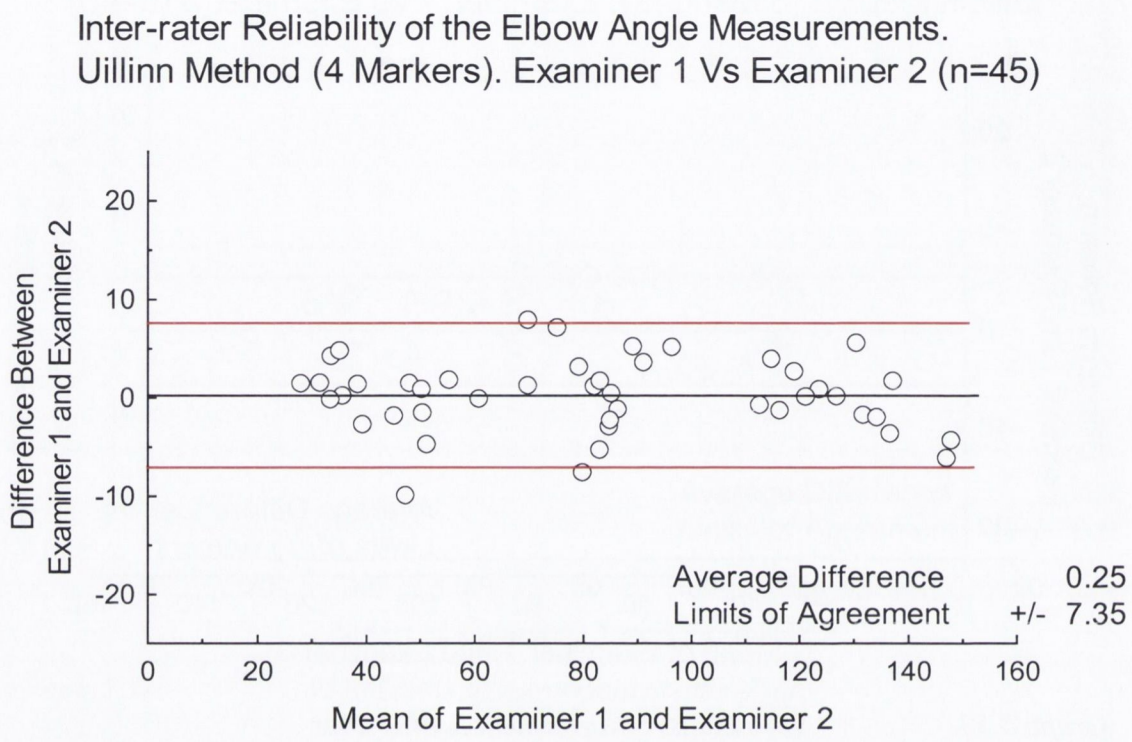
5.4.26 Uillinn Method (4-Markers) for Elbow Measurements

As with the result for all other approaches to the UM that used adhesive skin markers, there was no SB between the examiners. The highest average difference was 0.95° which was below the 2° SB limit. TE ranged from 2.27° to 2.7° between the paired examiners and was 2.46° across all the examiners (multi-variant). The scatter plots show a similar spread of the data for each separate analysis (see Graph 5.34, 5.35, 5.36). The ICC (3,1) between the 3 examiners for this type of measurement was calculated as 0.93.

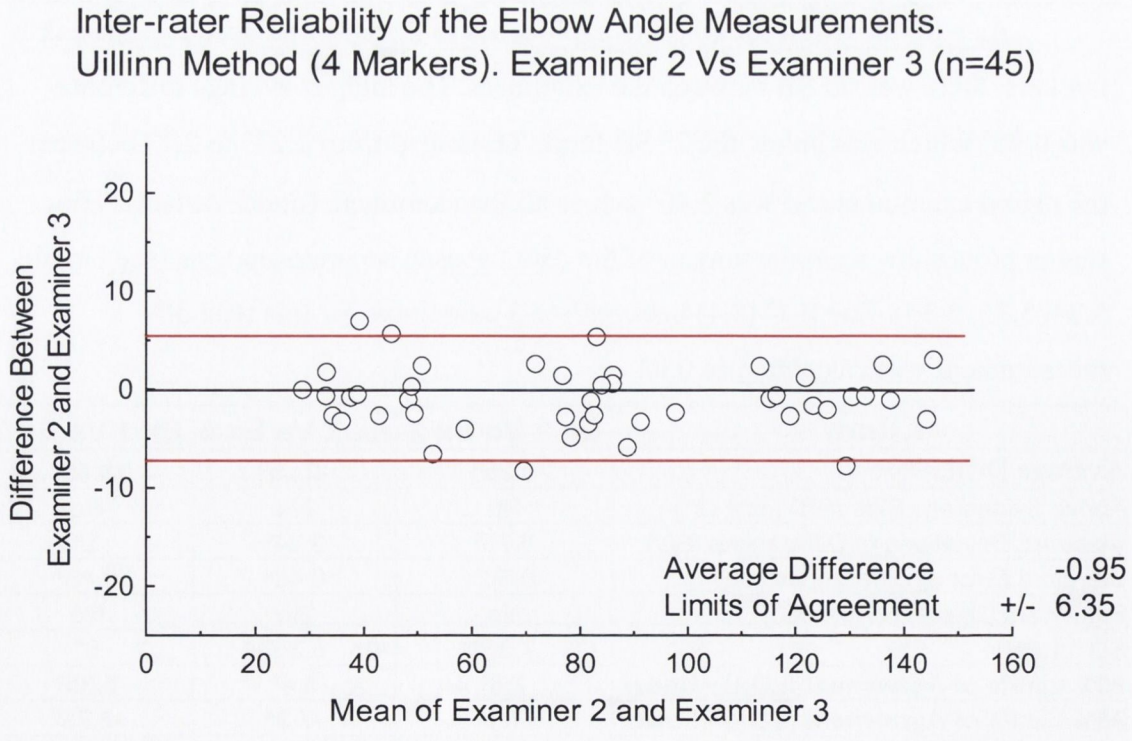
Elbow	Ex 1 Vs Ex 2	Ex 2 Vs Ex 3	Ex 1 Vs Ex 3
Average Difference	**0.25°**	**-0.95°**	**-0.69°**
Above Systematic Bias (SB) Limit (2°)	No	No	No
Standard Deviations of Differences (SD)	3.75°	3.24°	3°
Standard Error of Differences (SE)	0.56°	0.48°	0.46°
Significant Difference (p=0.05)	No	No	No
SD * 1.96	7.35°	6.35°	6°
95% Limits of Agreement (LOA) - Upper	**7.6°**	**5.4°**	**5.35°**
95% Limits of Agreement (LOA) - Lower	**-7.1°**	**-7.3°**	**-6.73°**
Typical Error (TE)	**2.7°**	**2.28°**	**2.27°**
Lower Confidence Limits of TE	*2.32°*	*1.95°*	*1.95°*

Upper Confidence Limits of TE	3.26°	2.74°	2.74°

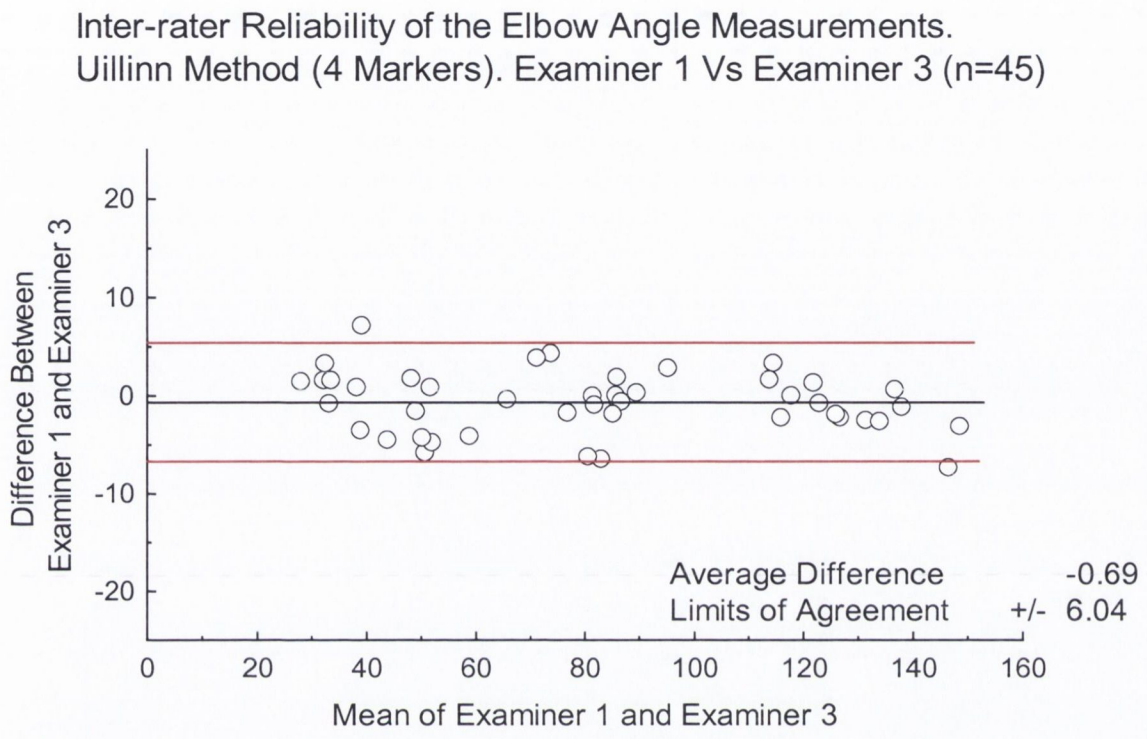
Table 5.25. Inter-rater Reliability of the Uillinn Method (4-Markers) (the Elbow)



Graph 5.34. *95% Limits of Agreement Scatter Plot.*



Graph 5.35. *95% Limits of Agreement Scatter Plot.*



Graph 5.36. *95% Limits of Agreement Scatter Plot.*

Typical Error Between All Examiners (3- Way)		
	Knee	Elbow
UG	3.99°	7.19°
No Markers	2.44°	3°
Parallel	2.8°	3.4°
Mid-Point	1.99°	2.1°
3-Markers	1.75°	2.16°
4-Markers	1.73°	2.46

ICC (3,1) Between All Examiners		
	Knee	Elbow
UG	0.89	0.90
No Markers	0.97	0.93
Parallel	0.95	0.90
Mid-Point	0.97	0.95
3-Markers	0.98	0.94
4-Markers	0.98	0.93

Table 5.26. Inter-rater Reliability of all Techniques. Typical error (TE) and ICC (3,1) between all examiners.

5.4.27 Confidence in the Data

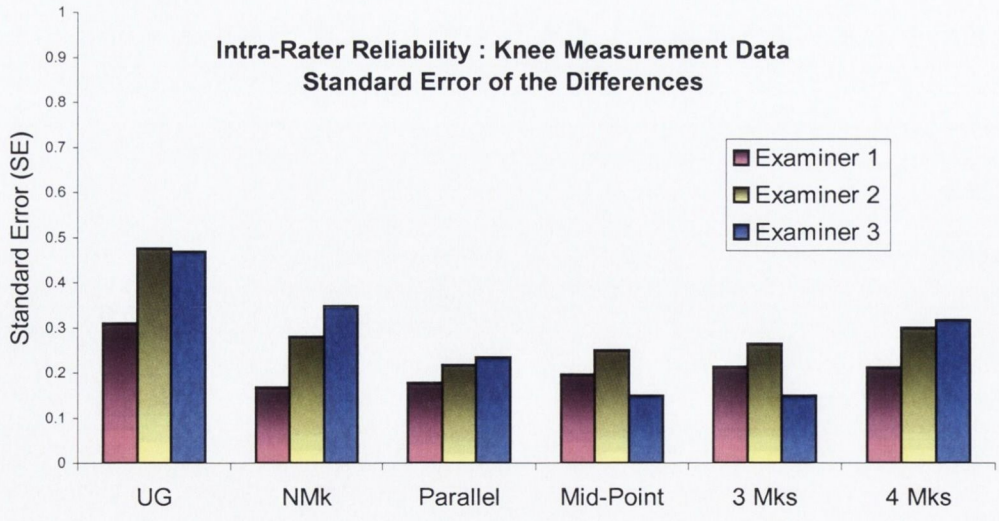
The certainty of the data upon which these calculations were based was established using a calculation of standard error (SE) in the data. This test was applied to the data for the differences between repeated observations, both within testers and between the testers.

The intra-rater reliability data showed that for the majority of the techniques further sampling would not produce variations greater than 0.4°. The UG data would produce greater changes of up to 0.62° (see Table 5.37, 5.38). Nonetheless, this magnitude of SE in the data would probably not have a dramatic effect the SD of the differences of the data, and therefore have little impact on the results.

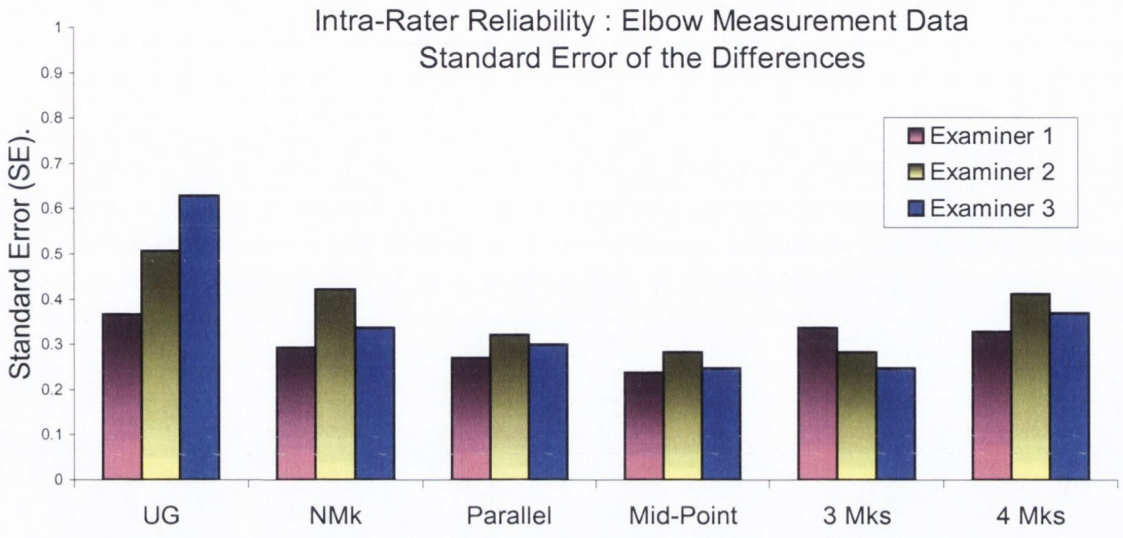
KNEES Standard Error of the Differences			
	Examiner 1	Examiner 2	Examiner 3
UG	**0.3°**	**0.47°**	**0.46°**
NMk	**0.16°**	**0.27°**	**0.34°**
Parallel	**0.17°**	**0.21°**	**0.23°**
Mid-Point	**0.19°**	**0.25°**	**0.14°**
3 Mks	**0.21°**	**0.26°**	**0.14°**
4 Mks	**0.21°**	**0.29°**	**0.31°**

ELBOWS Standard Error of the Differences			
	Examiner 1	Examiner 2	Examiner 3
UG	**0.36°**	**0.5°**	**0.62°**
NMk	**0.29°**	**0.42°**	**0.33°**
Parallel	**0.27°**	**0.32°**	**0.3°**
Mid-Point	**0.23°**	**0.28°**	**0.24°**
3 Mks	**0.33°**	**0.28°**	**0.24°**
4 Mks	**0.32°**	**0.41°**	**0.36°**

Table 5.27. Intra-rater Reliability: Standard Error of the Differences for the different measurement techniques.



Graph 5.37. Intra-rater Reliability: Standard Error within the Data for the Difference between the Repeated Assessments.



Graph 5.38. Intra-rater Reliability: Standard Error within the Data for the Difference between the Repeated Assessments.

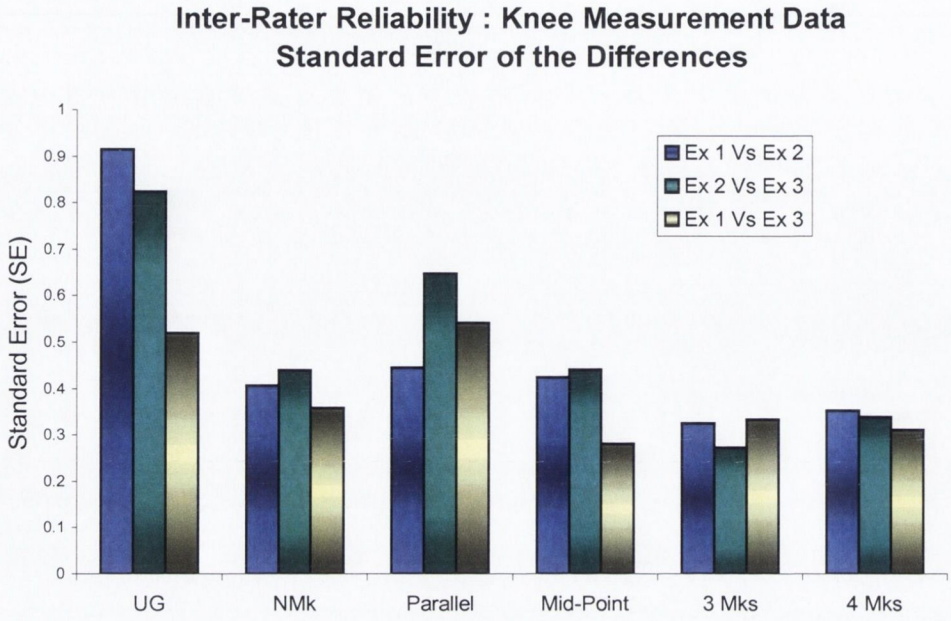
For inter-rater reliability the SE analysis showed that with the exception of the UG data, all tests were less than 0.7°. The SE of the UG data was up to 0.91°. This implied that further sampling could have produced additional differences of these amounts (see Table 5.32; Graph 5.39, 5.40). The effect of an additional difference of this magnitude (< 1°) on such a large number of the SD of the differences (n≈50), would again probably be minimal.

Upper and lower confidence levels for TE calculations were presented in the appropriate tables in the results section.

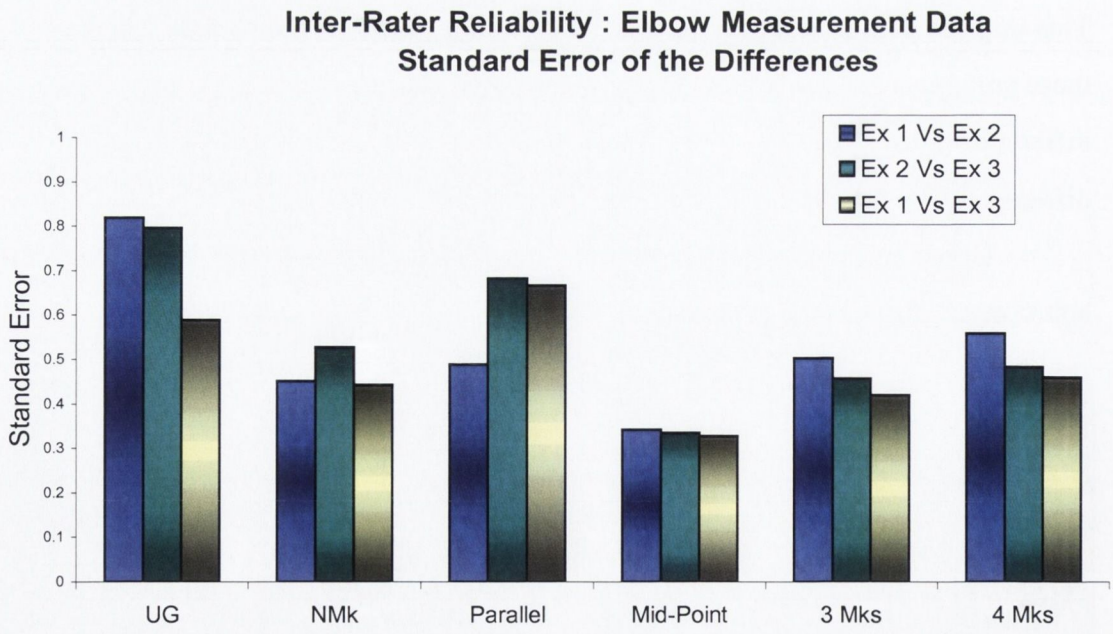
KNEES Standard Error of the Differences			
	Ex 1 Vs Ex 2	Ex 2 Vs Ex 3	Ex 1 Vs Ex 3
UG	0.91°	0.82°	0.51°
No Markers	0.4°	0.4°	0.35°
Parallel	0.44°	0.64°	0.54°
Mid Point	0.42°	0.44°	0.28°
3-Markers	0.32°	0.27°	0.33°
4-Markers	0.35°	0.33°	0.31°

ELBOWS Standard Error of the Differences			
	Ex 1 Vs Ex 2	Ex 2 Vs Ex 3	Ex 1 Vs Ex 3
UG	0.81°	0.79°	0.58°
No Markers	0.45°	0.52°	0.44°
Parallel	0.48°	0.68°	0.66°
Mid Point	0.34°	0.33°	0.32°
3-Markers	0.5°	0.45°	0.41°
4-Markers	0.56°	0.48°	0.45°

Table 5.32. Inter-rater Reliability: Standard Error of the Differences for the different measurement techniques.



Graph 5.39. Inter-rater Reliability: Standard Error within the Data.



Graph 5.40. Inter-rater Reliability: Standard Error within the Data.

5.5. Discussion

Chapter 2 outlined the development of the UM and proposed 5 different possible techniques toward its application to the human model. The ability of the system to correctly calculate inanimate angles was tested in Chapter 4. Those experiments found that angular measurements would typically be accurate to less than 0.3° and for 95% of measurements would be within 0.8° of a mathematical control.

This current set of experiments set out to establish the reliability of the UM for measurements of human joints. The study also aimed to assess the reliability of the standard UG and compare its performance to that of the UM. The aims of the study were to investigate both int*ra*-rater and int*er*-rater reliability of the UG and the 5 different approaches of the UM for joint angle measurements at the elbow and the knee. The final aim of the study was to report on issues regarding the reliability of re-applying adhesive skin markers for the purpose of joint angle measurements.

The results show that non-random error was not present in any of the 6 different methods when measurements were made by the same examiner. However, the data did support the notion of SB for some of the different methods when the between examiners data were compared. Specifically, SB was present in one of the methods of the knee measurements (No Marker) and 3 assessment methods for upper limb (UG, No Marker, Parallel) (see Graph 5.45). The worst SB was seen for the UG at the elbow where there were differences of over 6° between two of the examiners. This meant that one of the examiners (Examiner 2) would consistently measure an additional 6° more than their colleague (Examiner 3) for assessments of the same elbow when using a UG. This figure did not include the influence of random error on the assessments.

The Parallel and Mid-point approaches showed the least non-random error between the examiners for knee measurements. While at the elbow, the 3-Marker and Mid-point approaches displayed the least SB. The Mid-point approach was therefore the common method for the lowest non-random error between examiners at both joints. This method was seen as the approach that shows the least SB for joint angle measurements (see Graph 5.45).

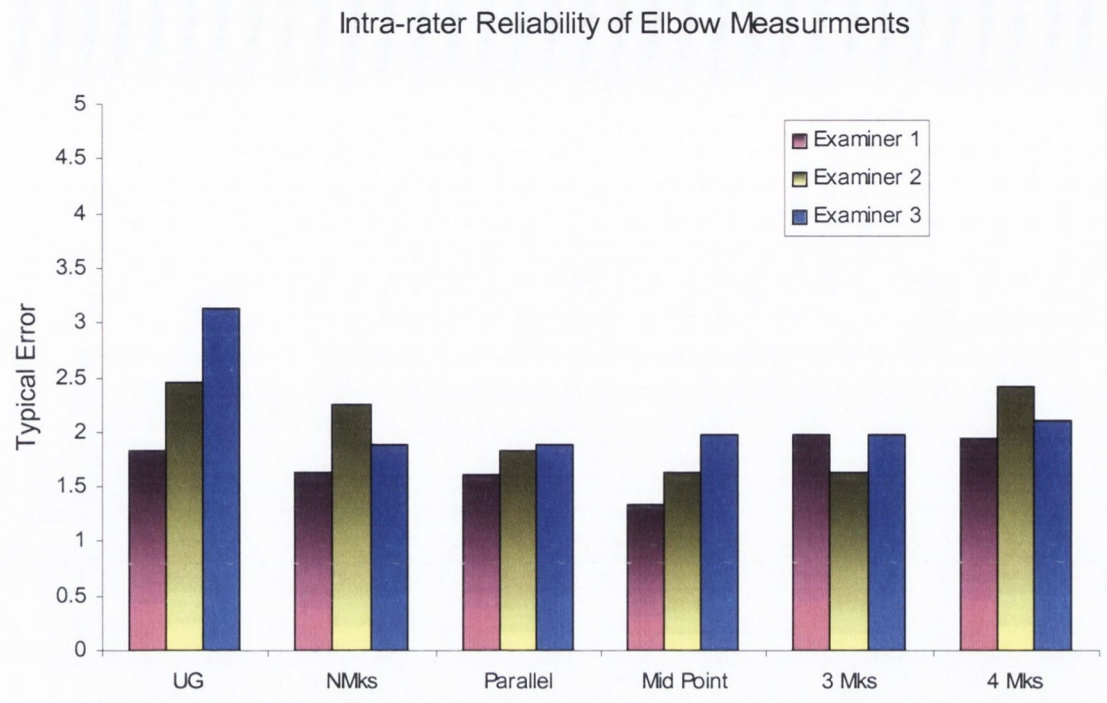
Repeatability Coefficient							
KNEES					ELBOWS		
	Examiner 1	Examiner 2	Examiner 3			Examiner 1	Examiner 2	Examiner 3
UG	4.33°	6.64°	6.54°		UG	5.04°	6.96°	8.62°
No Markers	2.34°	3.91°	4.86°		No Markers	4.02°	5.78°	4.61°
Parallel	2.48°	3.03°	3.27°		Parallel	3.7°	4.41°	4.12°
Mid Point	2.74°	3.49°	2.08°		Mid Point	3.27°	3.88°	3.4°
3-Markers	2.97°	3.69°	2.08°		3-Markers	4.62°	3.88°	3.4°
4-Markers	2.95°	4.18°	4.43°		4-Markers	4.51°	5.65°	5.06°

Table 5.27. Intra-rater Reliability of All Techniques (Repeatability Coefficient)

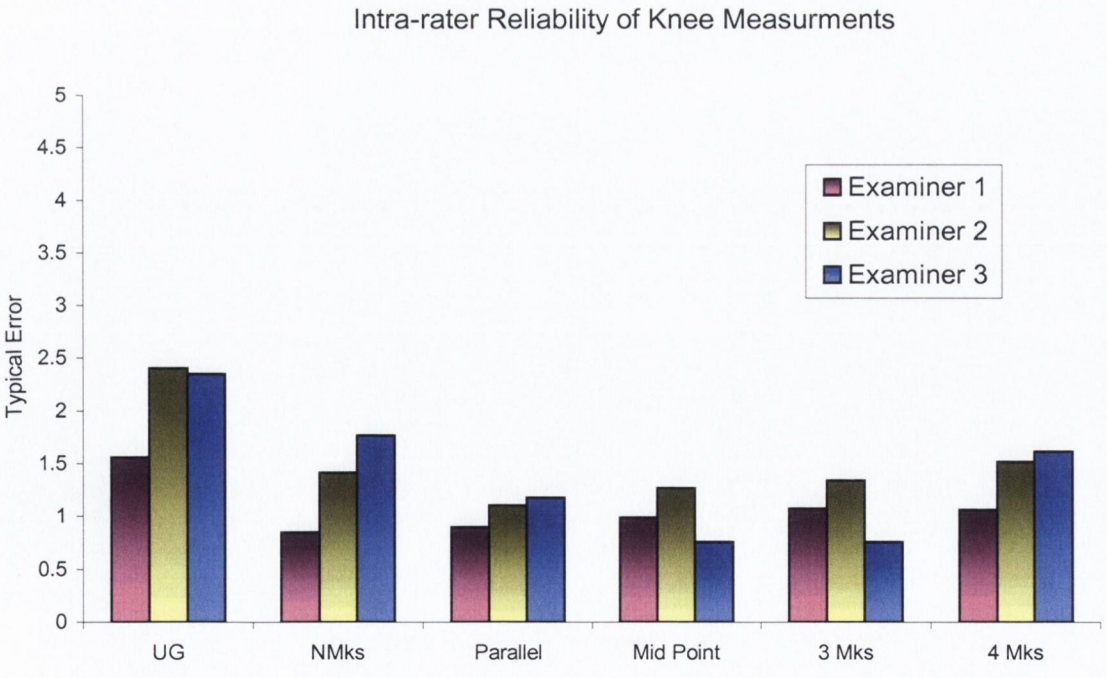
Typical Error							
KNEES					ELBOWS		
	Examiner 1	Examiner 2	Examiner 3			Examiner 1	Examiner 2	Examiner 3
UG	1.56°	2.4°	2.35°		UG	1.82°	2.45°	3.14°
No Markers	0.85°	1.41°	1.77°		No Markers	1.63°	2.26°	1.88°
Parallel	0.89°	1.1°	1.17°		Parallel	1.61°	1.83°	1.88°
Mid Point	0.99°	1.27°	0.76°		Mid Point	1.34°	1.63°	1.98°
3-Markers	1.07°	1.34°	0.75°		3-Markers	1.97°	1.63°	1.97°
4-Markers	1.06°	1.51°	1.61°		4-Markers	1.94°	2.41°	2.09°

Table 5.28. Intra-rater Reliability of All Techniques (Typical Error).

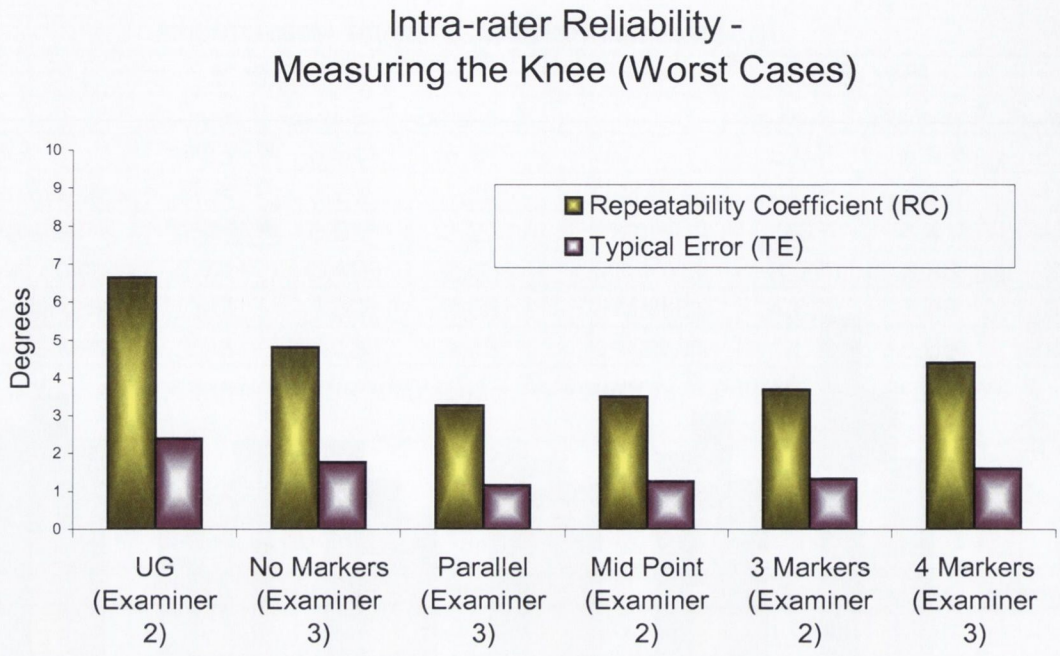
Because there was no evidence of SB in the intra-rater reliability tests, the RC was used in addition to TE. Assessments of RC (intra-rater) showed that the UG had the worst maximum observed results of 6.6° (Examiner 2) at the knee and of 8.62° (Examiner 3) at the elbow. If the maximum variation from the group of 3 examiners are inspected, i.e. the least favourable observations, the Parallel method had the best results (RC of 3.27° at the knee (Examiner 3) and 3.7° at the elbow (Examiner 1)). These results were echoed in the TE data (see Table 5.28; Graph 5.36a to 5.39).



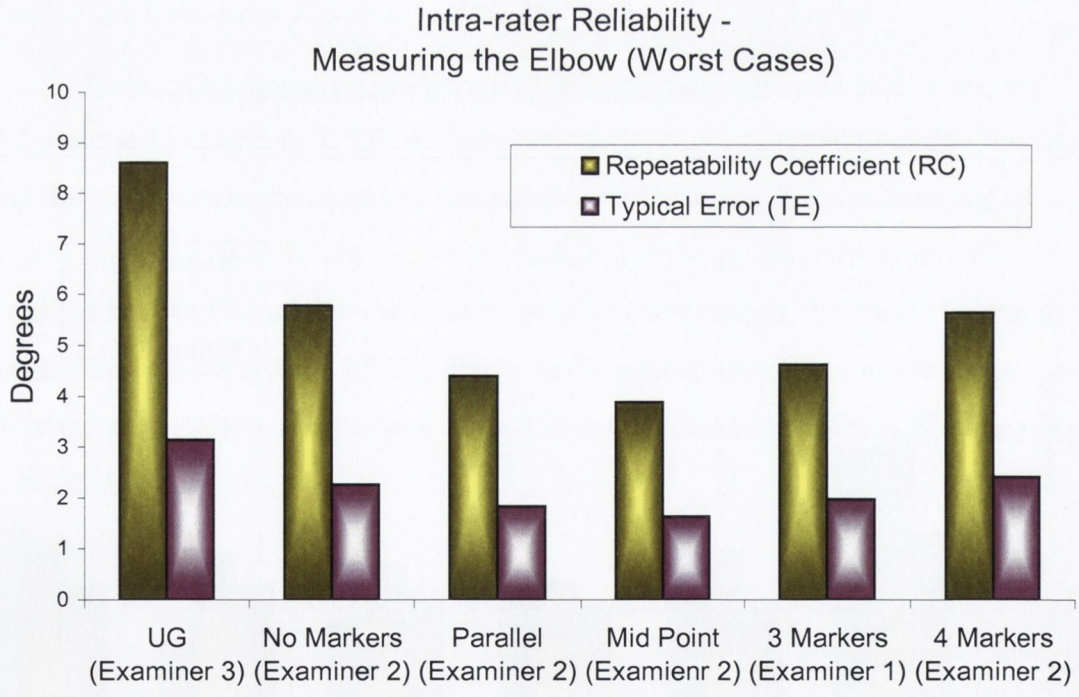
Graph 5.41 Intra-rater Reliability: Typical Error for Different Measurement Techniques.



Graph 5.42. Intra-rater Reliability: Typical Error for Different Measurement Techniques.



Graph 5.43. Intra-Rater Reliability at the Knee – Worst Observed Cases



Graph 5.44. Intra-Rater Reliability at the Elbow – Worst Observed Cases

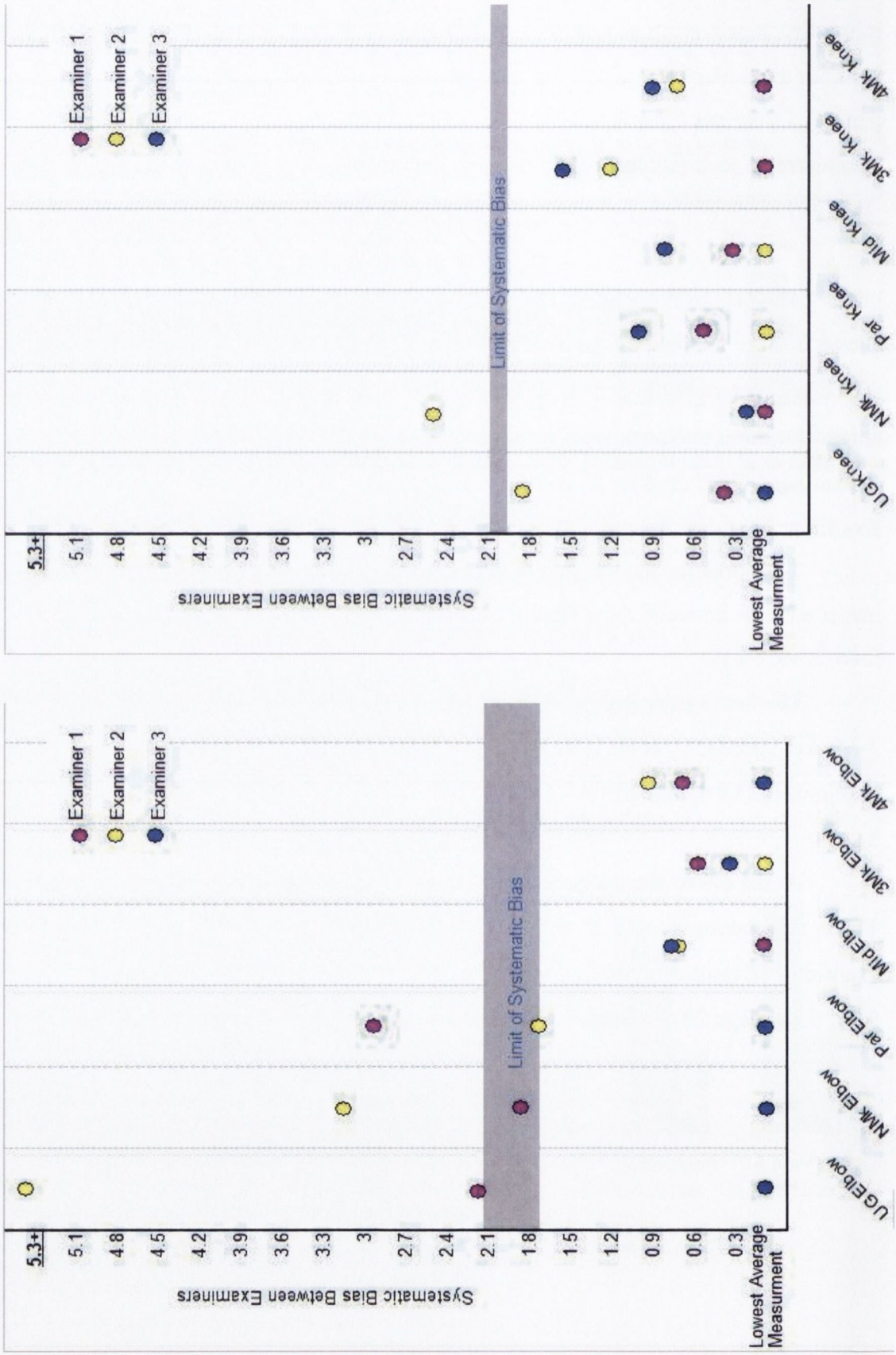
The Parallel method showed the smallest random error for intra-rater reliability, however, there was little real difference between it and the error observed for both the 3-Marker and Mid-point approaches (see Graph 5.36a to 5.39). In general, Examiner 1 displayed the best intra-rater reliability over the different measurement techniques for both the knee and elbow.

Analysis of the random error between the examiners (inter-rater reliability) shows that the UG had the poorest agreements. The TE between the examiners was 4° for knee measurements and 7.2° at the elbow. The maximum observed LOA spread for these measurements at the knee were -11.2°, 14.4° (2 SD's of 12.8°) (Examiner 1 Vs Examiner 2) and -7°, 14.5° (2 SD's of 10.7°) (Examiner 1 Vs Examiner 2) at the elbow.

TE scores for the UG and the Parallel method had the poorest agreement for measurements between the different examiners (see Table 5.26 (page 227); Graph 5.46, 5.40, 5.41).

The best agreement between examiners were by both the Mid-Point and the 3-Marker approach. At the knee 4-Marker approach had a TE score of 1.73° with a maximum LOA spread of –3.1°, 4.6° (2 SD's of 4.65°) (Examiner 1 Vs Examiner 3).

At the elbow the best observed scores were for the Mid-Point approach. The TE for this measure was 2.1° with a LOA spread (maximum between paired examiners (Examiner 1 Vs Examiner 2)) of –3.6°, 5.3° (2 SD's of 4.48°) (see Table 5.29, 5.30 (page 237); Graph 5.46, 5.40, 5.41 (page 237-238)).



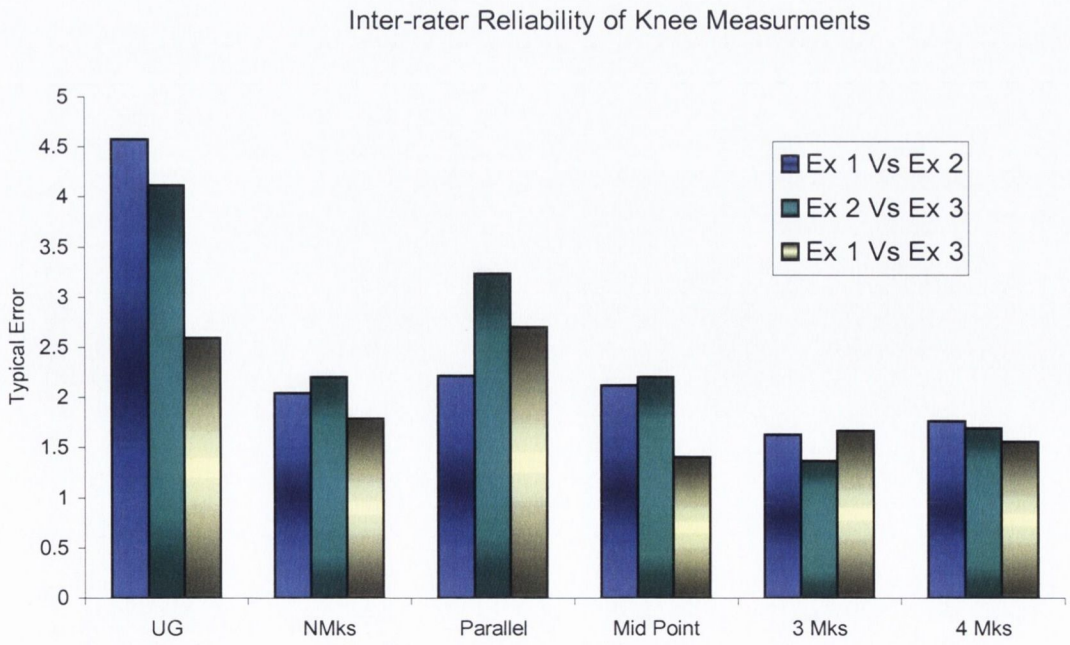
Graph 5.45 Systematic Bias Between Examiners for the Different Measurement Techniques.

Repeatability Coefficient							
KNEES					ELBOWS		
	Ex 1 Vs Ex 2	Ex 2 Vs Ex 3	Ex 1 Vs Ex 3		Ex 1 Vs Ex 2	Ex 2 Vs Ex 3	Ex 1 Vs Ex 3
UG	12.8°	11.52°	7.26°	UG	10.76°	10.45°	7.74°
No Markers	5.69°	6.14°	5°	No Markers	5.92°	6.93°	5.81°
Parallel	6.21°	9°	7.57°	Parallel	6.42°	8.97°	8.75°
Mid Point	5.93°	6.17°	3.93°	Mid Point	4.48°	4.4°	4.29°
3-Markers	4.54°	3.82°	4.65°	3-Markers	6.6°	6°	5.51°
4-Markers	4.94°	4.74°	4.35°	4-Markers	7.35°	6.35°	6°

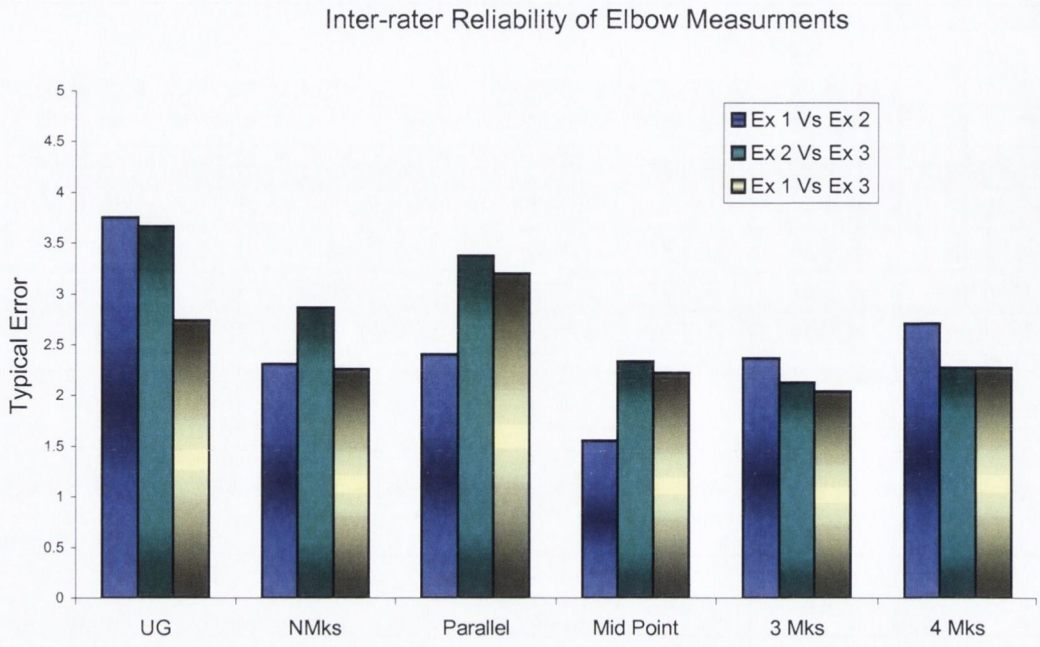
Table 5.29. Inter-rater Reliability of All Techniques (Repeatability Coefficient)

Typical Error							
KNEES					ELBOWS		
	Ex 1 Vs Ex 2	Ex 2 Vs Ex 3	Ex 1 Vs Ex 3		Ex 1 Vs Ex 2	Ex 2 Vs Ex 3	Ex 1 Vs Ex 3
UG	4.58°	4.12°	2.59°	UG	3.75°	3.66°	2.74°
No Markers	2°	2.2°	1.79°	No Markers	2.31°	2.86°	2.26°
Parallel	2.22°	3.23°	2.7°	Parallel	2.4°	3.37°	3.2°
Mid Point	2.12°	2.21°	1.41°	Mid Point	1.55°	2.33°	2.22°
3-Markers	1.62°	1.36°	1.67°	3-Markers	2.36°	2.12°	2°
4-Markers	1.76°	1.69°	1.55°	4-Markers	2.7°	2.27°	2.27°

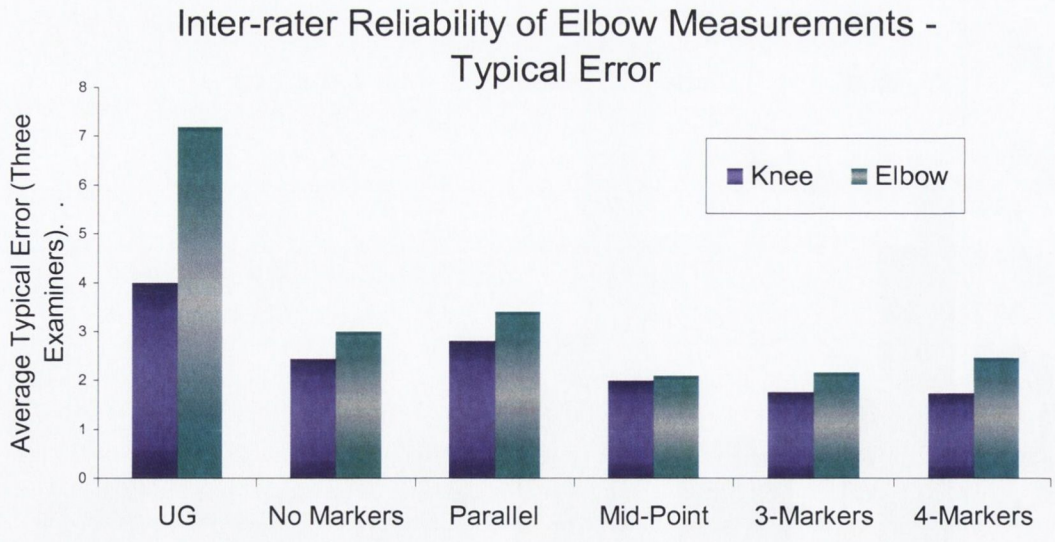
Table 5.30. Inter-rater Reliability of All Techniques (Typical Error).



Graph 5.46. Inter-rater Reliability: Typical Error for Different Measurement Techniques, Between Examiners.



Graph 5.47. Inter-rater Reliability: Typical Error for Different Measurement Techniques, Between Examiners.



Graph 5.48. Inter-rater Reliability: Typical Error for Different Measurement Techniques, Between Examiners.

5.5.1. The Most Reliable Method

When judging which was the most reliable method overall, the amount of random and non-random error for repeated assessments by the same and different examiners were considered. It was appropriate to consider that the method which had the lowest of these criteria for both joints would have the most versatile reliability.

The No Marker method had SB between examiners at both the knee and elbow. The UG and Parallel approach showed SB between examiners for elbow measurements only. Therefore, on the basis of high non-random error alone, these methods were discounted. Only the Mid-Point approach, the 3-Marker and 4-Marker methods had no SB in their measurements.

The multi-variant analysis of TE for inter-rater reliability show similar results between these remining measurements techniques, with the 4-Marker approach slightly higher at the elbow (see Table 5.26, page 227). A multi-variant TE was not possible for the intra-rater reliability. In order to gauge comparative performance of the TEs for each technique the data from each of the examiners was averaged. These averages were then presented beside the inter-rater TE (multi-variant) (see Table 5.33). This format was chosen for simplicity of understanding.

Again this comparison show how similar the performances of the Mid-Point and 3-Marker approaches were. The Mid-Point approach had a slightly better intra-rater reliability (averaged TE) compared to the 3-Marker approach, while these slight differences were reversed for inter-rater reliability of the knee.

While these two approaches achieved the strongest overall results, the results generally showed low random error over repeated assessments, with the exception of the UG at the elbow. However, on balance the Mid-Point and 3-Marker approaches were proposed as the most reliable.

Typical Error	Intra- Rater (Averaged)		Inter- Rater	
	Knee	Elbow	Knee	Elbow
UG	**2.1°**	**2.4°**	**3.99°**	**7.19°**
No Markers	**1.34°**	**1.92°**	**2.44°**	**3°**
Parallel	**1.06°**	**1.77°**	**2.8°**	**3.4°**
Mid Point	**1°**	**1.64°**	**1.99°**	**2.1°**
3-Markers	**1.05°**	**1.86°**	**1.75°**	**2.16°**
4-Markers	**1.39°**	**2.15°**	**1.73°**	**2.46**

Table 5.33 Typical Error for all Examiners. Both the Intra-rater and Inter-rater Reliability Data.

The issue regarding the relative reliability between using adhesive skin markers and measurement techniques that do not, was raised in the literature review. A number of authors have reported validity and reliability limitations associated with the practice of placing adhesive skin markers for the purpose of osteokinematic measurements (Marks and Karkouti, 1996; Sati et al, 1996; Karkouti and Marks, 1997; Szulc et al, 2000). This current methodology did aim to provide further information on the reliability of this practice.

The reliability of measurements that used adhesive landmarkers was higher than conventional UG. However, all the approaches to the UM showed somewhat similar random error results, and did not indicate that adhesive markers had superiority. However, adhesive markers did control non-random error between the different examiners. Only the Mid-point approach and the two adhesive marker methods successfully control SB across both joints. Therefore, while adhesive markers can be useful, they should not be considered compulsory for controlling SB between examiners and they do not improve random error over repeated measurements.

5.5.2. Other Work

Although there are many published studies that have investigated the reliability of osteokinematic measurement tools, the possibility of carrying out in depth comparisons between those studies and this current work were limited. This is due to the fact that the data analyses in the majority of these studies have emphasised the use of correlation coefficients such as the ICC and PPMCC. No study review distinguished between analysis of non-random and random error for the reliability of any elbow measurement technique (see Table 1.4).

Marks and Karkouti (1996) assessed the reliability of a photographic technique of knee angle measurements using an analysis of 'method error' (ME), which appears to be the same calculation as the TE analysis used in this current study. Those authors based their measurements on lines drawn on a set of conventional photographs. Apart from their restriction to measuring outer ranges knee angles only (0° to 30°), their methodology was comparable to the 3-Marker

intra-rater reliability test in this current trial. They found that their measurements had an additional error of 1.35° when measurements were repeated over a 4-week period compared to measurements repeated at a 5-minute interval. That study report ME (TE) for the immediately (5 minute) re-applied skin markers as between 1.32° to 1.5° (average 1.41°), while the current study reported a TE of 0.75° to 1.34° (average 1.05°). This indicates that the 3-Marker approach to the Uillinn Method© had showed some improvement in performance (circa 0.35°) over a conventional photograph measurement system. The issues regarding time related changes in reliability were not addressed by this current study.

Russell et al (2003) investigated the reliability of the internet based goniometer (IBG) for the remote measurement of knee angles. That system was developed to allow a physiotherapist to gain a digital photograph of a joint in a location any distance from the subject, once both had access to the internet and an internet camera. The IBG acts in a similar way to the UM, however, Russell et al, did not explore different methods of applying their system to the human model, instead their system seemed based of a mix between the 3-Marker method and the No Marker method, as no actual adhesive markers were used, rather the subject was asked to point to the bony landmarks themselves. They used one healthy subject and 2 different examiners to assess 9 different joint positions, in a range from 3° to 97°. The agreement between repeated tests was assessed using LOA analysis. Results for the intra-rater reliability of the IBG displayed a mean LOA of -1.0°, 1.6° and a mean LOA of -1.0°, 1.4° for inter-rater reliability.

The No Marker approach to the UM, which is the most similar to Russell et als' (2003) procedure, did not show results as strong as those reported by that researchers. The current intra-rater trial had a LOA of up to –5.1°, 4.5° (Examiner 3), while the largest LOA spread for the inter-rater results was –8.4°, 3.8° (Examiner 2 Vs Examiner 3). The difference in results between the current work and Russell et al (2003) may be due to their limited numbers and uncertainties regarding their procedures (see 1.2.9, page 48).

Nigel-Keenan et al (2004) examined the intra-rater and inter-rater reliability of a number of lower limb joint angles using the UG. Two of their angles were the knee extension angle in lying and the popliteal angle, which is a measure of passive

knee extension with the hip flexed to 90° and is used clinically to measure hamstring length. They use 2 examiners for the repeated tests and 20 patient subjects. They analysed their data using the LOA test, however, they based their conclusions mainly on random error and did not sufficiently explore the influence of non-random error in their data.

The intra-rater results show a LOA of -12.3°, 12.3° for the Popliteal Angle and a LOA of -6.5°, 5.0° for knee extension. The current trials intra-rater results for the UG, when expressed in the LOA form rather than the RC configuration are from –4.6°, 4° (Examiner 1) to –6.1°, 7.2° (Examiner 2), which are in a similar band to Nigel Keenan's knee extension measurements. The much larger spread of the data for the Popliteal angle seems to indict it as having a particular poor intra-rater reliability for knee angle measurements.

The results for Nigel Kennen's inter-rater reliability test show that repeated assessments by 2 examiners had a LOA of -13.6°, 7.8° for the Popliteal Angle and a LOA -12°, 5.0° for knee extension. The current study used 3 examiners and therefore generated more LOA data. This ranged from –7.5°, 6.9° to -11°, 14.3°, which were again in a similar spread to those seen in the previous study for both the Popliteal Angle and knee extension.

The most appropriate comparison for elbow measurements from this current work is against Fish & Wingate (1985) who report the SD of the differences between examiners (inter-rater reliability, examiner=46, subject=1) for the UG as approximately 4°. From there published data TE may be calculated as approximately 2.8° and random error (1.96 * SD differences) of approximately 7.84°. The TE seen in this current study (examiners=3, subject=49) found that TE for inter-rater reliability of the UG was 3.75°, with a LOA of –6.9°, 14.4°. The comparative 1.96 * SD differences was 10.6°. In addition to these there was a SB between the different examiners of up to 6°.

These results imply the current study showed a poorer performance for the UG elbow measurements than those found by Fish & Wingate (1985). The methodologies and data analysis were somewhat different between the two studies.

5.5.3. Reliability in Different Joint Ranges

Another question raised in the literature review was whether or not measurements of a joint's outer range (relative extension) performed worse than measurements of its inner or middle range measurements (Rothstein et al, 1983, Enwemeka, 1986; Watkins et al, 1991; Marks and Karkouti, 1996; Karkouti and Marks, 1997; Brosseau et al, 1997, 2001). To address this question the data was divided into 3 sections relative to the range that the data was in, for both the knee and the elbow. The data was divided by the values given in the UG datasets. The criteria for the divisions were made in an attempt to keep similar numbers in the 3 groups. At the knee the 3 sections were measurements from -3° to 59° (outer) (n=17), 60° to 109° (n=18) (mid-range) and 110° to 150° (n=16) (inner). At the elbow the 3 sections were 20° to 59° (n=17) (outer), 60° to 99° (n=16) (mid-range) and 100° to 140° (n=16) (inner). The TE of these three datasets were compared and the results were graphed for each of the measurement techniques.

UG measurements at the knee were identified in the literature as have poorer measurements of joint extension (Rothstein et al, 1983; Brosseau et al 1997, 2001). However, the TE for repeated measurements by the same examiner and between the different examiners in this current work showed no evidence to support this notion. This was also reflected in the elbow data. In fact, rather than increased TE in the outer range, there may have been a trend towards increased TE for mid-ranges (see Table 5.34, page 244; Appendix P).

There was minimal increase in TE for the outer range of the Mid-Point and 3-Markers approaches to the UM for both the intra-rater and inter-rater reliability data. However, the size of these increases were small and inconsistent between examiners and therefore could be attributed to random error rather than an association with the particular joint range (see Table 5.34, 5.35, Appendix P).

Further evidence to dismiss the notion of poorer performance of the UG, or any of the other measurements, for joint extension is provided by the scatter plots that were drawn up to illustrate the LOA. If there was further random error in joint extension the plots would show a heteroscedastic spread of the data, this was not seen. The homoscedastic nature of the data indicates that the magnitude of the

measurement value is not related to the amount of random error in the measurement (see Graph 5.1 to 5.36, Appendix F).

Knees	Intra Rater Reliability - Average TE					
	UG	No Markers	Parallel	Mid-Point	3-Markers	4-Markers
Outer (0-59) (n=17)	**2.13°**	**1.32°**	**0.99°**	**1.19°**	**1.18°**	**1.32°**
Mid-Range (60-109) (n=18)	**2.44°**	**1.29°**	**0.93°**	**0.91°**	**1.1°**	**1.42°**
Inner (110-150) (n=16)	**1.49°**	**1.37°**	**1.17°**	**0.85°**	**0.82°**	**1.36°**
Knees	**Inter Rater Reliability - Average TE**					
Average	UG	No Markers	Parallel	Mid-Point	3-Markers	4-Markers
Outer (0-59) (n=17)	**3°**	**2.25°**	**3.15°**	**1.74°**	**1.61°**	**1.9°**
Mid-Range (60-109) (n=18)	**4.42°**	**1.53°**	**1.69°**	**1.47°**	**1.38°**	**1.53°**
Inner (110-150) (n=16)	**2.58°**	**2.08°**	**2.69°**	**1.91°**	**1.4°**	**1.4°**

Table 5.34 Average Typical Error Between all Examiners Different Ranges of the Knee.

Elbows	Intra Rater Reliability - Average TE					
Average	UG	No Markers	Parallel	Mid-Point	3-Markers	4-Markers
Outer (0-59) (n=17)	**2.48°**	**2°**	**1.51°**	**1.68°**	**1.92°**	**2.39°**
Mid-Range (60-109) (n=16)	**3.04°**	**1.89°**	**1.79°**	**1.45°**	**1.61°**	**2.07°**
Inner (110-150) (n=16)	**1.8°**	**1.89°**	**1.93°**	**1.66°**	**1.92°**	**1.88°**
Elbows	**Inter Rater Reliability - Average TE**					
Average	UG	No Markers	Parallel	Mid-Point	3-Markers	4-Markers
Outer (0-59) (n=17)	**2.97°**	**2.58°**	**2.74°**	**1.82°**	**2.8°**	**2.61°**
Mid-Range (60-109) (n=16)	**3.74°**	**2°**	**2.32°**	**1.9°**	**2.45°**	**2.88°**
Inner (110-150) (n=16)	**2°**	**2.76°**	**2.91°**	**1.68°**	**1.95°**	**2.14°**

Table 5.35 Average Typical Error Between all Examiners Different Ranges of the Elbow.

The inter-rater reliability results were consistently worse than the results for the intra-rater reliability (see Table 5.26). This is a feature reported by the majority of previous researchers (see Table 1.2 to 1.5).

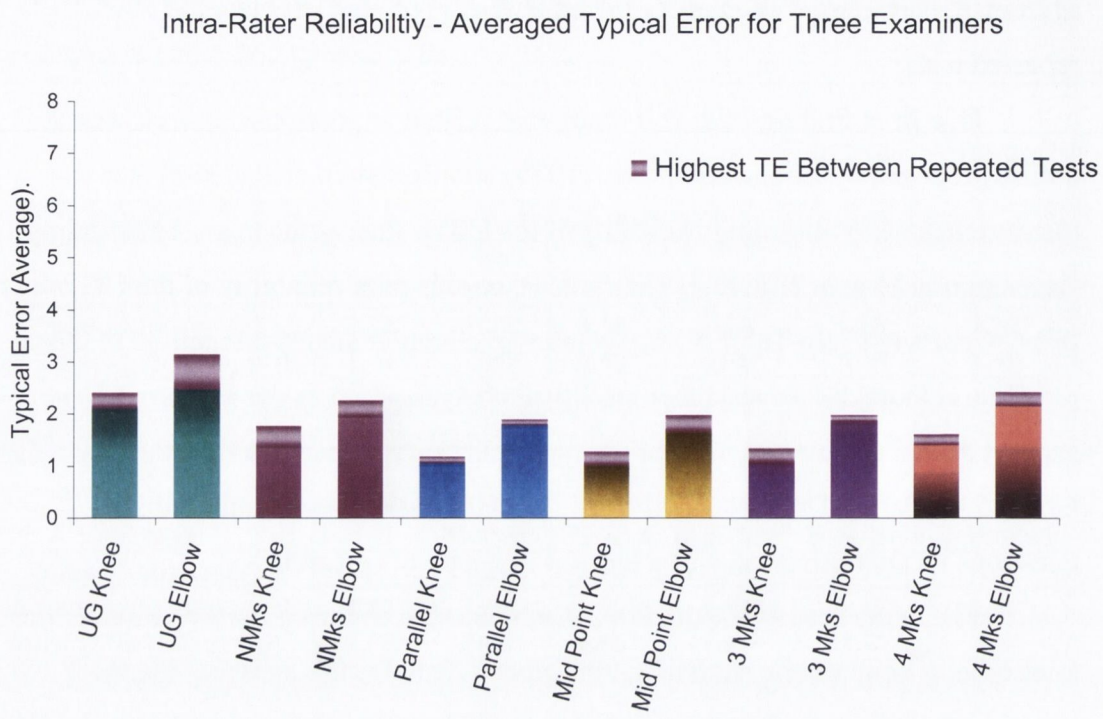
The fact that there were large variations in non-random error between examiners (see Graph 5.45) clearly shows that the individual approach of each of the examiners measurement technique is a major contributing factor to the poorer

agreement between examiners compared to the agreement between their own repeated tests.

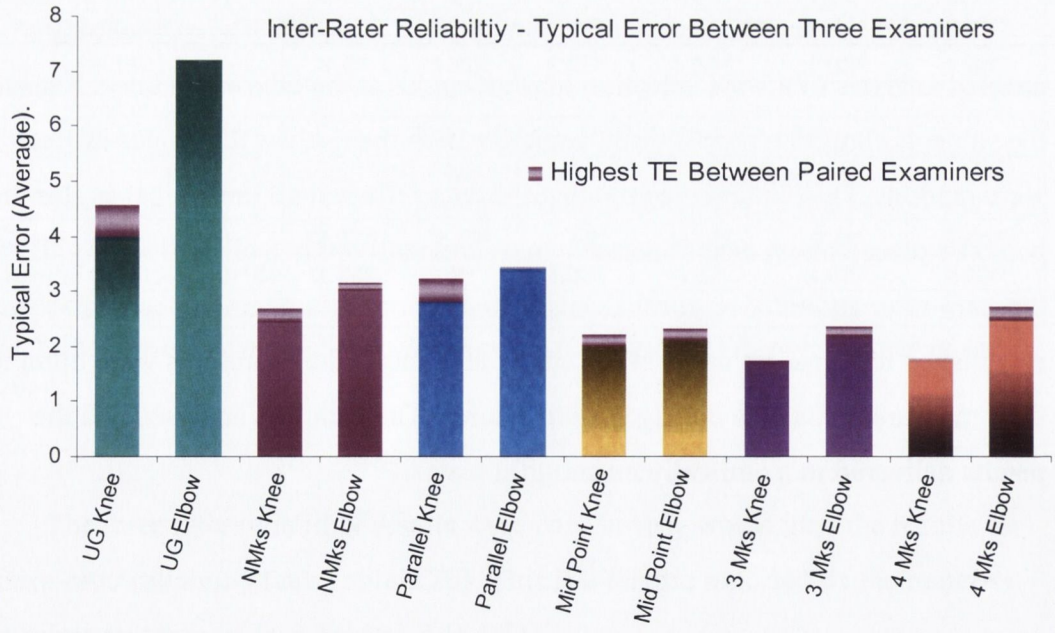
One final finding from this study that related to previous research was the notion proposed by Boone and Azen, (1979) and Rothstein et al (1983) that measurements have stronger reliability at the elbow than at the knee. This finding was supported by the ICC (3,1) results for the inter-rater reliability of the UG only in this current work (see Table 5.26), which was a similar statistical method to those previous authors. However, other statistical analysis, such as the TE data, do not support such a notion and showed stronger reliability at the knee than the elbow (see Graph 5.49, 5.79). The fact that the ICC (3,1) confused these results further questions its appropriateness as a tool for this type of research.

The better results for the lower limb over the upper limb may be due to the knee having longer bone shafts than the elbow. This feature may provide the examiner with better reference points from which to take their measurements. In addition the supination and pronation movement of the forearm may produce changeable reference points for elbow measurements which may negatively effect reliability.

All of the proposed aims of the study were achieved. The methodology enabled analysis of a wide selection of joint angles at the elbow and knee. Because the subjects chose the precise joint angle position themselves the angles assessed were random. This feature ensured that the data reflected all the typical angles that could be placed on an osteokinematic measurement tool in a clinical setting. Both the intra-rater and inter-rater reliability of 6 different measurement techniques were established for measurements at the elbow and knee. The examiners were blind to their measurements data during the procedures. The data was analysed and the results delivered in a clinically meaningful way.



Graph 5.49 Intra-rater Reliability: Average Typical Error with the Highest Observed Error for all Measurement Techniques



Graph 5.50 Inter-rater Reliability: Typical Error with Highest Observed Error for Paired Measurements.

5.5.4. Limitations of the Study

The author was one of the 3 examiners employed by the study (examiner 3). Using the author to take measurements was done for practical reasons. A big commitment in personal time was needed by each of the examiners which made recruitment of professional staff difficult. It was recognised that some readers may suspect the author would favour one of the UM approaches over the UG on the grounds possible personal reward and therefore open the study to criticism on the grounds of examiner bias. However, great efforts were made to blind all the instruments (UG and all UM approaches) during the examinations making it impossible for one examiner to influence results. In addition, inspection of the results do not support the presence of a single examiner bias for the UG or any of the measurements. There is much fluctuation in TE scores for intra-rater reliability at both joints (see Table 5.1 to 5.12). While examiner 3 did have the greatest TE scores for the UG at the elbow, examiner 2 had the greatest scores for this device at the knee.

Inter-rater reliability scores also do not support favouritism to the UM over the UG for any of the examiner pairings. The largest random error score between examiners for the UG at both joints did not involve examiner 3 (the author), but were between examiner 1 and examiner 2 (see Table 5.29, 5.30, page 237). The non-random error data show that examiner 3's UG results had the closest agreement to examiner 1, which was the general trend for all approaches (see Graph 5.45, page 236). In short, scrutiny of all results support the view that bias by one examiner was not a feature of this study.

The examiners were not given instruction towards the use of the standard UG. This was avoided because each of the examiners were over 2 years qualified and had call to use the UG on a regular basis as part of their work, hence typical professional skill was assumed. The objective of the study was to relate its finding towards routine clinical practice, therefore, controlling the examiners techniques may have artificially influence their normal performance.

It is acknowledged that there were some limitations to the study. The fact that only 3 examiners were used to conduct the examinations was one limitation. While the statistical degree of freedom was high due to the number of test subjects

used, the small examiner number may reduce the inferences regarding the reliability of the measurement techniques to the general population of physiotherapists. However, this study design, of few examiners and large numbers of subjects is a similar one to those used in the majority of previous studies (see Table 1.2 to 1.5, page 22 to 26) and is endorsed by Hopkins (2000a). Nonetheless, the author feels that a larger number of examiners would have made the results more robust.

Another limitation is the fact that the study only repeated measurements on two joints, the elbow and the knee. Therefore, the results can only be inferred to those joints (Stratford et al, 1984; Fish and Windgate, 1985; Riddle et al, 1987; Hayes et al, 2001, (1.4.1, page 61)) and will provide only limited knowledge regarding measurement of other joints, especially small joints, such as those in the hand. Furthermore, both joints assessed had predominantly uni-planar motion which may have further limited inferences to the measurement of large joints that are multi-planar, such as the shoulder or hip.

As discussed previously there are pros and cons to using healthy subjects in a reliability trial (see 1.4.3, page 64). Based on this particular study design, which included a long period on inactivity, it was felt that using a healthy population was most appropriate. Subjects were asked to remain in one position, without any movement for up to 20 minutes. It was felt that it would be unreasonable to ask a patient, who may have an uncomfortable or painful joint, to partake in such a task. It was supposed that a patient population might present a higher risk of movement during the examination procedures (Grohmann, 1983; Nicol, 1989), which would add unnecessary noise to the measurements and blur the results due to increased biological error (Burton et al, 1999; Hopkins, 2000a). It is recognised that there may be some limitations inferring these results to a patient population (Rheault et al, 1988).

The use of a patient versus a healthy population presents a difficult question for a researcher designing a reliability trial. One reason for poor inferences between these two subject groups may be that a patient population will have less available joint ranges than their healthy counter parts. Therefore a trial that used end of range (EOR) movements only on a healthy population will not have tested their equipment appropriately. This current methodology used a wide range of joint angles and therefore may have avoided that particular limitation. Although it is acknowledged

that elbow measurements in this study did not go below 20°, which was due to the random choices made by the subjects.

It may be argued that using adhesive markers on a patient population may be substantially more unreliable than using them on healthy subjects. It has been proposed that identifying bony landmarks can be difficult, but especially difficult on over-weight subjects (Marks and Karkouti, 1996) or subjects with bony deformities. Most of the healthy subjects in the current study were not over-weight, nor did they report any orthopaedic abnormalities. As one of the risk factor for joint dysfunction is obesity (Aoyagi et al, 2002), this implies that it is likely that there is a greater ratio of over-weight people in a patient population seeking physiotherapy intervention than in a healthy population. It is not clear from this current study how this extra challenge would be met by any of the joint angle measurement methods, especially the approaches that used adhesive skin markers.

Another relevant issue not dealt with in this study and apparently absent in the literature, is how the size of markers effects the reliability of photographic based measurement. It seems logical that it may be more difficult to be precise when placing larger markers rather than smaller ones. The marker size may also effect the placement of the measurement arms on the photograph. Therefore, this issue is worthy of further investigation.

One of the principle reasons for choosing healthy subjects over patient subjects, as stated above, was to minimise the possibility of subject movement during the repeated measurements, which were to take up to 20 minutes per joint. That equated to the subject being involved in the study for up to 40 minutes. During those periods it was emphasised to the subjects that they must remain perfectly still, naturally, this proved to be a difficulty for some subjects.

When the measurements were inspected, 4 out-lyers in the elbow data were detected. The photographs for these repeated measurements were scrutinised using the Adobe® Photoshop® 7.0 [Adobe Systems Inc., Minnesota, USA]. That software was used to 'overlay' one photograph over the other and check for actual movement by the subject, or to verify if the out-lying data was a real feature of the measurement systems. The paired photographs were merged, re-sized and rotated

until the posterior surface of the elbow and forearm from both photographs were aligned. The orientation of the images were checked by other static features in the photographs, such as the line of the plinth and the subject number label. The overlay images (see Figure 5.10, 5.11, 5.12, 5.13) clearly show gross movement by the subjects between their repeated assessments. For this reason the data from these 4 subjects were excluded from the calculations of reliability for all elbow measurement techniques. This meant that while 49 elbows had been assessed only 45 were included in the data analysis. It was felt that less movement had occurred during the second round of measurements and it was for this reason the data from that the 2nd round of measurements were used to calculate the inter-rater reliability results.

Only these 4 subjects, who had large irregularities in their measurements, were investigated using the photographic overlay method. It is likely that other subjects also displayed some movement, of a lesser magnitude, between the repeated trials. Such movement was attributed to unavoidable biological error and accepted as inevitable to the testing procedures (Bovens et al, 1990; Hopkins, 2000a).

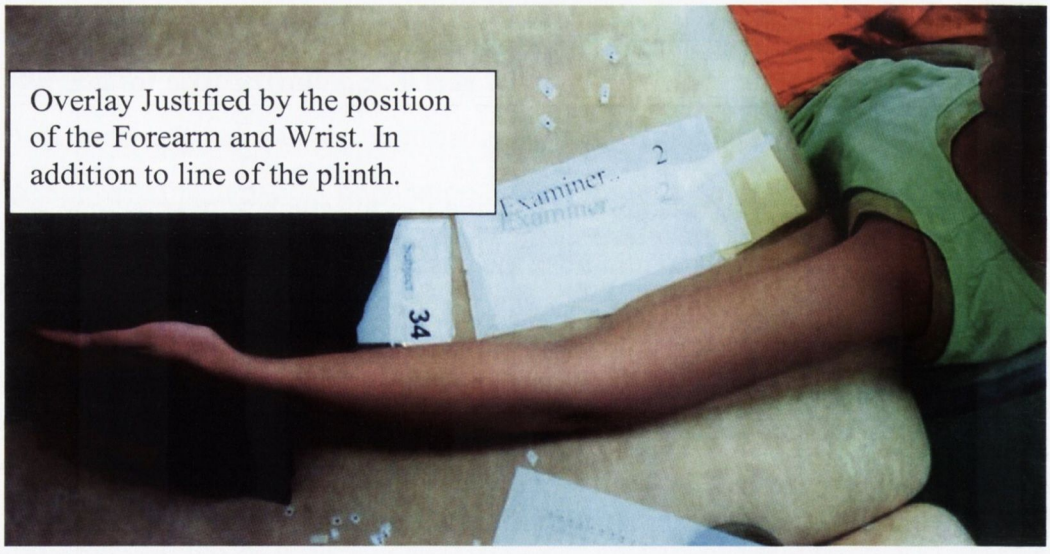


Figure 5.10. Overlay of Photographs, Movement of Subject 34.

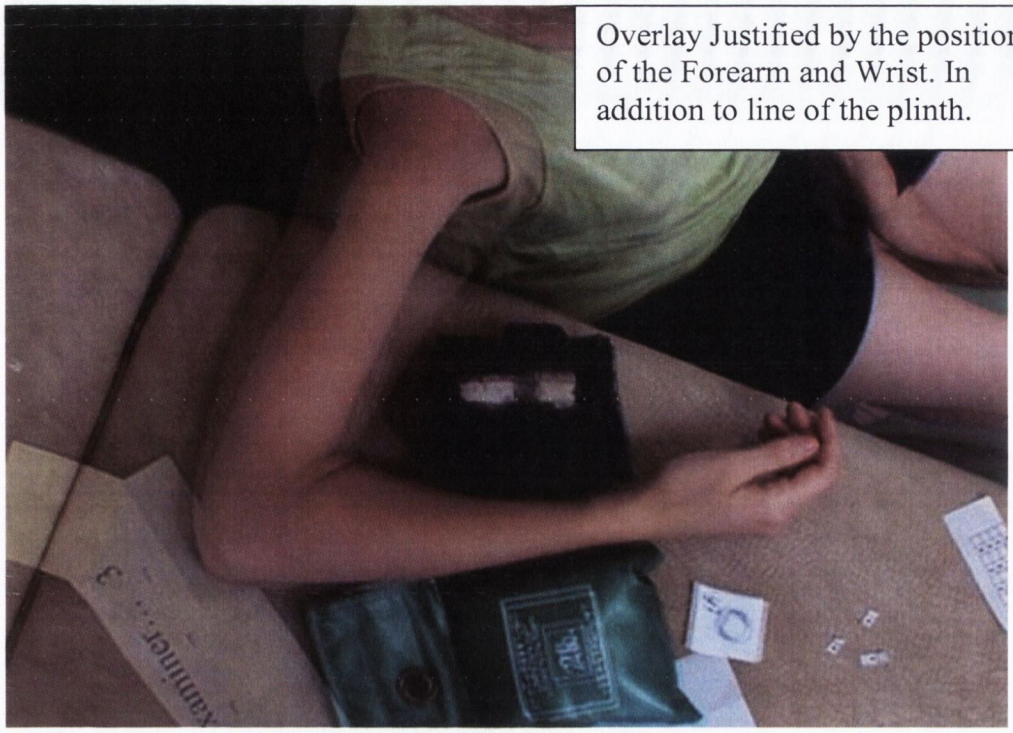


Figure 5.11. Overlay of Photographs, Movement of Subject 50.

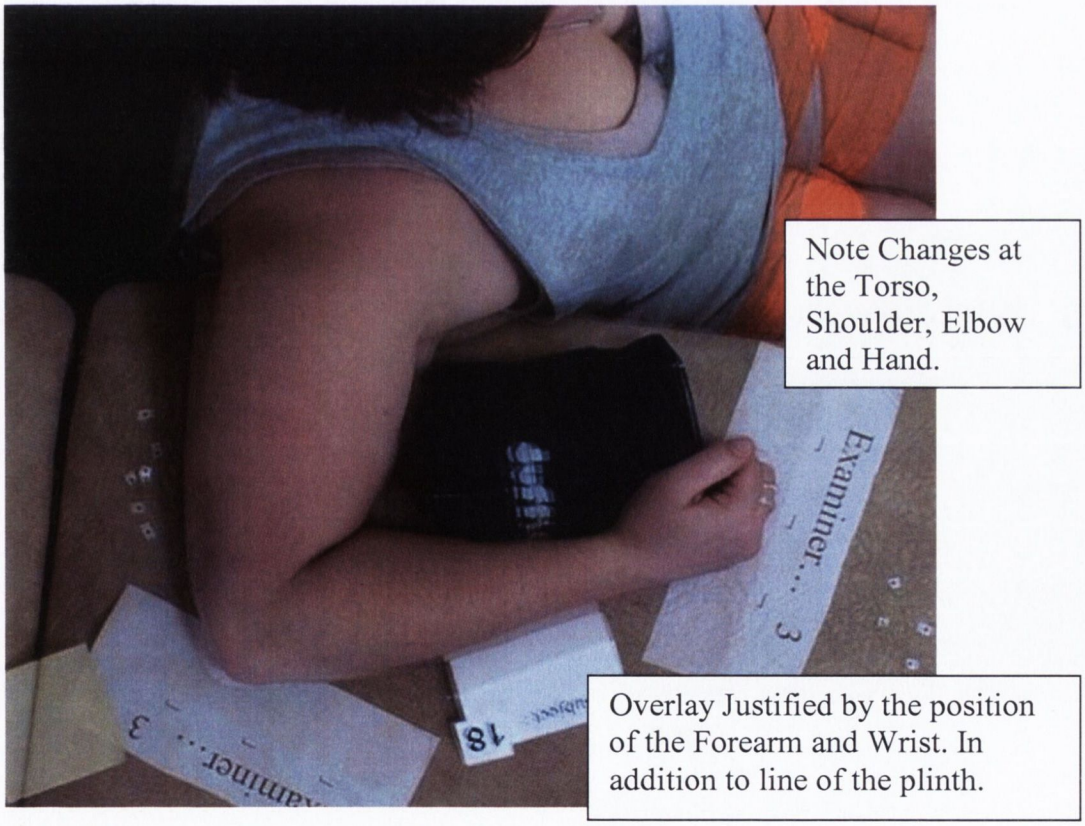


Figure 5.12. Overlay of Photographs, Movement of Subject 18.

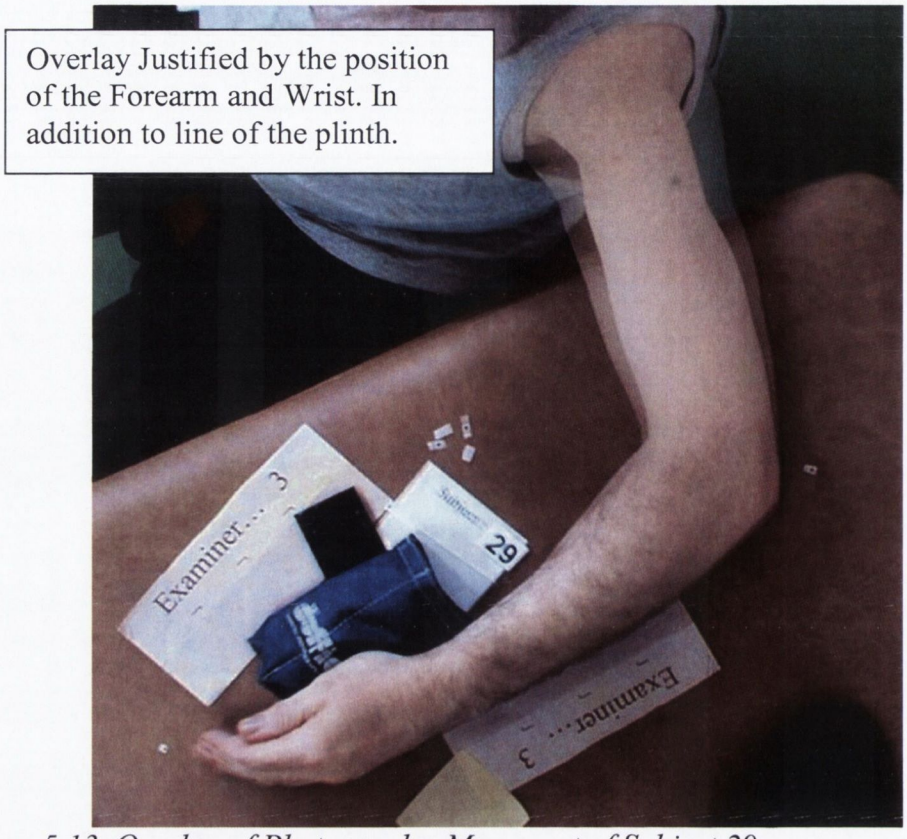


Figure 5.13. Overlay of Photographs, Movement of Subject 29.

In summary, this study upholds the view that measuring human movement is a challenging science. No joint angle assessment method is perfectly reliable, and this situation is unlikely to change (Lantz et al, 1999; Rothstein, 2001). The reason for this limitation is due to the presence of uncontrollable 'noise' which is inherently associated with human measurement (Hopkins, 2000a).

The movement of the subjects illustrated in Figures 5.10, 5.11, 5.12, 5.13 is one example of these difficulties. This was present despite every effort being made to minimise biological variations in the methodology and thereby isolating the reliability tests to the criteria of both examiner skill and the measurement tools.

There is evidence that all types of measurement noise are present in the data including biological variations, equipment error and examiners error. These are features that are present in a clinical environment (Armstrong et al, 1998; Burton et al, 1999; Hopkins, 2000a) and therefore were expected in this current study.

The results found that 2 approaches to the UM, the Mid-Point approach and the 3-Marker approach had the most versatile reliability, i.e. they both had low random and non-random error for both joints when assessments were made by the same and different examiners.

Based on the poorest observed results of typical error (TE) for these methods, when the same examiner conducts the examinations the TE will be of the order of 1.3° at the knee (see Table 5.28; Graph 5.45, page 234) and 2° at the elbow (see Table 5.31; Graph 5.46, page 234). These methods carry no systematic bias (SB) between examiners and the typically (TE (multi-variant)) examiners will by less than 2° at both joints (see Table 5.26, page 227).

In comparison the UG showed that TE for the same tester was 2.4° at the knee and 3.14° at the elbow. There was a substantial SB between examiners of up to 6° (see Graph 5.45, page 234), in addition TE between examiners was an extra 4° at the knee and 4.13° at the elbow.

From these results there is clear evidence that the UM offers the clinician an improvement in the reliability of joint angles measurements over the currently used UG. Furthermore, the study has successfully provided the clinician with tangible data regarding the expected errors (both random and non-random) for the UG at the elbow and the knee.

This type of improvement in clinical measurement may be highly valuable to clinicians and researchers alike. The improved error levels associated with repeated measurements made by the UM should allows us to improve our confidence in measurement data and make research both more sensitive and accessible.

The main reason for the limitation of comparisons between this work and older studies was the fact that the majority of previous research based their findings on analysis by correlation coefficients. As previously discussed (see 1.4.4, page 66), this type of analysis is largely insufficient (Daly and Bourke, 2000), therefore conclusions drawn between this current work and previous studies may be flawed. The current data analysis was exhaustive and fulfilled its aim of presenting the data in a clinically meaningful way.

A standardised approach to data analysis may help to maximise information obtained from this type of research. With this in mind it there may be a need to make recommendations regarding specific methodologies or areas of research to clinicians. One avenue to achieve such a paradigm is for relevant professional bodies to become involved and orchestrate research into wanton areas. This it may help to avoid some of the limitations to future research, enable better comparisons

between studies and therefore promote the development of the physiotherapy profession.

5.6. Conclusion.

The Uillinn Method© provides the user with a virtual goniometer with which to measure a relevant joint angle. The performance of this method on a human population had to be ascertained if this method is to be recommended to clinical practice. The aim of this set of experiments was to provide knowledge on the reliability of the UM for static measurements of sagittal knee and elbow angles. The study also assessed the reliability of the UG and provided information on the reliability of re-applying adhesive markers.

Reliability is a core consideration to osteokinematic measurement science. The concept of reliability deals with the ability of a measurement to provide consistency during repeated measurements. Both intra-rater reliability (repeated measurements by the same examiner) and inter-rater reliability (repeated measurements by different examiners) were assessed in this study.

The results show that the universal goniometer (UG) had the poorest results for both random and non-random error. The most favourable results were displayed by the 3-Marker approach and the Mid-Point approach. The former approach uses adhesive skin markers, while the latter approach guides the examiner to find the centre point of the limb at any point on the limb segment.

The expected TE for the Mid-Point approach to the UM at the knee was 1.3° (intra-rater) and 1.5° (inter-rater), at the elbow these scores were 2° (intra-rater) and 2.1° (inter-rater). There was no systematic bias between the examiners measurements.

Other findings from the study refute claims by some authors that measurements of the elbow are generally more reliable than measurements of the knee (Boone and Azen, 1979; Rothstein et al, 1983). The results also disagree with other researchers who have claimed worsening reliability in the outer joint range at the knee, than middle or inner ranges (Rothstein et al, 1983; Watkins et al, 1991; Brosseau et al, 1997, 2001).

Finally, the results show that while the practice of anatomical landmarking for photographic based measurements improves upon the reliability of the UG, they are not necessary to achieve optimal reliability when using the UM. In addition reliability may be seen to decrease with an increased number of markers. Similar

reliability can be gained without the use of skin markers using the Mid-Point approach to the UM.

The study provided valuable new knowledge regarding the reliability of the UG for elbow and knee measurements. In terms of TE, the intra-rater reliability of this device at the knee was 2.1° and at the elbow TE was 2.4°. The inter-rater reliability (TE) for the UG at the knee was 4° and 7.2° at the elbow.

Chapter 6.

CONCURRENT VALIDITY OF UNIVERSAL GONIOMETRY AND THE UILLINN METHOD©.

6.1. Introduction.

The preceding work has described the development and operation of the UM, established the sensitivity of it's angular calculations and tested the reliability of different approaches of its measurement procedures. In addition the reliability of the UG has been tested in order to provide better knowledge about that tool, which is the most widely used tool in clinical practice today (Coxford et al, 1998; Burton et al, 1999). This Chapter deals with issues regarding the validity of both the UM and the UG.

LoBiondo-Wood and Haber (2002) define validity as a determination of whether a measurement instrument actually measures what it is purported to measure. They state that there are 3 major kinds of validity; content, criterion-related and construct validity. Within these types of validity there are further subcategories such as; face, concurrent, predictive, convergent and divergent validity. An investigation that compares objective measurements of the same entity, made by different measurement tools (one of which is a gold-standard), when the measurements are made at approximately the same time is an investigation into concurrent validity. This is the strongest type of validity (Norkin and White, 2003) and is a subcategory of criterion-related validity.

Norkin and White (2003) and Altman (1999) state that in order to properly test the validity of a tool that is used to determine objective clinical data, it is best to assess the agreement between it and a criterion or 'gold-standard' measurement. The best gold-standard available for comparisons with an osteokinematic measurement tool is a method that allows the researcher to map the exact orientation of the bony tissue under the skin, such as radiographic based calculation of the joint angle (Enwemeka 1986; Gogia et al, 1987; Brosseau et al, 1997, 2001; Rowe et al, 2001; Berryman Reese and Bandy, 2002; Sprigle et al, 2003; Norkin and White, 2003).

In short, the most appropriate study design used to establish validity of both the UG and the UM would be to make direct comparisons against calculations taken from X-rays.

A study design such as this has already been used by a number of previous researchers (Enwemeka, 1986, Gogia et al, 1987, Granberry et al, 1990, Paquet et al, 1991, Ishii et al, 1995, Brosseau et al, 1997, 2001) (see Table 1.6). The different

measurement techniques can be compared in terms of validity relative to each other and conclusions can be drawn regarding the most appropriate tool for measuring human joints. Similar to the reliability trial described in Chapter 5 the elbow and knee joints were investigated in this study.

The rationale supporting this work is based on the lack of clear knowledge regarding the validity of the commonly used UG (see Table 1.6, page 26). In addition, as the UM is a novel method of joint angle measurement it's validity had to be established. This should be done before clinicians or researchers would have confidence in it as a legitimate alternative to the UG.

As discussed previously there are at least 5 different approaches to the application of the UM and its virtual measurement arms (VMA) to the human model (5.1; 2.3). The validity testing would have to consider each of these approaches and establish which was the most trustworthy. As part of a methodology designed to examine all 5 UM approaches, the study would provide data regarding the validity of placing adhesive landmarkers of anatomical reference points, the study would attempt to interpret the validity of this practice.

6.2. Specific Aims of the Study.

1. To investigate concurrent validity of both the UG and the UM, for joint angle measurements of the elbow and knee by making comparisons with a radiographic procedure.
2. To report the validity data of the UM and UG with appropriate descriptive and statistical tools, including the Bland and Altman 95% limits of agreement and measurement of typical error (TE).
3. To compare between the performances of each of the 5 different approaches to the UM and the UG against the radiographic data and to elect the most valid measurement method for both the elbow and the knee.
4. To investigate whether applying adhesive skin markers to anatomical landmarks improves the validity of photographic based measurement.

6.2.1. Justification of the Aims.

Aim (1). To investigate concurrent validity of both the UG and the UM, for joint angle measurements of the elbow and knee by making comparisons with a radiographic procedure.

A number of studies reviewed have investigated the validity of knee measurements made using the standard UG and radiography as a gold-standard (Enwemeka, 1986, Gogia et al, 1987, Brosseau et al, 1997, 2001). However, similar to the published research on the reliability of the UG, the validity studies can be accused of a number of methodological limitations. These include low subject numbers (Enwemeka, 1986) and the use of healthy a population (Enwemeka, 1986, Gogia et al, 1987, Brosseau et al, 1997). In addition, all of these studies used statistical analyses that were inappropriate or incompatible with clinical interpretation (Daly and Bourke, 2000).

Furthermore, all of those studies mentioned used the knee as the focus of measurement. A Medline search performed in July, 2004, using the terms 'validity', 'goniometer' and 'elbow' did not provide any criterion based studies. Other studies have provided evidence that measurement tools, including the UG, will have different precision capabilities at different joints (Boone and Azen, 1979; Rothstein et al, 1983; Fish and Windgate, 1985; Bovens et al, 1990; Williams and Callaghan,

1990), which implies that inferences regarding the validity of the UG at the elbow drawn from its performance at the knee are limited and there is an urgent need to assess the concurrent validity of the UG at the elbow using a radiographic methodology.

Therefore, justification for this aim of the study is given by the fact that there is no evidence to validate of the UG at the elbow and limited evidence for its use at the knee. In addition, there does not appear to be criterion based evidence available for photographic based systems, such as the new UM. There is a need to recommend one of the different approaches to the UMs operation in order to avoid confusion and promote best practice toward its operation.

Aim (2). To report the validity data of the UM and UG with appropriate descriptive and statistical tools, including the Bland and Altman 95% limits of agreement and measurement of typical error (TE).

As mentioned above, one of the main reasons why it is difficult to extract useful information regarding the validity of the UG from the previously published research is due to the fact that the statistical analysis is focused on the use of correlation coefficients, which provides incomplete data regarding the performance of measurement tools (1.4.4). Therefore, the validity of measurement techniques needed to be revisited with different, more informative analysis tools, such as 95% LOA, the RC and TE (Bland and Altman, 1986; Altman, 1999; George et al, 2000). This analysis would reveal levels of random and non-random error between experimental and control data.

Aim (3). To compare between the performances of each of the 5 different approaches to the UM and the UG against the radiographic data and to elect the most valid measurement method for both the elbow and the knee.

The control data for this study would be gained from radiographic studies, while the experimental data would be gained from each of the 6 external measurement techniques (UG and 5 UMs). Because the study design would test the different methods under the same circumstances, it would be in a strong position to compare the performance of each of the methods. The method that has the strongest agreement and the lowest TE would be considered as having the best validity. It may

be that the best technique at the knee may not be the best method for elbow measurements and vice versa.

Aim (4). To investigate whether applying adhesive skin markers to anatomical landmarks improves the validity of photographic based measurement.

Controversy in the literature regarding the precision of using adhesive skin markers to improve joint angle measurements taken from photographs extends to the question of validity (Fish and Wingate, 1985; Marks and Karkouti, 1996; Sati et al, 1996; Hayes et al, 2001). Analysis of the agreement (LOA) and TE between angles gained from the radiographs and measurements based on adhesive skin landmarks may help to provide further evidence for the validity of the practice.

Specifically, the 3-Marker approach to the UM is based on the same model as the No Marker approach. They both use the same anatomical landmarks as the to perform their calculations. Therefore, if the 3-Marker approach had less validity error than the No Marker approach, it might be considered that adhesive landmarking techniques do improve photographic based osteokinematic measurements.

6.3. Methodology.

The study was designed in conjunction with the radiography department at the Adelaide and Meath Hospital incorporating the National Children's Hospital (AMNCH), Dublin, Ireland. The proposed study design would attempt to take lateral UG measurements and the digital photographs (both with and without adhesive markers) at approximately the same time as a lateral X-ray of the joints.

However, on consultation with a number of department radiographers it became clear that it would not be possible to take lateral external measurements of the knee at the same time and in the same position as the X-rays. This is because the standard procedure for lateral view X-ray films of the knee involve the patient lying on their side with the radiographic plate underneath their lower-most leg (see Figure 6.2, page 267). The roentgen rays are then exposed to the knee from the medial side, this beam gives exposure to the plate beneath the joint. Having the patient in this position would make UG and photography of the lateral aspect of the knee impossible. As these X-rays were to be drawn from a routine patient population, alteration of the standard X-ray procedures was not an acceptable option. A system was devised which attempted to immobilise the knee and allow the patient to roll from the supine position to the side lying position without changing the angle at the knee.

The standard radiographic procedures for assessment of the elbow exposes the medial side of the joint rather than the lateral aspect. The subject routinely sits beside a plinth with their elbow resting on it (see Figure 6.10), similar to the position used in the reliability trial (5.3.3), the X-ray plate is then put under the medial aspect of the joint and the roentgen beam is projected from the lateral side. The resultant X-ray portrays the medial side of the joint. This scenario provided an excellent position for the UG and UM measurements without requiring the subject to change position. However, it also presented a disparity between the X-ray view (medial) and that seen by the external measurements (lateral).

The radiography department in the AMNCH provided enthusiastic support, in the form of supplying the location, technical assistance and access to the study population. Approval for the study was granted by local ethics committee at the AMNCH.

6.3.1. Equipment.

An adjustable lower limb cast was fabricated to prevent movement of the patient's knee during the required change in position from side lying to supine. The materials use to construct the cast had to be radiolucent so as not to limit the quality of the X-ray. Two casts (one left and one right) were fabricated using thermoplastic material, moulded in 2 sections about the posterior aspect of the thigh and shank. The 2 sections of the cast were connected via a hinge mechanism made of plastic tie-wraps. A wooden fixing rod was used to ensure the knee hinge remained fixed once the specific knee flexion angle was decided upon, this fixing rod would be adjusted to a different position for each subject. Finally, a number of large Velcro straps were used to fix the limb to the cast (see Figure 6.1).

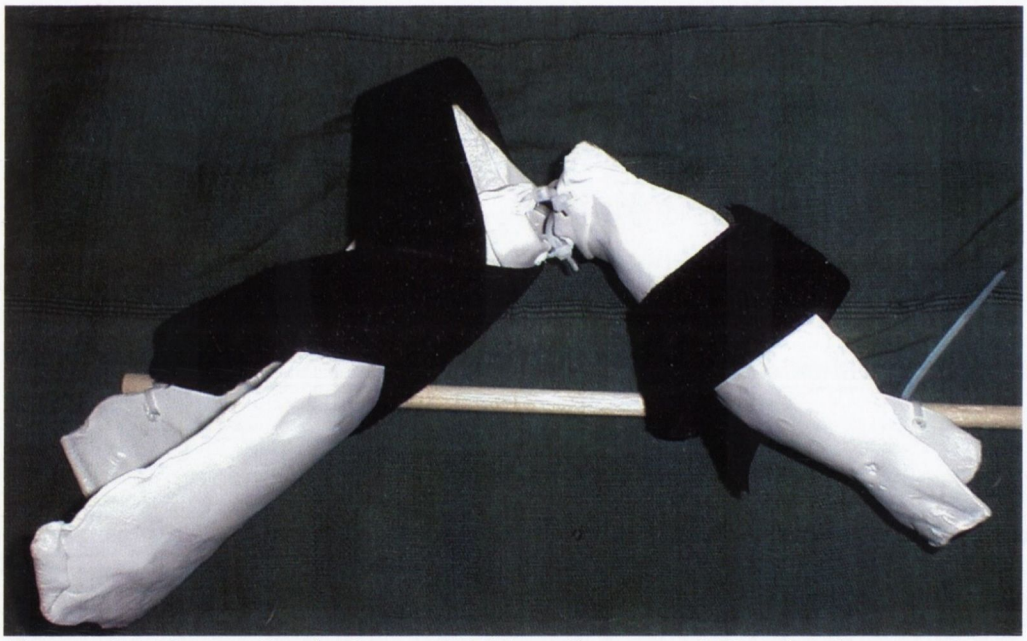


Figure 6.1. Lower Limb Cast with Velcro Straps and Wooden fixing rod.

The procedure for the application of the casts to the lower limb and the change in position, with the cast was in place, was piloted using a healthy female volunteer. From this pilot investigation it was discovered that the position of the wooden fixing rod had to be altered as the original position was uncomfortable to wear. It was also felt that the examiner should hold the limb with both hands during the move to provide feedback to the subject and encourage stability.

During the X-ray process standard radiographic safety techniques were observed, which included protecting the patient and staff from the potentially

harmful exposure of X-rays. This equipment included, lead clothing and protective screens.

When the X-ray was taken it was processed using the Picture Archiving and Communications System (PACS), which is a modern method of handling radiographs using information technology (IT). This system has numerous benefits to the end user, including software driven interfaces that allow image manipulation and software driven measurement tools, such as angular assessment tools (Oakley, 2003).

The external measurement procedures required that the entire limb be exposed, where necessary, sleeveless vests and loose fitting shorts were provided to the patient so that they felt comfortable at all times.

The photographic equipment used was an Olympus D-340L 1.3 Megapixel digital camera [Olympus Optical Co., London, UK]. The UM software (version 2.3, 2.4) was run on a laptop personal computer (PC) with SmartMedia card drive manufactured by SanDisk [Sunnyvale, CA, USA.] which was used to transfer the data from the digital camera to the PC (see Figure 4.1, page 137).

When all the measurements were made, the data was inputted into Microsoft® Excel 97 SR-1, spreadsheets [Redmond, Washington, USA], and when all the data was completed the spreadsheets were complied into one sheet (see Appendix H). That computer application and a statistical computer application, KyPlot©, 2000 [Koichi Yoshioka], were both used to perform statistical and descriptive analysis and to plot data graphs.

6.3.2. Subjects.

The patient subjects that were recruited for this study were those attending for routine radiographs of the knee or elbow at AMNCH, it was therefore a random sample of convenience. Part of the criteria set down by the study was that the participants limb had to be exposed in order to properly complete the external measurements. Therefore, those patients who were immobilised in a plaster-of-Paris (POP), back slab, were heavily strapped or had an objection the revealing their entire limb for measurement and photography were excluded.

The subjects had a wide range of complaints, including non-specific joint pain, recent trauma, osteoarthritis and joint replacement, internal fixation or other

surgeries. The diagnosis or suspected diagnosis was not readily available on the X-ray request form and patients were not quizzed regarding their diagnosis. Consequently the specifics of their conditions were not recorded and no patient was excluded on the grounds of his or her medical history or current diagnosis. The elbow and knee cohorts were assessed over different periods almost 1 year apart. The knee assessments were conducted first followed by the elbow trial.

A total of 43 knees were assessed (15 males and 28 females; 23 left, 20 right), in this group there were 2 repeat assessments and 7 subjects had both sides assessed during the same examination. The elbow group consisted of 50 subjects (31 males and 19 females; 22 left, 28 right), there was a total of 9 repeat assessments, and no paired assessments.

6.3.3. Procedures.

The receptionists in the radiography department initiated the recruitment procedure. They were asked to contact the examiner by an internal hospital bleep system when any appropriate patients attended for X-ray. When the examiner arrived in the X-ray reception area, they initially screened the referral forms for the X-ray procedure that was to be undertaken. If a sagittal X-ray view was requested the patient was approached for inclusion to the study. Other more complicated radiographic procedures that may have combined standard sagittal views, such as skyline or weight bearing views at the knee, were not possible in the X-ray room that was designated to the study and therefore these patients were excluded.

The subjects were approached informally and the examiners background was explained to them. The nature of the research was then described to them succinctly. If they were amenable to inclusion in the study, an information sheet was presented to the patients with more specific details regarding their participation in the study (see Appendix I). Each subject was given adequate time to read it and any questions they had were answered. If they remained agreeable to participation a written consent form was presented to them which asked for their name, the date and their signature (see Appendix J). Completed consent forms were kept as a record of the patients' agreement to be in the study. They were aware that they could terminate their participation at any time without consequence. Only when the subjects fully

understood their expected level of involvement in the study were they brought into the appropriate X-ray room to commence the investigations.

During the knee examinations, as soon as the patients were taken into the X-ray room they were asked to don a pair of loose fitting shorts. They were then assisted onto the X-ray examination table and requested to assume a knee position that was comfortable to them. The radiographers had requested that the knee be neither in full extension or more than approximately 90° knee flexion in order that a good lateral description of the knee was gained from the X-ray. This is standard radiographic procedure, in the AMNCH and is done in order to best visualise the anterior joint space.

When the subjects were in position the lower limb cast was applied to their leg and secured with two wide Velcro straps. The wooden fixing rod was then fastened using plastic tie-wraps (see Figure 6.3).

A total of 6 adhesive black dots, each with a diameter of 2 mm (see 5.3.3), were attached to the limb. Two were placed on the medial side (medial malleolus and upper medial thigh) (see Figure 6.2), and 4 on the lateral side of the limb. The medial skin markers were applied in order to further control and monitor knee flexion. The distance between these 2 dots was measured, recorded and monitored throughout the procedure in an attempt to further control unwanted movement during the change in position.



Figure 6.2. Medial View of the Knee with Adhesive Markers.

The 4 lateral markers were placed on the inferior aspect of the greater trochanter, the inferior tip of the lateral epicondyle, the superior tip of the head of the fibula and the inferior tip of the lateral malleolus, three of these point are described by Norkin and White (1995) as those that should be used as guides for knee flexion measurements, the exception being the head of the fibula (see Figure 6.3). These markers would be used for the 3-Marker and 4-Marker approaches to the UM.

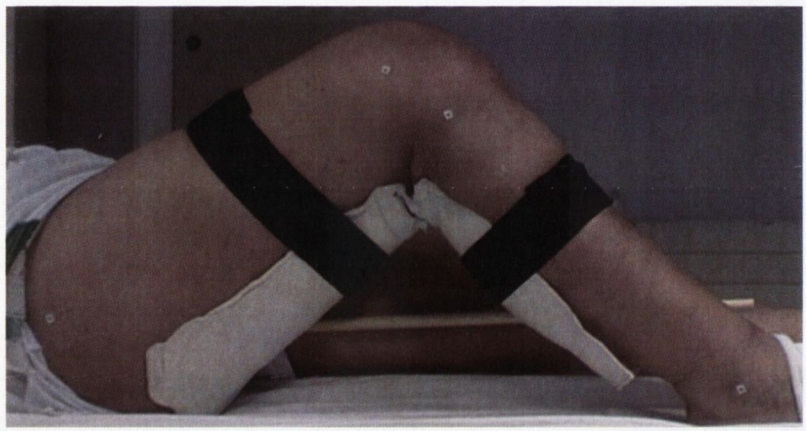


Figure 6.3 Four Lateral Limb Markers on the Greater trochanter, lateral epicondyle, head of the fibula and the lateral malleolus.

The importance of complete immobility of the knee during the examinations was explained to the patient. The patient was asked to practice rolling onto their side without moving the position of their bottom or heel. The examiner held onto the patient's thigh and ankle in order to stabilise and encourage immobility of these 2 limb segments. The roll was repeated a number of times if necessary until it could be performed without changing the angle of knee flexion. This was monitored by the distance between the 2 medial skin markers. Only when the examiner and the patient were happy with this procedure were the measurement procedures commenced.

A small sign was placed beside the patient containing the individual patient's medical record number, this was used to identify the patients from the photographs. The examiner and the digital camera were positioned so that the camera's lens was approximately centred on the posterior aspect of the knee joint and with a perpendicular distance of approximately 1 m. This was done in order to keep as much of the limb as possible central in the lens field of view. If more than one limb was assessed, or assessments were repeated this information was also recorded in the photograph.

The 4 lateral skin markers were then removed and the limb photographed again, from the same position. The knee flexion angle was then assessed and recorded using a UG. The distance between the 2 medial points was then re-measured to ensure that no knee movement had taken place. No subject displayed any movement during this part of the procedure. The entire measurement procedure, from the time the subject got on to the examination table, until the time the last photograph was taken, took less than 5 minutes.

An X-ray plate (Fuji IP Cassette 3, Fuji Photo Film Co. LTD. Japan.), of dimensions 24 cm X 30 cm, was then placed on the table beside the subject. The patient was then asked to roll on to their side, as practiced before hand, so as to rest the lateral aspect of their knee of the X-ray plate. A number of subjects required a small sponge wedge to be placed underneath their ankle in order to keep their knee comfortably against the X-ray plate. This is a standard practice for radiographers as it helps in obtaining a good lateral knee X-ray.

When the radiographer was content with their position and they were covered appropriately with protective equipment, the distance between the 2 medial skin markers was again re-measured in order to detect any movement that occurred during the positional change. The examiner had to make manual adjustments to correct the distance between the markers for 10 patients, for whom the distance had altered to no greater than 2 cm.

The examiner and radiographer stood behind the protective screen and the X-ray was taken. The X-ray procedure that was conducted was a standard lateral knee X-ray protocol. The film's focus distance was 100 cm and delivered at 55 kVp (kilo Volts) and 3.2 mA/s (milliamps per second).

In the case of one subject, after they had removed the lower limb cast and stood up, the radiographer discovered that she was not happy with the X-ray exposure as the 2 condyles of the femur were not superimposed directly over each other. The patient was agreeable for the entire procedure to be repeated. This resulted in 2 different knee flexion angles being available for one patient. Both measurements were dealt with individually and included in the data analysis. During the X-ray procedure, another patient was marked for possible exclusion as it was felt that he moved just prior to when the X-ray was taken. This patient did in fact provide out-lying data and was rejected on those grounds.

The elbow procedures were very similar to those of the knee, the patients were recruited in the same manner, from the same population. They consented to be included in the study in the same manner outline for the knee recruitments. They were then taken to the appropriate radiographic room. Here they were provided with a sleeveless tee-shirt, if they required it. Then they were seated on a chair beside the X-ray table, with their involved arm placed on the table in an abducted position (see Figure 6.10). An X-ray plate, similar to the type used for the knee was placed between the joint and the table. They were asked to slouch in the chair slightly and the table was then raised or lower to be in complete comfortable contact with the medial aspect of their upper limb. Their forearm was supinated and sometimes supported with a foam wedge at the level of their wrist. These are standard operational procedures for the X-ray department in the hospital.

The X-ray was taken first, to optimise the radiographers time involvement in the procedures. All standard safety procedures were followed as outlined previously. Once this procedure was complete, the first digital picture of the exposed upper limb was taken from the exact position of the X-ray beam. Adhesive skin markers were then placed on the acromion process, the lateral humeral epicondyle, the head of the radius, and the radial styloid. When these were in place, a second digital photograph was taken. Again from the same view point. Finally, the markers were removed and the UG was used to gain a measurement. The X-ray plate was then checked by the radiographer. The patient redressed, but was not released until the radiographer was happy with the radiographic picture.

At the conclusion of the measurement procedures, each limb examined had one X-ray, one UG measurement and two digital photographs (one with skin markers and one without skin markers). The digital photographs were downloaded on to a PC and sorted appropriately. After a period of at least two months (up to 5 months) were the 5 different approaches of UM used to assess each of the joint angles.

The part of the screen that displayed the angular data (see Figure 2.10, page 105) was covered with a card in order to blind the examiner to the calculated data. When the examiner was content with the position of the virtual measurement arms (VMA) the card was removed and the data recorded. The *No Marker* method (see

Figure 2.11, page 106) was the first approach to be calculated. The entire set of photographs (without markers) was assessed before calculations by the next approach were commenced. The *Parallel* method (see Figure 2.5, page 100) was next to be calculated using the same procedure, then the other set of photographs (containing markers) were used to calculate the *3-Marker* approach (see Figure 2.7, 2.8, pages 103) and the *4-Marker* approach (see Figure 2.9, 2.10, pages 104, 105). Analysis in this order ensured the examinations were not biased and that the examiner did not know the palpable position of the anatomical points.

A number of months after these assessments were done, the *Mid-Point* approach (see Figure 2.6, 2.15, pages 101, 110) was developed and used to test the photographs that did not contain the adhesive landmarkers. Again the section of the screen that contained the result was covered during the calculation to blind the examiner and prevent bias.



Figure 6.4 - The PACS System Being Used for Angular Assessment of the Knee.

6.3.4. Calculation of the X-ray data

The PACS system provides the user with 'measurement assistant tools' (MATs) which includes an angle calculation tool (see Figure 6.4). The measurement

arms of this virtual goniometer were aligned with the longitudinal axis of the femur and tibia at the knee, and the humerus and ulna at the elbow, much like the technique used by Brosseau (2001) (see Figure 6.16).

In an attempt to maximise the precision of the results assessments were carried out twice for most of the joints involved. Each of the repeated elbow assessments were separated by one week. There was over 1 year between repeated PACS assessments of the knee. The average value was taken for the true joint angle (see Appendix K). Only 22 out of the 42 knee subjects were assessed twice due to a technical issue.

When using the PACS angle calculation tool the results are displayed over the radiographic image at the apex of the angle which is a variable point on the screen (see Figure 6.4). This made it impractical to attempt to blind the calculation data from the examiner.

Approximately 3 weeks after the UM calculations were made, the same examiner assessed the criterion data for the knee flexion angle using the PACS X-ray system. The X-ray calculations at the elbow were not begun until approximately 3 months post the final UM assessment. Due to the substantial time between the UG, UM and PACS measurements and only medical record numbers were used to identify the patients in the X-rays, the examiner had no way of differentiating between the patients or even distinguish gender simply by looking at the subjects X-ray image. Therefore, examiner bias was not seen as a substantial threat to the data calculations.

As a quality assurance exercise and in order to ascertain the magnitude of expected error of the PACS system, a small study of both intra-rater and inter-rater reliability of the PACS systems was carried out at both the knee and the elbow. The intra-rater reliability test used the 22 repeated knee data and 50 elbow measurements. The inter-rater reliability calculations used a much smaller group of 7 knees and 7 elbow, randomly sampled from the volunteer subjects. These X-rays were assessed by both the main examiner (a physiotherapist - CGD) and a consultant orthopaedic surgeon (IK). The intra-rater reliability of the PACS system were compared to a similar study published by Sanfridsson et al (1997).

6.3.5. Data Analysis.

The data was similar to those described in chapter 1 (see 1.4.4). In summary, non-random and random error were examined using the 95% limits of agreement (LOA), the repeatability coefficient (RC) and typical error (TE). Other data analysis such as t-tests and correlation coefficients would be included in the study in order to allow some comparisons with previous studies.

6.4. Results.

A total of 43 knee patients were examined during the study. However, one subject was seen to move during the assessments and was excluded on these grounds as an outlier. Of the 42 remaining knee patients, the X-ray data revealed that the angles ranged from 25.9° to 94.43°, with a mean of 64.6° (± 15.7°).

No elbow subject was excluded after the assessments. The data for the 50 elbow subjects ranged from 63.2° to 112.6°, with a mean angular value of 89.5° (± 10.6°).

A value of 2.5% of the mean measured data represents the limit of SB for a given sample. For these datasets the average knee value was 64.6°, which set the limits of SB at 1.6°. At the elbow the SB value was 2.2° (average value 89.5°). If the mean of the differences between the PACS angles and each of the other measurement methods matched or exceeded these figures, it would be proposed that there was a SB between the two measurement techniques.

6.4.1. Reliability of the PACS system

Analysis of the intra-rater reliability of the PACS system showed that there was no SB between the repeated assessments. The RC was similar for both the knee and the elbow. This data revealed that 95% of repeated assessments would be within approximately 2.5°. Typically the difference between measurements would be less than 1° (see Table 6.1).

The limited data collected for the analysis of inter-rater reliability showed similarly favourable results at the elbow. Ninety-five per cent of measurements between examiner would be within 2.7° with a TE of again less than 1°. This implied that repeated measurements by different examiners agreed almost as closely as repeated assessments made by the same examiner (even with a limited degree of freedom). However, between examiner analyses of the knee showed larger disagreements. There was a SB of 2.2° indicating that CGD generally measured a larger angle than his colleague (IK). The random error between the two examiners was also higher than those observed at the elbow. For 95% of measurements the 2

examiners would agree to with 4° (LOA = -2°, 6°), the TE between them would be 1.5° (see Table 6.2).

These data suggest that angular measurements between examiners, based on radiographs are substantially more reliable at the elbow than at the knee, although this conclusion is based on a small sample size.

Intra-rater Reliability of PACS	Knee (n=22)	Elbow (n=50)
Average Difference Between Repeated Measures	0.54°	-0.18°
Standard Deviation of the Differences	1.41°	1.21°
Repeatability Coefficient (RC)	2.76°	2.38°
Typical Error (TE) of the Differences	0.99°	0.86°

Table 6.1. Intra-rater Reliability of the PACS system of Joint angle measurements.

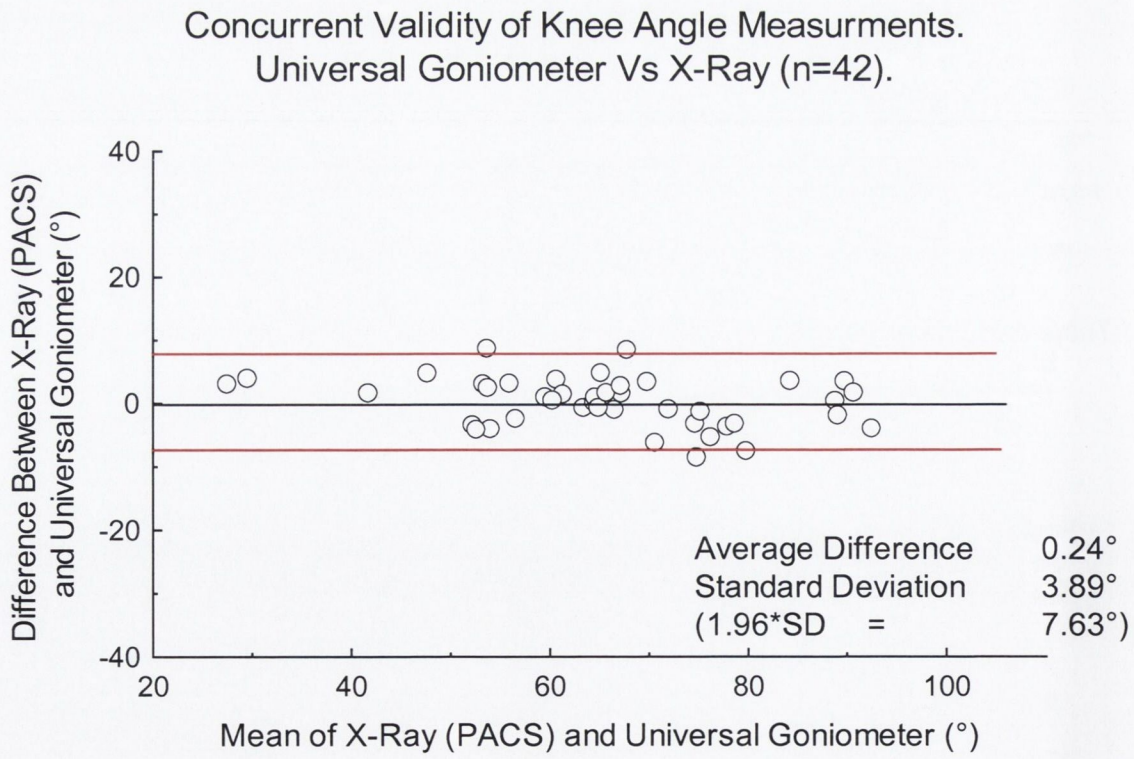
Inter-rater Reliability of PACS	Knee (n=7)	Elbow (n=7)
Average Difference Between Repeated Measures	2.19°	-0.3°
Standard Deviation of the Differences	2.06°	1.36°
Repeatability Coefficient (RC)	n/a	2.67°
Limits of Agreement (LOA)	-2°, + 6°	n/a
Typical Error (TE) of the Differences	1.45°	0.96°

Table 6.2. Inter-rater Reliability of the PACS system of Joint angle measurements.

6.4.2. Validity Results for Measurement of the Knee

Knees (Difference from PACS)	UG	No Markers	Parallel	Mid-Point	3-Markers	4-Markers
Maximum Range (from, to..)	8.52° -8.73°	8.27° -8.43°	10.07° -3.42°	11.91° -6.93°	5.85° -14°	10.98° -11.24°
Average Difference	**0.24°**	**-0.17°**	**-3.58°**	**-2.75°**	**1.75°**	**-0.89°**
Systematic Bias (SB)	*No*	*No*	*Yes*	*Yes*	*Yes*	*No*
Std. Dev. (SD)	3.89°	3.88°	3.63°	3.67°	4.77°	4.11°
Significant Difference Of Means (pair t-Test; p=0.05)	*No*	*No*	*Yes*	*Yes*	*Yes*	*No*
Pearson's Product Moment Correlation Coefficient (PPMCC)	0.97	0.96	0.97	0.97	0.95	0.96
Limits of Agreement (LOA) (SD *1.96)	7.63°	7.6°	7.13°	7.2°	9.35°	8.06°
Upper LOA	**7.87°**	**7.43°**	**3.55°**	**4.44°**	**11.1°**	**7.17°**
Lower LOA	**-7.39°**	**-7.77°**	**-10.7°**	**-9.96°**	**-7.59°**	**-8.96°**
Typical Error (TE)	**2.75°**	**2.74°**	**2.57°**	**2.59°**	**3.37°**	**2.9°**
SB Limit: 1.6° (Perfect sample: 1.75°)						

Table 6.3. Data Analysis for each of the Experimental Measures, Comparisons with X-ray (PACS) Measurements of the Knee.



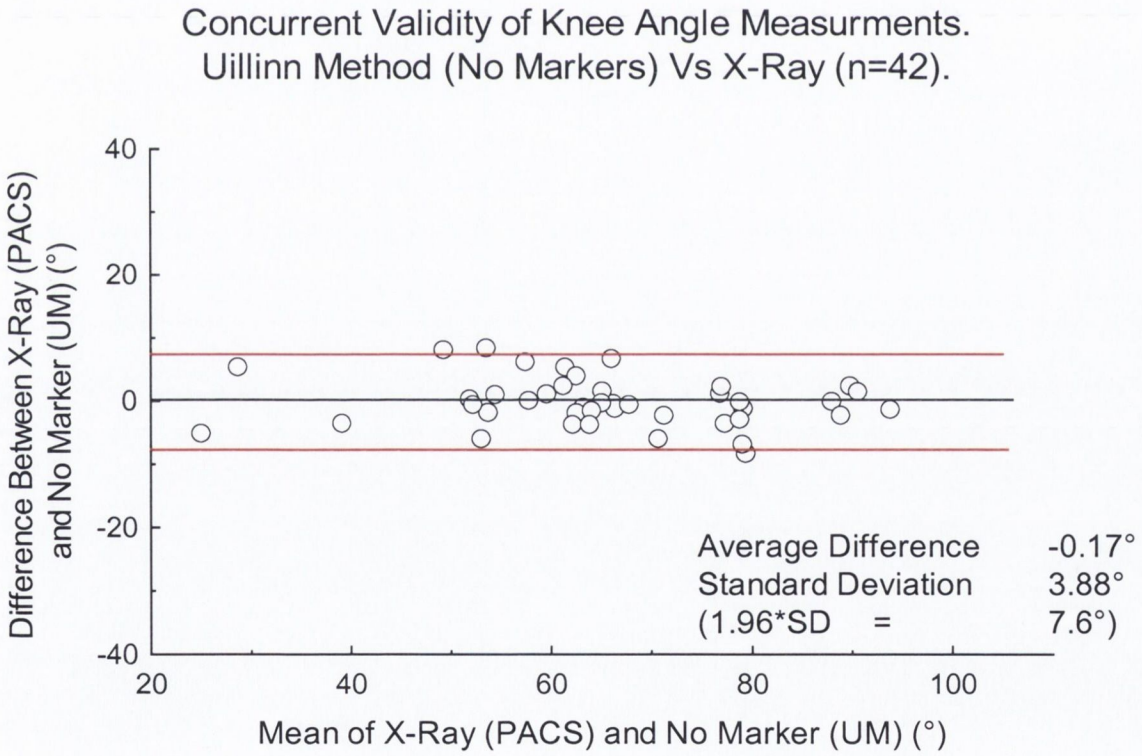
Graphs 6.1. Limits of Agreement between PACS data (X-ray) and Universal Goniometer for Knee measurements.

6.4.3. The Universal Goniometer (UG) at the Knee.

The average difference between the UG data and the radiographic measures was -0.24°, which indicated no SB between the two data sets. The random error between these two measurement systems as calculated by agreement testing (LOA) indicated that 95% of measurements would be within 7.8° of the angle calculated by X-ray (see Table 6.3; Graph 6.1), typically this error would only be 2.75°.

6.4.4. The Uillinn Method (No Markers Approach) at the Knee.

There was no SB between the PACS data and the No Marker approach to the UM. The random error data for this method was very similar to that of the UG. The TE was 2.74° and the 95% LOA would be in a spread from –7.7° to 7.4° (see Table 6.3; Graph 6.2).

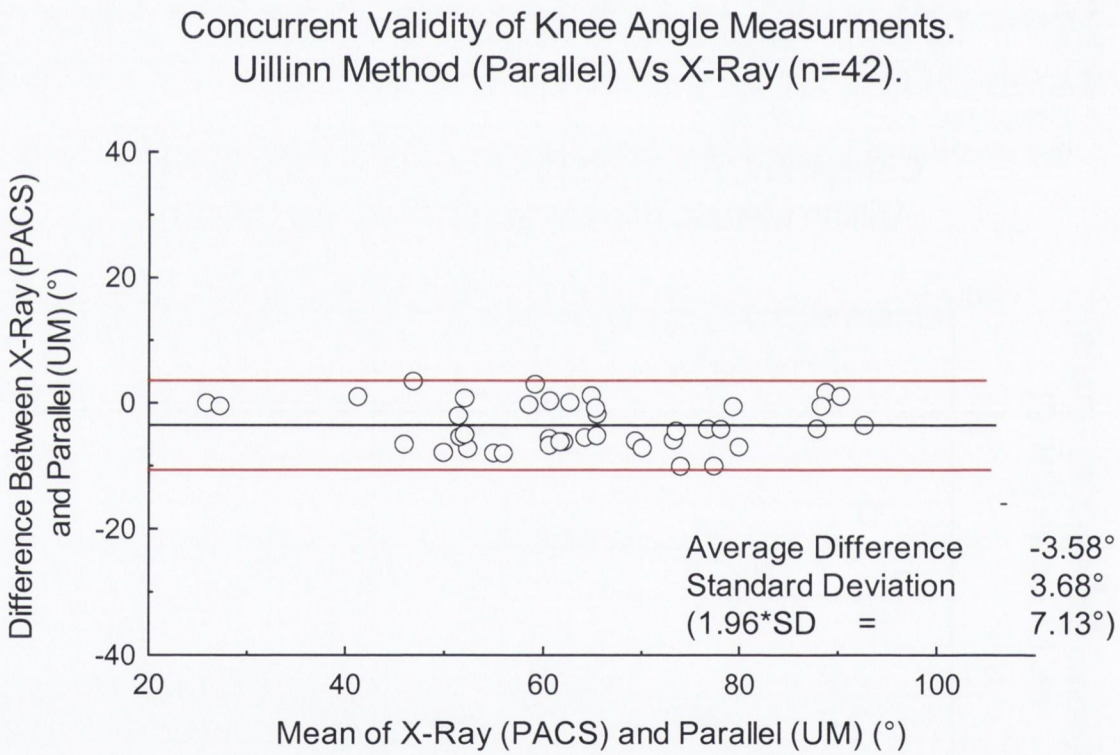


Graphs 6.2. Limits of Agreement between PACS data (X-ray) and the No Markers Approach to the Uillinn Method for Knee measurements.

6.4.5. The Uillinn Method (Parallel Approach) at the Knee.

The Parallel approach to the UM differed significantly (p=0.05) from the PACS knee measurements. This significance difference was reflective of a substantial SB of -3.6° between the cohorts. This reflects a generally lower measurement value by the Parallel approach in comparison to the PACS measurements.

However, random error between these two systems was low with a TE of 2.57° and a 95% LOA spread of –10.7°, 3.5°, which was the narrowest scatter of data over all the methods assessed (see Table 6.3; Graph 6.3).

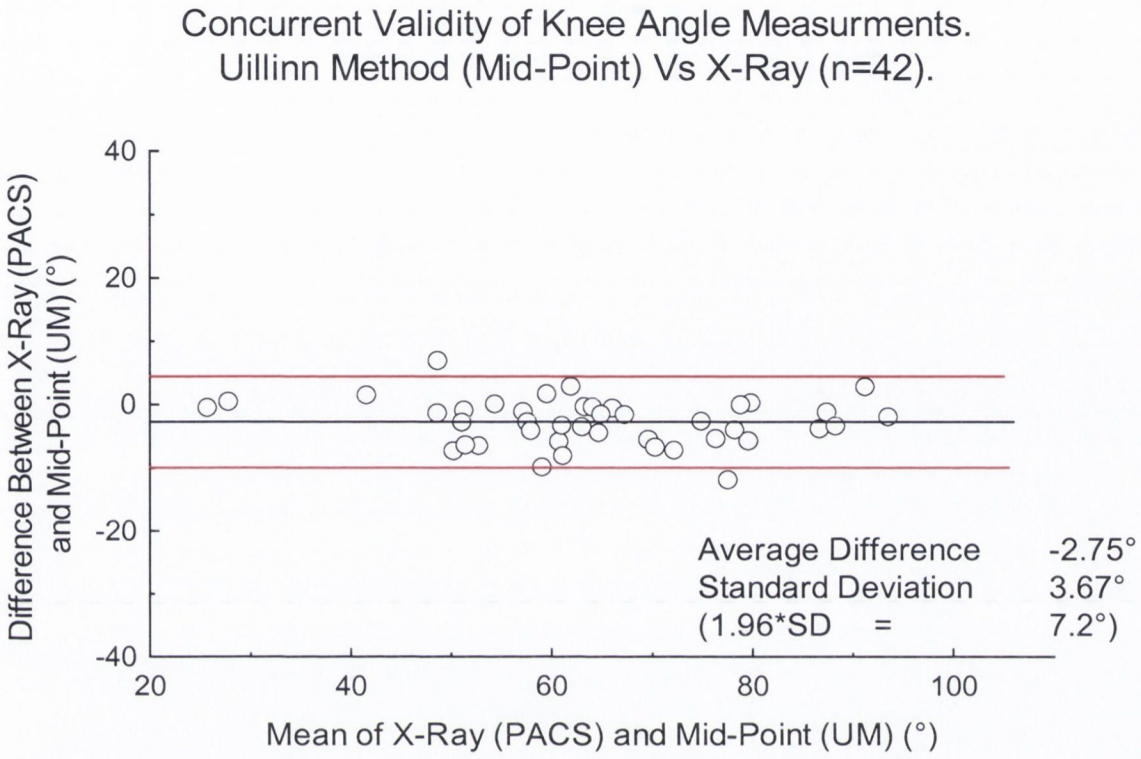


Graphs 6.3. Limits of Agreement between PACS data (X-ray) and the Parallel Approach to the Uillinn Method for Knee measurements.

6.4.6. The Uillinn Method (Mid-Point Approach) at the Knee.

There was also a significant difference (p=0.05) between the Mid-Point approach and the X-ray data. The SB bias was calculated as –2.75°, which meant that in general the Mid-Point approach would measure that amount less than the PACS system. The LOA spread was –9.96°, 4.4° and the TE between the 2 datasets

was 2.6° which was the second lowest random error score after the Parallel approach (see Table 6.1; Graph 6.4).



Graphs 6.4. Limits of Agreement between PACS data (X-ray) and the Mid-Point Approach to the Uillinn Method for Knee measurements.

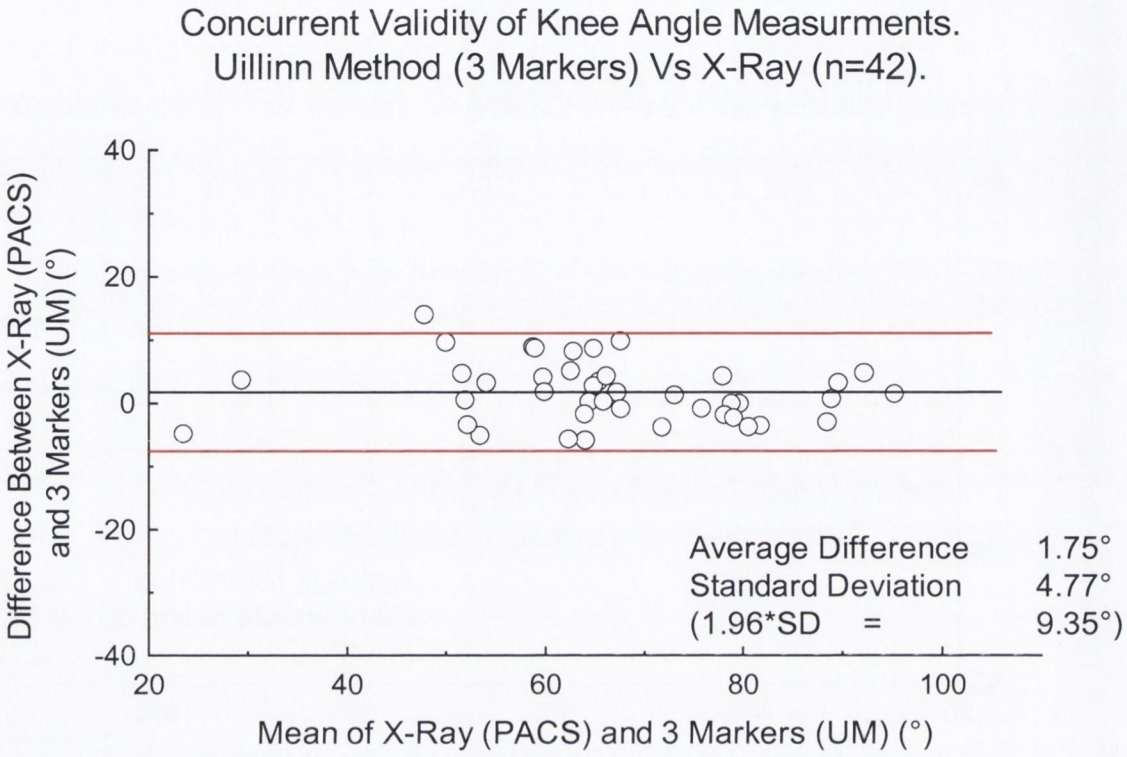
6.4.7. The Uillinn Method (3-Markers Approach) at the Knee.

There was a significant difference (p=0.05) between the 3-Marker approach to the UM and the radiographic measurement data. The SB for this difference was 1.75° which illustrated that the 3-Marker approach would generally yield slightly larger measurements that those calculated using X-rays. The agreement spread between the datasets was widest at –7.6°, 11.1° and the TE was the highest observed at 3.3° (see Table 6.3; Graph 6.5).

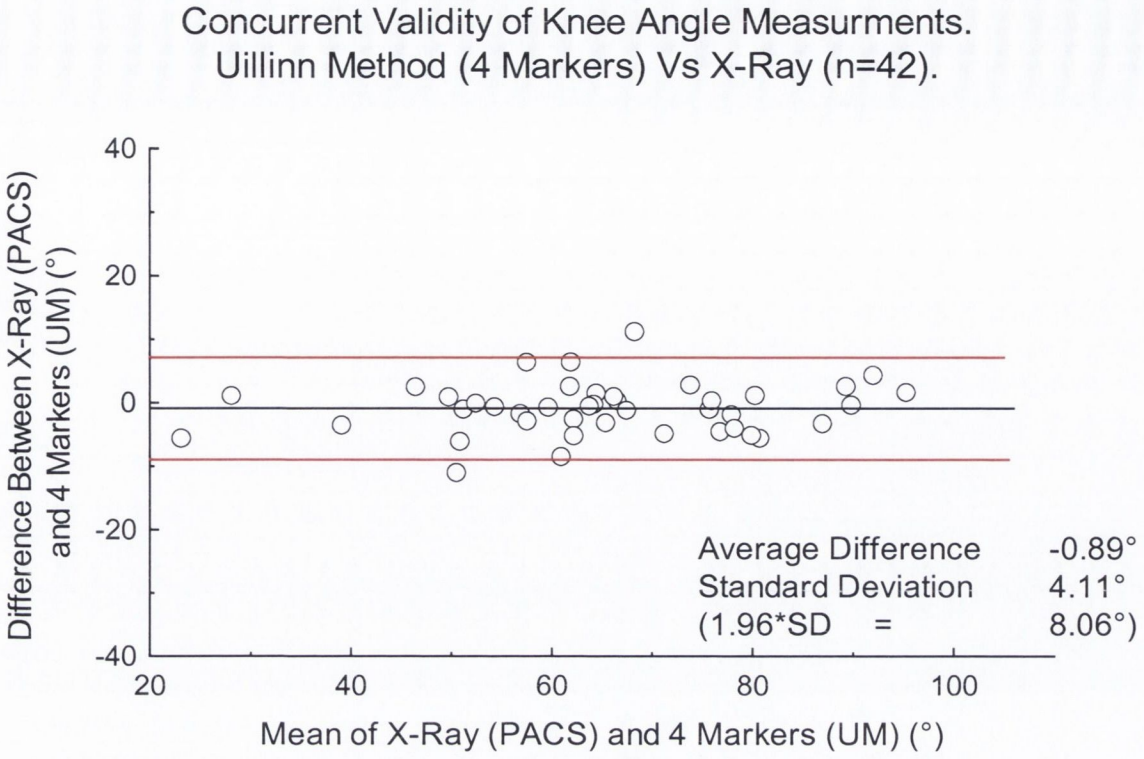
6.4.8. The Uillinn Method (4-Markers Approach) at the Knee.

Unlike the 3-Marker approach to the UM, the 4-Marker approach did not have a significant difference to the PACS data and there was no SB between the datasets, with an average value of –0.9°. The 95% LOA spread was –8.9°, 7.17°,

with a TE between it and the X-ray data of 2.9°. This was the second highest TE seen out of all the methods assessed (see Table 6.1; Graph 6.6).



Graphs 6.5. Limits of Agreement between PACS data (X-ray) and the 3-Marker Approach to the Uillinn Method for Knee measurements.



Graphs 6.6. Limits of Agreement between PACS data (X-ray) and the 4-Marker Approach to the Uillinn Method for Knee measurements.

6.4.9. Validity Results for Measurement of the Elbow

Elbows (Difference from PACS)	UG	No Markers	Parallel	Mid-Point	3-Markers	4-Markers
Maximum Range (from, to..)	13.56° -28.88°	11.31° -20.22°	15° -10.72°	23.5° -4.75°	11.87° -21.4°	13° -20.72°
Average Difference	**2.2°**	**-1.36°**	**4.27°**	**11.94°**	**-3.3°**	**-2.13°**
Systematic Bias (SB)	*Yes*	*No*	*Yes*	*Yes*	*Yes*	*No**
Std. Dev.	8.33°	7.31°	6.15°	6.55°	7.74°	7.14°
Significant Difference Of Means (pair t-Test; p=0.05)	*No*	*No*	*Yes*	*Yes*	*Yes*	*Yes*
Pearson's Product Moment Correlation Coefficient (PPMCC)	0.70	0.73	0.81	0.78	0.70	0.74
Limits of Agreement (LOA) (SD *1.96)	16.33°	14.34°	12.07°	12.84°	15.19°	14°
Upper LOA	**18.54°**	**12.98°**	**16.35°**	**24.78°**	**11.87°**	**11.87°**
Lower LOA	**-14.13°**	**-15.7°**	**-7.79°**	**-0.9°**	**-18.48°**	**-16.14°**
Typical Error (TE)	**5.89°**	**5.17°**	**4.35°**	**4.63°**	**5.47°**	**5°**
SB limit: 2.2°; * Yes if based on a perfect sample (perfect sample SB limit: 1.75°)						

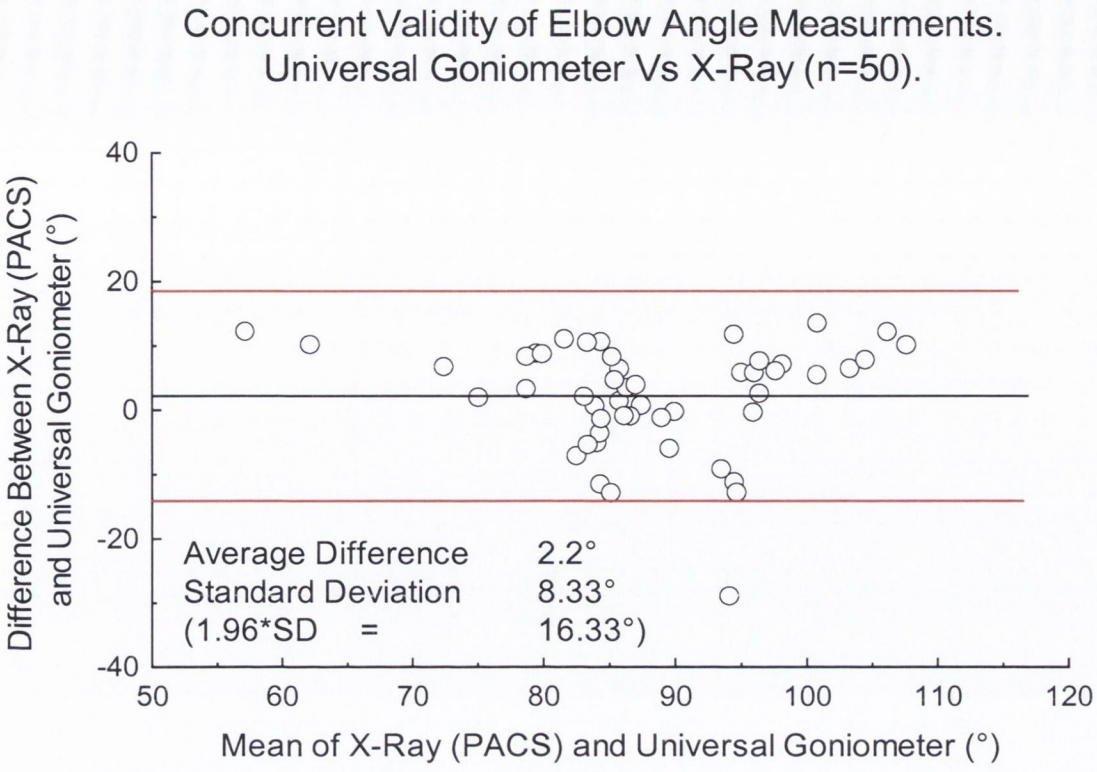
Table 6.4. Data Analysis for each of the Experimental Measures, Comparisons with X-ray (PACS) Measurements of the Elbow.

6.4.10. The Universal Goniometer (UG) at the Elbow.

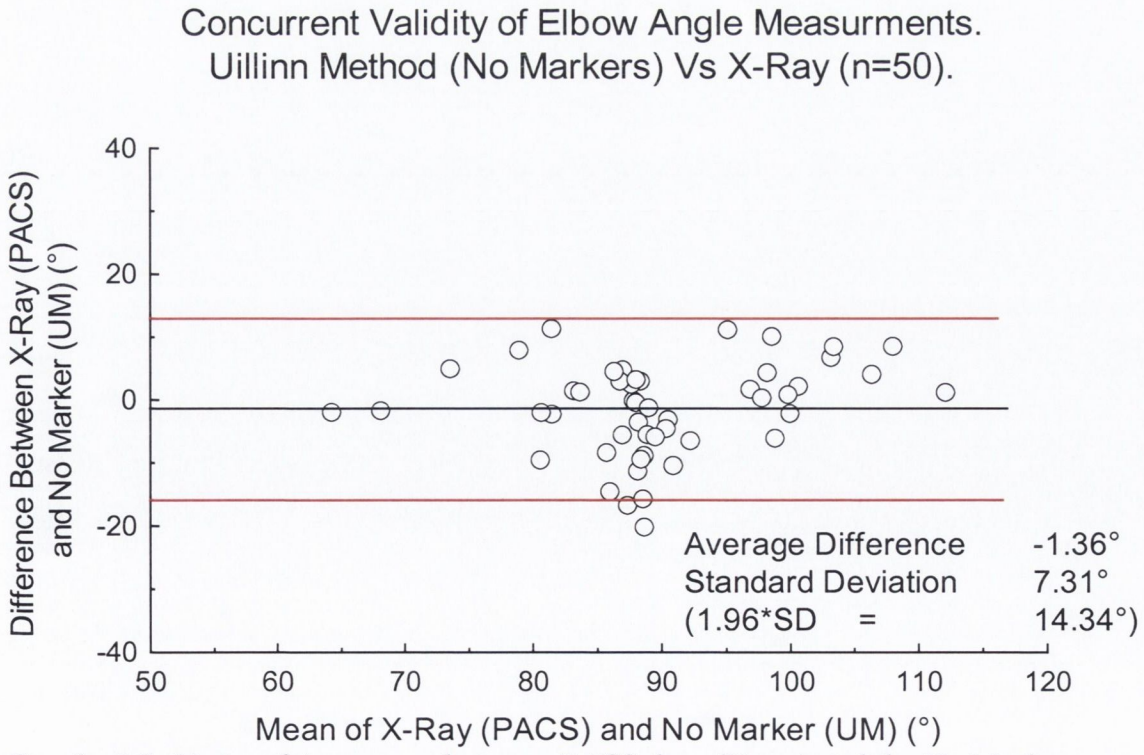
There was a systematic bias (SB) between the control data (PACS) and the experimental data (UG). This shows that the UG will generally measure 2.2° greater than the PACS system. Despite this SB, there was no significant difference (p=0.05) between the two datasets. The TE between the two systems was 5.9° and the 95% LOA showed a spread between –14.1°, 18.5° (see Table 6.4; Graph 6.7).

6.4.11. The Uillinn Method (No Markers Approach) at the Elbow.

There was no SB or significant difference (p=0.05) between the No Marker approach to the UM and the PACS data. The TE for this method was 5.17° and the 95% LOA spread was –15.7°, 13° (see Table 6.4; Graph 6.8).



Graphs 6.7. Limits of Agreement between PACS data (X-ray) and the Universal Goniometer for Elbow measurements.

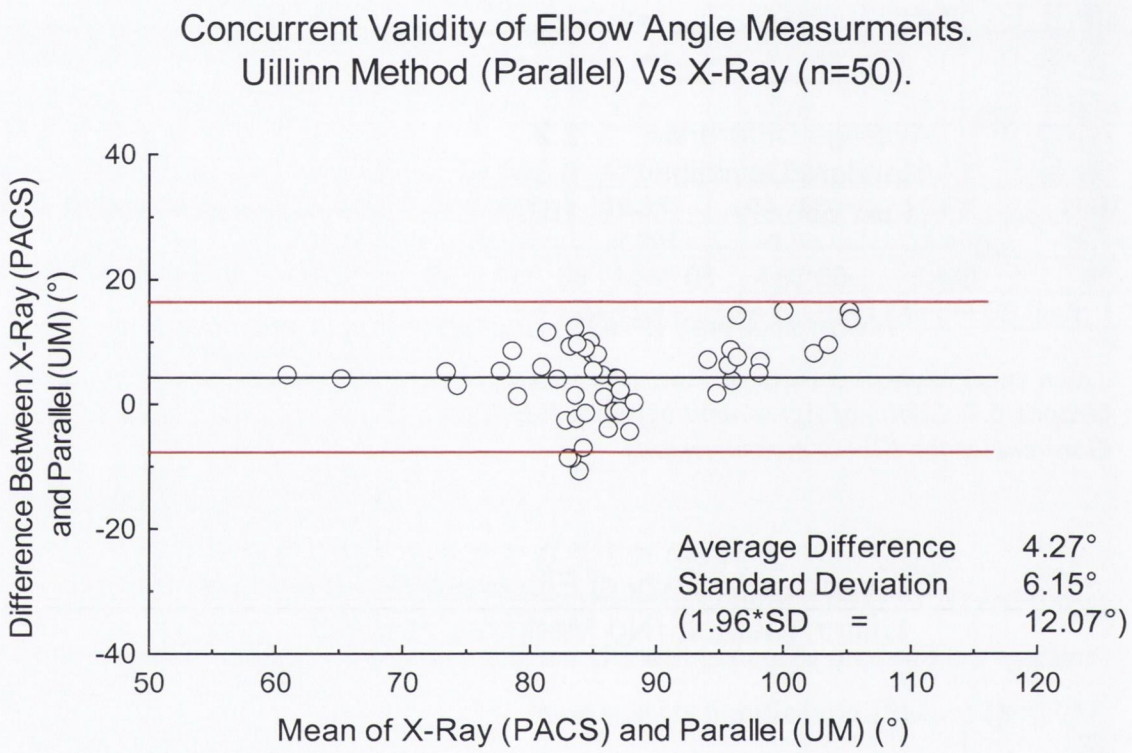


Graphs 6.8. Limits of Agreement between PACS data (X-ray) and the No Marker Approach to the Uillinn Method for Elbow measurements.

6.4.12. The Uillinn Method (Parallel Approach) at the Elbow.

The Parallel approach generally measured 4.3° greater than the PACS system (see Table 6.4). This was the second largest SB among the different methods. There was also a significant difference (p=0.05) between the datasets.

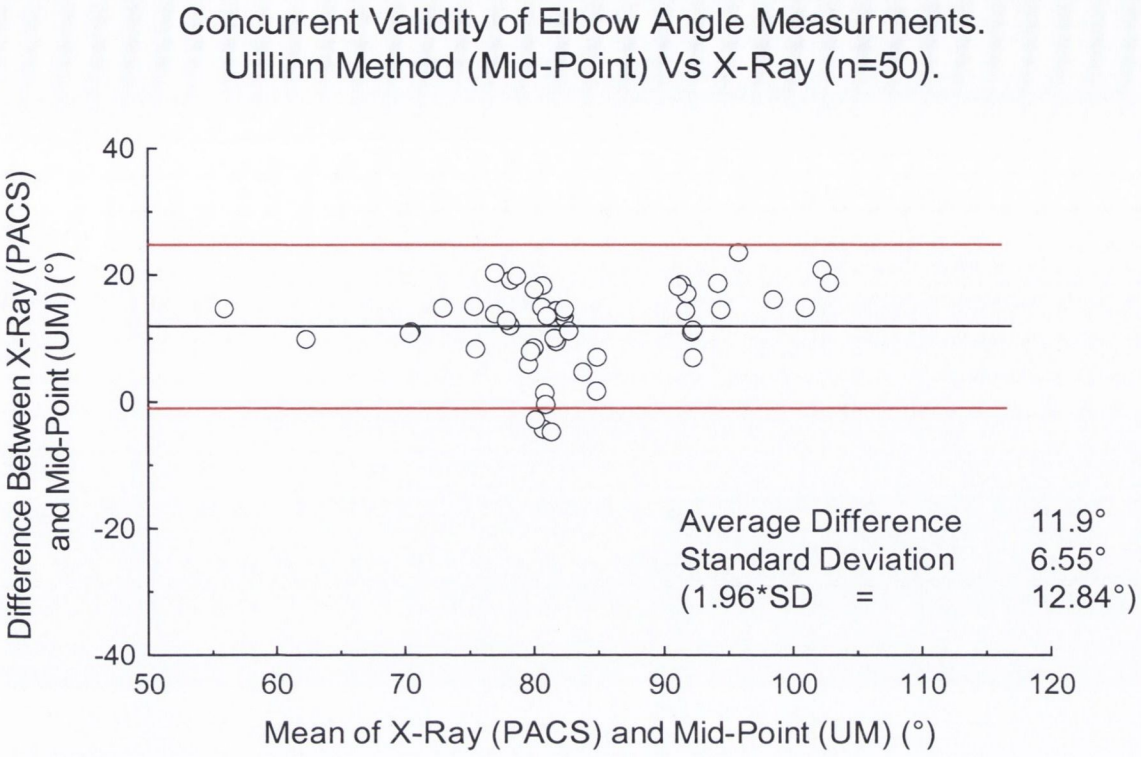
Assessments of random error showed the lowest TE (4.3°) and narrowest 95% LOA spread (-7.7°, 16.3°). In addition, it had the highest correlation coefficient 0.81 between the datasets (see Table 6.4, Graph 6.9).



Graphs 6.9. Limits of Agreement between PACS data (X-ray) and the Parallel Approach to the Uillinn Method for Elbow measurements.

6.4.13. The Uillinn Method (Mid-Point Approach) at the Elbow.

The Mid-Point approach had a very large SB of 12° greater than the PACS measurements. This was almost three times as much as the next largest SB (Parallel; 4.3°). As expected with this magnitude of difference of the means reflected a significant difference (p=0.05) between the datasets. Nonetheless, random error as assessed by the 95% LOA and the TE were the second lowest observed for these measurements (TE = 4.6°; 95% LOA, -1°, 24.7°) (see Table 6.4; Graph 6.10).



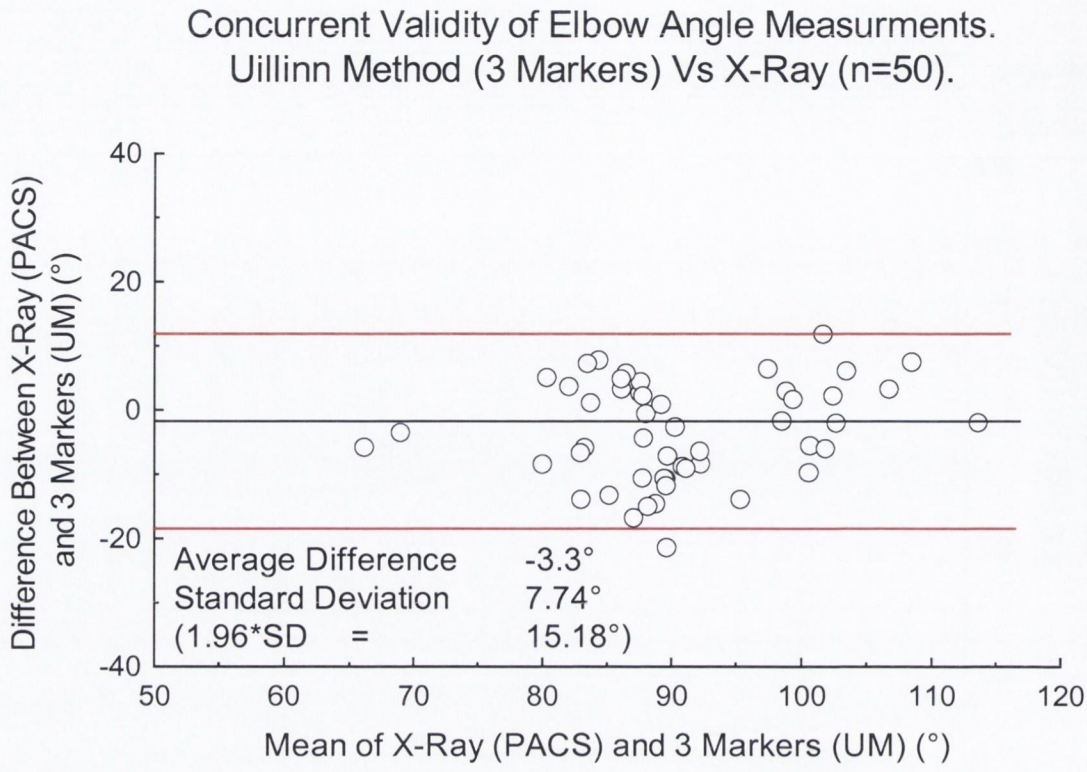
Graphs 6.10. Limits of Agreement between PACS data (X-ray) and the Mid-Point Approach to the Uillinn Method for Elbow measurements.

6.4.14. The Uillinn Method (3-Markers Approach) at the Elbow.

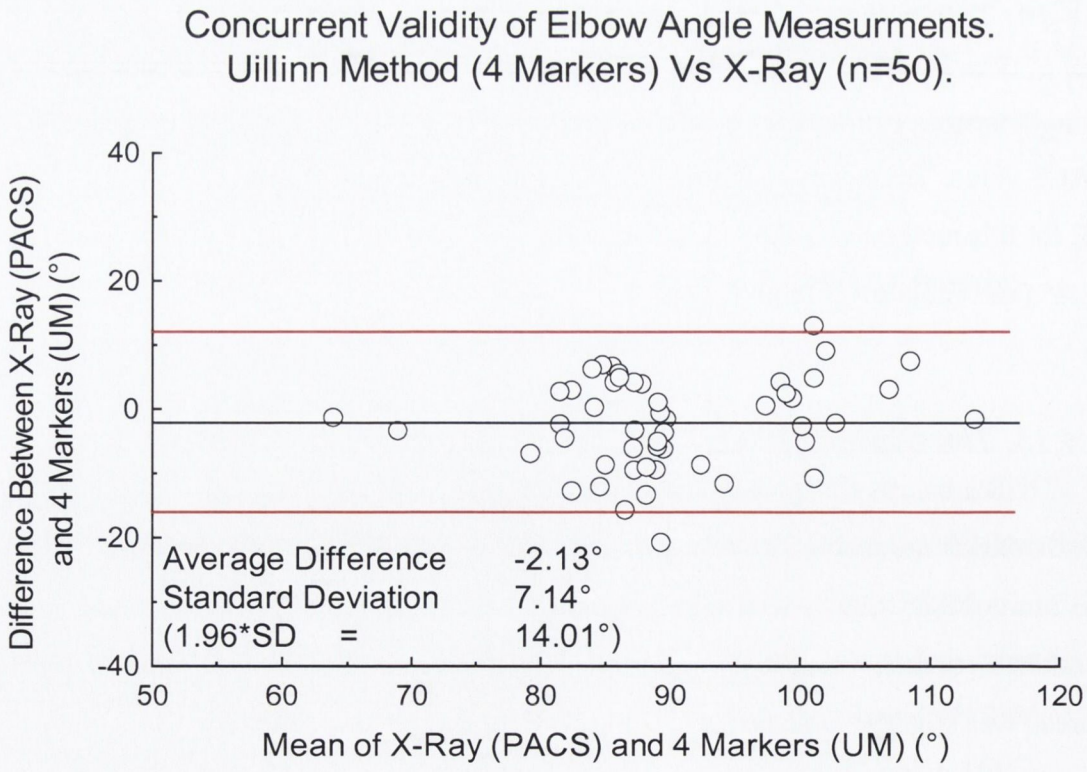
The 3-Marker method had a negative SB value of –3.3° which meant that using 3 anatomical markers would generally produce smaller measures than the PACS value. There was a significant difference between the datasets (p=0.05). The TE for this method was the second highest at 5.5° and the LOA spread was –18.5°, 11.8° (see Table 6.4; Graph 6.11).

6.4.15. The Uillinn Method (4-Markers Approach) at the Elbow.

The results for the 4-Marker approach to the UM, were slightly better than the 3-Marker approach. The average difference of –2.13° was on the border of the SB limit of 2.2°, there was a significant difference (p=0.05) between the means. The random error data was also similar to the 3-Marker method. TE was 5° and the LOA spread was slightly narrower at –16.1°, 11.8° (see Table 6.4; Graph 6.12).



Graphs 6.11. Limits of Agreement between PACS data (X-ray) and the 3-Marker Approach to the Uillinn Method for Elbow measurements.



Graphs 6.12. Limits of Agreement between PACS data (X-ray) and the 4-Marker Approach to the Uillinn Method for Elbow measurements.

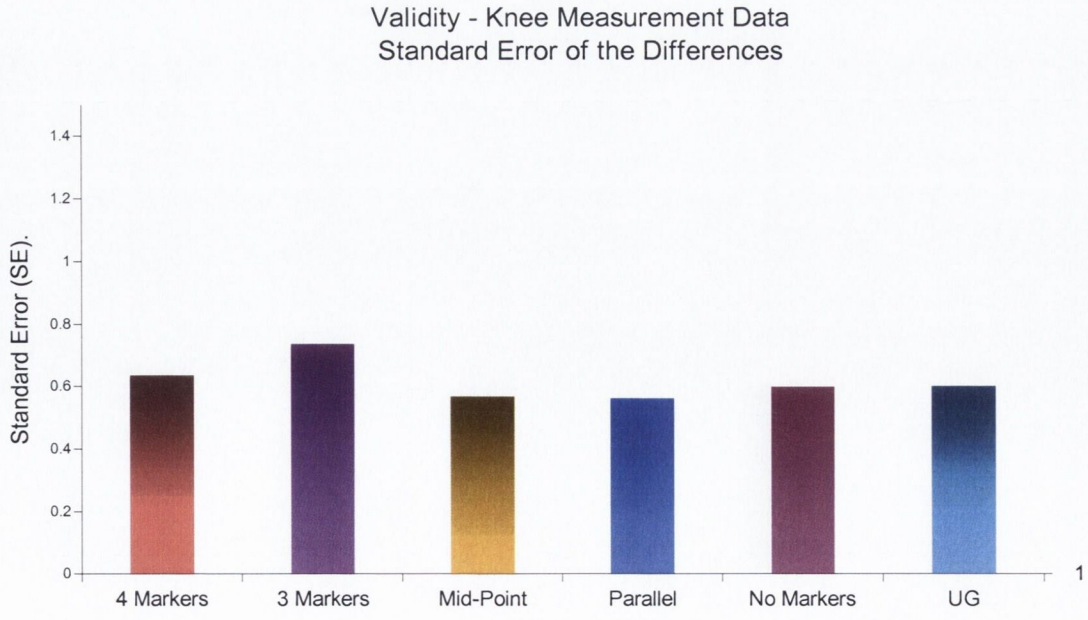
6.4.16. Confidence of Results Data.

For the knee measurements the standard error (SE) of the differences between the data sets ranged from 0.56° to 0.73° (see Table 6.5), this meant that additional differences as high as these figures could be seen if sampling was continued. However, all SEs were small (less than 1°) implying that individual differences of this magnitude would probably not have a substantial impact on the relatively large datasets.

The confidence limits of the TE ranged by about 1°, which meant that the TE figures upon which the results are based could in fact vary by approximately ±0.5° (see Table 6.3, 6.4; Graph 6.13).

Knees (Difference from PACS)	UG	No Markers	Parallel	Mid-Point	3-Markers	4-Markers
Standard Error (SE) of Differences	0.6°	0.59°	0.56°	0.56°	0.73°	0.63°
Lower Confidence Limits (TE)	2.33°	2.32°	2.18°	2.2°	2.86°	2.47°
Upper Confidence Limits (TE)	3.37°	3.36°	3.15°	3.18°	4.13°	3.56°

Table 6.5. Standard Error and Confidence Limits for the Knee Data Calculations.

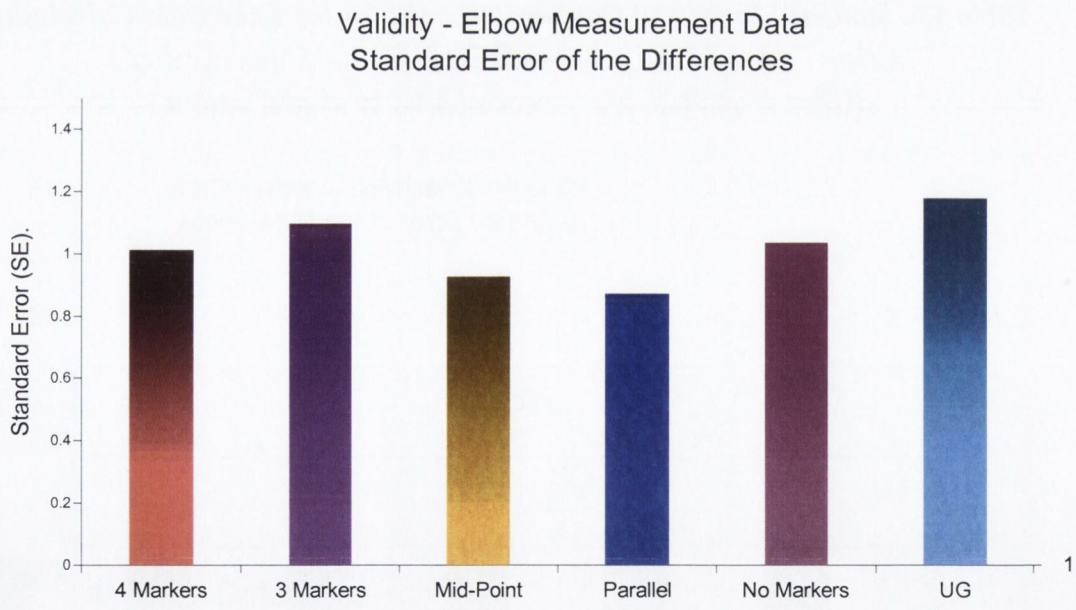


Graphs 6.13. Standard Error in the Data; Differences between PACS and the other measurement techniques.

Larger SEs were seen in the elbow measurement data, differences between the PACS and other measurement systems ranged from 0.87° to 1.17°. This meant that there could be larger samples, of an addition 1.17° in the UG data upon further examinations. However, as with the knee data additional data of this magnitude in the relatively large sample would probably not have a noticeable effect on the overall data. The true value of the TE data may have varied by approximately ± 1° (see Table 6.4, 6.6; Graph 6.14)

Elbows (Difference from X-ray)	UG	No Markers	Parallel	Mid-Point	3-Markers	4-Markers
Standard Error (SE) of Differences	1.17°	1.03°	0.87°	0.92°	1.09°	1°
Lower Confidence Limits (TE)	5°	4.4°	3.74°	3.98°	4.7°	4.34°
Upper Confidence Limits (TE)	7.08°	6.2°	5.24°	5.56°	6.58°	6.07°

Table 6.6. Standard Error and Confidence Limits for the Elbow Data Calculations.



Graphs 6.14. Standard Error in the Data; Differences between PACS and the other measurement techniques.

6.5. Discussion

The validity of sagittal elbow and knee joint angles as measured by the UM and the UG were addressed in this study. Five different approaches to the UM were investigated (2.3), two of which involved the use of adhesive skin markers. The objective of the study was to assess the agreement of each of the 6 measurement methods with a criterion (control) measure. The criterion measure chosen had to be a method that had strong validity for the provision of accurate joint angle data. Most authors believe that angles calculated from radiographs meet this demand and should be considered a gold-standard against which other external measurement techniques should be assessed (Enwemeka 1986; Gogia et al, 1987; Brosseau et al, 1997, 2001; Rowe et al, 2001; Berryman Reese and Bandy, 2002; Sprigle et al, 2003; Norkin and White, 2003).

6.5.1. Examination Procedures

A concurrent validity trial must try to gather all the comparative measurements at approximately the same time (LoBiondo-Wood and Haber, 1998). The methodology used in this study attempted to achieve this by carrying out UG measurements, photographic image capture with and without adhesive skin markers and radiographic pictures in close succession.

Because a patients population under-going routine investigations were used, standard X-ray procedures had to be followed. This made the recommended testing procedures straightforward for the elbow trial, but difficult at the knee. The procedures at the knee required the subject to be positioned in side lying (lateral side down) during the X-ray procedure thus making external lateral measurements impossible.

After a number of solutions to this problem were considered, the most practical approach appeared to be the use of a radiolucent lower limb splint (see Figure 6.1) to immobilise the knee and to facilitate a changed in subject position. The splints used were fabricated from thermoplastic material similar to that used by Fish and Windgate (1985). Those authors used a splint to stabilise the elbow in a reliability study for the UG and measurements taken from conventional photography. While the change in subject position for the knee measurements were

not ideal, the splints and other controlling procedures were thought have been successful at stabilising the knee position.

6.5.2. Value of the Research

One of the driving forces for this set of investigations was to address the information gap regarding the performance of the commonly used UG. There seem to be a number of specific gaps in knowledge regarding the validity of UG measurements. Some validity studies did not use a strong enough criterion measure (Petherick et al, 1988; Rheault et al, 1988; Paquet et al, 1991; Russell et al, 2003). While, no gold-standard criterion research was identified for UG measurement of the elbow.

Another important rationale for the study was provided by the fact that no criterion validity study appears to have been carried out on a photographic based measurement system. This had implications for the UM, but also for all other photographic based research in general.

A number were identified that dealt with the validity of knee joint measurements using the UG, these all had at least some methodological flaws. Not least of these were the types of data analysis used. Daly and Bourke, (2000) state that it has been common place in the medical literature in general to have used the wrong statistical tools to establish agreement. It is felt that this notion is especially true in the field of osteokinematic research.

Therefore, this study attempted to use analysis tools that are currently recommended and justified by the statistical medical literature. It was hoped that the data provided by these analysis tools would be readily interpretable by clinicians and thereby promote professional knowledge towards the joint angle assessment tools involved in this study. Holm et al (2000) believe that it is important to use at least 2 or 3 different analytical calculations to give true, clinically relevant information about measurement discrepancies. Therefore, further more traditional analysis tools were used which would allow some level of comparison with previous studies.

6.5.3. Non-Random Error (Systematic Bias)

The limit of SB of 2.5% of the average measurement value was used in this study (Batterham and George, 2000). Because the average angle measured was different for both the knee and the elbow, there were different limits of SB at these joints (knee 1.6°; elbow 2.2°).

However, Berryman Reese and Bandy (2002) hold that the elbow and knee have approximately the same available range of motion (ROM) of about 140°. This may imply that when assessing the measurement error associated with these joint, the limits of SB should be equitable. If a completely homogenous test model were used, i.e. if all joint angles were assessed equally, the average value for the sample would be 70°. Therefore, in a perfect sample the SB limit for both the knee and elbow should be set at 1.75°.

The use of such a limit would ensure that the SB limit would not change from one sample to the next, and could be thought of as a constant. A perfect sample SB limit as opposed to one drawn from the average sampled value was not used here, but if it were it would have only changed one conclusion. The data for the 4-Marker approach to the elbow would change from no SB to one of a SB, generally measuring 2.1° less than the radiographic data (see Table 6.3, 6.4, page 276, 282).

The current method of SB calculation found that at the knee only 3 methods had a SB, the Mid-Point and Parallel approach to the UM would generally measure 2.7° and 3.6° less than the PACS system, i.e. a smaller value, while the 3-Marker approach would usually provide a value of 1.75° greater than the PACS system.

At the elbow all but the No Marker had a high level of SB. By far the largest value was seen for the Mid-Point approach which was nearly 12°. This meant that virtually all of the Mid-Point data would be much higher than the true measurement value. The Parallel approach had the second highest SB of 4.3°.

It may be possible to account for these SBs in future versions of the software by facilitating a calculation off-set. However, currently this feature is not an option and more work may need to be done to comprehensively understand all of the factors associated with the SB.

6.5.4. Random Error

The measurement techniques that had the lowest TE and narrowest LOA scatter plots in relation to the PACS system at both the knee and the elbow were the Parallel and Mid-Point approaches to the UM (see Table 6.3, 6.4). This implied that the relationship between these approaches and the PACS calculations were the most consistent and predictable. This finding presented a paradox as those two methods had the highest non-random error with the PACS data.

Only the No Marker and the 4-Marker approaches were absent from SB at both joints (see Table 6.3, 6.4). TE for both of these methods were comparable, however there was a significant difference ($p=0.05$) between the PACS and the 4-Marker approach. This implied that the No Marker approach had better agreement with the X-ray data.

The poorest result at the knee was seen for the 3-Marker method, which had a SB and the highest TE and LOA spread. These results are surprising as using a 3-Marker model over the standard bony landmarks (as proposed by Norkin and White (1995)) is the photographic methodology recommended by most authors (Fish and Windgate, 1985; Karkouti and Marks, 1997; Szulc et al, 2000). Based on the evidence of this current study this model is the least valid for photographic knee measurements.

At the elbow the least favourable approach was the UG. This method also had a SB and the poorest agreement with X-ray. This finding is in agreement with Fish and Windgate (1985) who reported that the UG performed worse than conventional photography for elbow measurements.

6.5.5. The Criterion Data

The Picture Archiving and Communications System (PACS), which is a digital X-ray system, was used to display all of the X-rays used in this study and calculate the joint angles via the systems virtual goniometer feature. This works in much the same way that the UM allows its user to manipulate virtual measurement arms (VMA) over the on-screen image.

In the past other authors have reported reliability error in joint angle calculations taken from radiographs (Jackson and Waugh ,1984; Gray et al, 1991; Ishii et al, 1995; Lonner et al, 1996; Lantz et al, 1999; Greene et al, 2001) (1.2.11).

For this reason the criterion data was calculated twice and the average value taken as the true value.

Examination of the repeated PACS data revealed that the TE between the repeated assessments by the same examiner was less than 1°, and the RC which presents the maximum range of differences for 95% of measurements was approximately 2.5° (see Table 6.1, 6.2, page 275; Appendix K).

In addition to this intra-rater trial a small sample of elbows and knees were assessed between 2 examiners. The error of greatest concern was seen for the inter-rater reliability at the knee. There was a SB between the examiners of 2.2°, creating a 95% LOA spread of -2°, 6°. There was a lower TE of 1.5°. Inter-rater agreement for elbow measurements were substantially stronger, there was no SB between the examiners, the RC just above 2.5° and the TE was below 1°. These latter results were more reflective of those seen in the intra-rater trial.

There were relatively large numbers used in the PACS intra-rater reliability trial (knees n=22; elbows n=50) but only a small sample in the inter-rater reliability trial (n=7), therefore the inter-rater reliability results should be viewed with some caution.

Only one article was found that investigated the reliability of angle measurements using the PACS system. Sanfridsson et al (1997) state that the reproducibility for measurements of slight knee flexion of 10 healthy volunteers (left and right) was ± 2.64° at a 95% confidence level. The current study found that the intra-rater RC for the PACS measurement of the knee flexion angle (RC= 2.76°) was identical to that reported by Sanfridsson et al (1997).

However, while Sanfridsson et al (1997) conclude that the PACS measurements fulfil the reproducibility requirement for clinical measurement, this study aimed to control procedural error as tightly as possible, and an error of this magnitude in the criterion data was seen as somewhat of a compromise to the precision of the results.

It was felt that one issue that may have helped to improve the accuracy of the PACS data may have been the use of larger X-ray plates. These would have provided an increased view of the bone shafts and thus may have assisted the examiner during the placement of the VMAs. However, the use of over-sized joint views for standard elbow and knee joint recordings is not normal radiographic

technique and therefore was not available during the experiments. Sanfridsson et al (1997) used full-length X-rays views for their angular calculations and did not manage to improve on the reliability results obtained in this study.

A problem for the elbow measurements was presented by the fact that the X-rays for this joint were actually medial view images, while all of the external measurements were taken from the lateral view. This presented a mismatch that could have added an additional amount of error to the results. It is likely that if error was added to the data it would be in the form of non-random error, i.e. systematic bias (SB). Indeed, there were larger amounts of SB in the elbow measurements than those seen in the knee data (that had a consistently lateral view of its joint) (see Table 6.3, 6.4).

It is not possible to definitively state why such a large amount of SB was present in the elbow data compared to the knee data. However, it was probably due to multiple factors, including the disparity between medial and lateral views, different joint size, the supination/ pronation available at the forearm and the concave nature of the humerus seen in some radiographs.

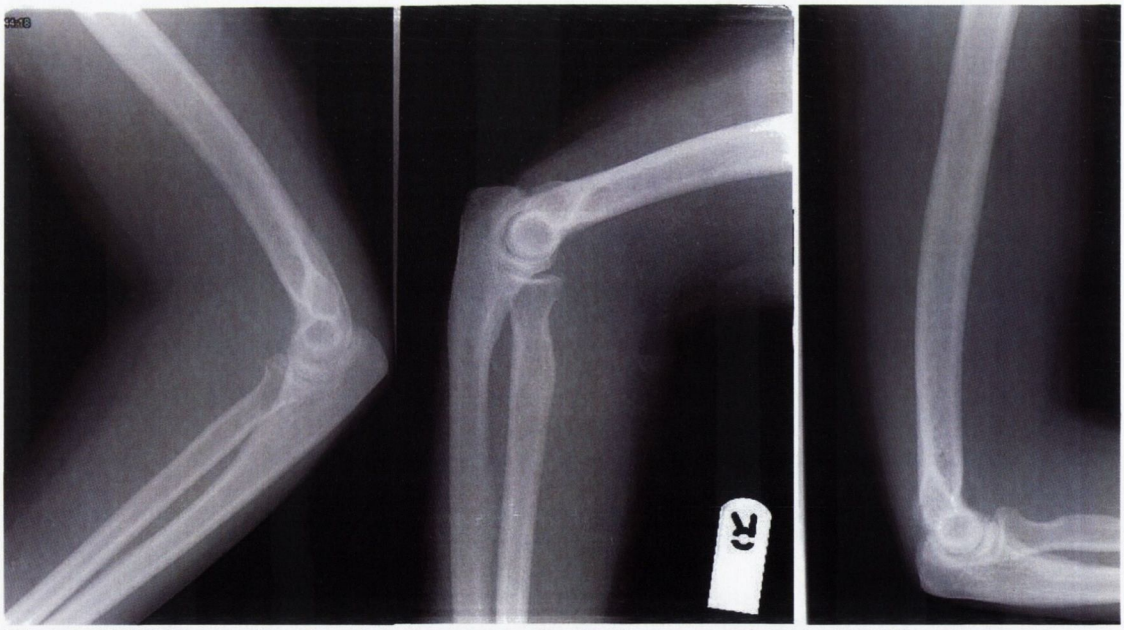


Figure 6.17 Example of concavities seen in the humerus during PACS assessments.

Forward concavity of the humerus is not a feature described in Gray's Anatomy (Soames, 1995) but one that was seen in this study (see Figure 6.17) and

may have been a source of non-random error. Further work may be needed to explain this phenomenon.

Precise in vivo joint measurements are difficult (Stokdijk et al, 1999). The PACS reliability trials from Sanfridsson et al (1997) and this current study show that there is some inevitable error in the criterion data. Hopkins (2000a) asserts that analysis of measurement validity is difficult due to the inevitable presence of error in the criterion measure, and Burton et al (1999) states that no system for joint angle measurement should be considered a certain gold standard.

The reliability of the PACS system was generally stronger at the elbow than the knee (see Table 6.1, 6.2). This reflects well with a view stated by Boone and Azen, (1979) that accuracy for measurements of the upper limb are better than those for the lower limb. However, this is in contrast with the results for the validity of the external measurement methods were substantially worse for the upper limb compared to the lower limb.

The disparity of measurement error at the knee and elbow is in keeping with the notion that the performance of measurement tools will vary at different joints (Fish and Windgate, 1985; Bovens et al, 1990). There was some consistency in the ranking of the external measurements at both joint. The Parallel and Mid-Point methods had the highest non-random error but the lowest random error at both joints. In short, these 2 methods had the best consistency with the PACS data, but had the poorest true value agreement with it.

6.5.6. Relationship of VMAs

The typical relationships between the orientation of the PACS VMAs (yellow) and the VMAs for each of the approaches to the UM (red and blue) can be seen in the Figures below (see Figure 6.5 to 6.14). These representations of the joint measurements were compiled using the 'overlay' feature in Adobe Photoshop 7.0. Figure 6.7, 6.8, 6.9 and 6.10 clearly show the tendency of the Parallel and Mid-Point methods to measure less than the PACS system.

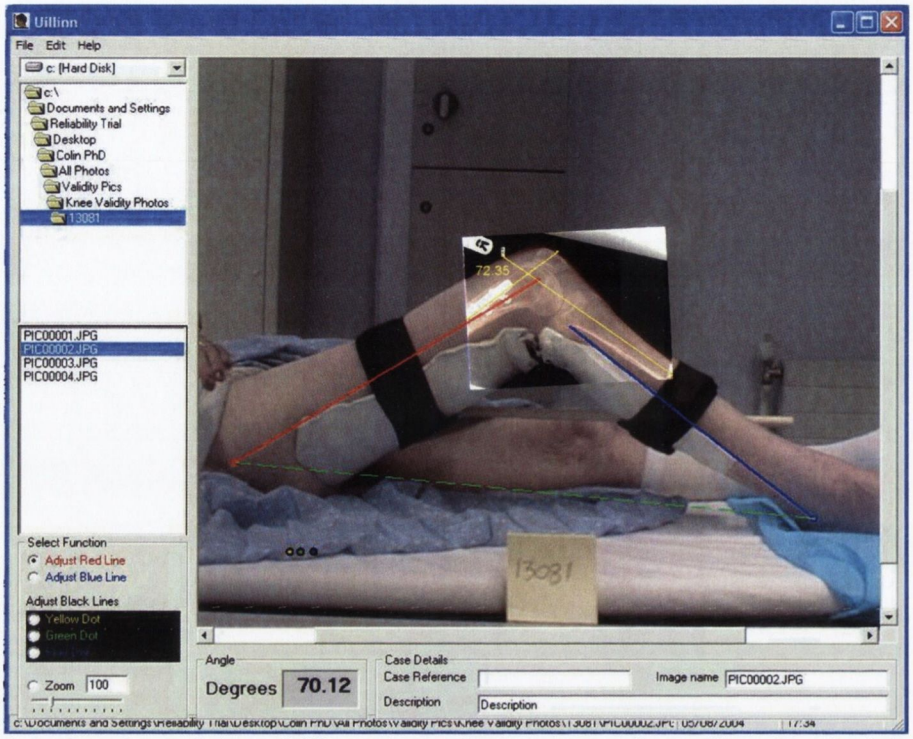


Figure 6.5. Overlay of the PACS angles calculation on the Uillinn Method (No Marker Approach) screen.

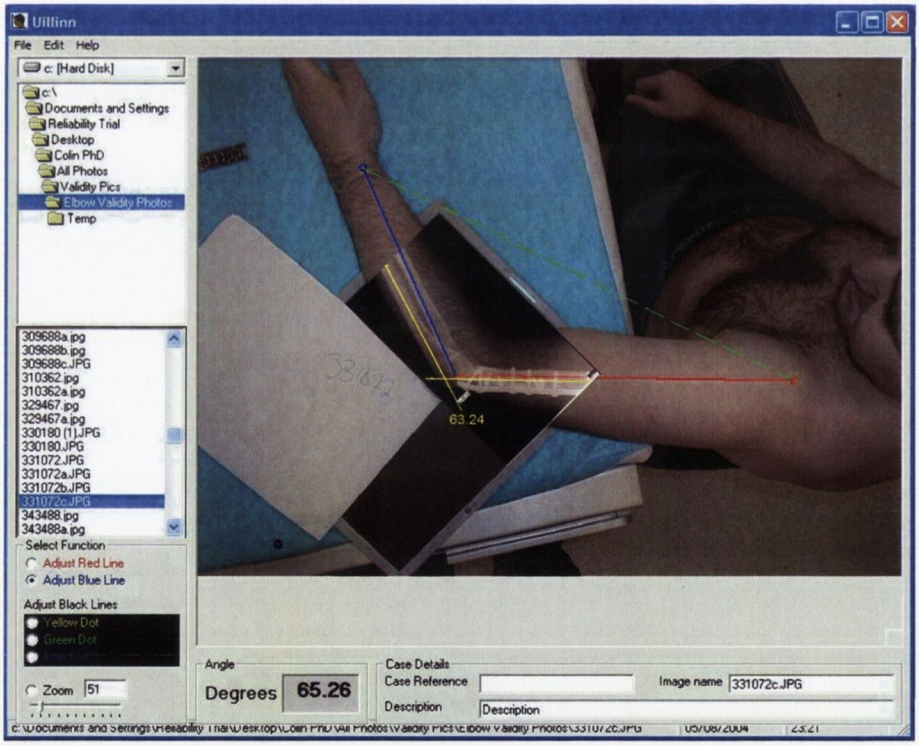


Figure 6.6. Overlay of the PACS angles calculation on the Uillinn Method (No Marker Approach) screen.

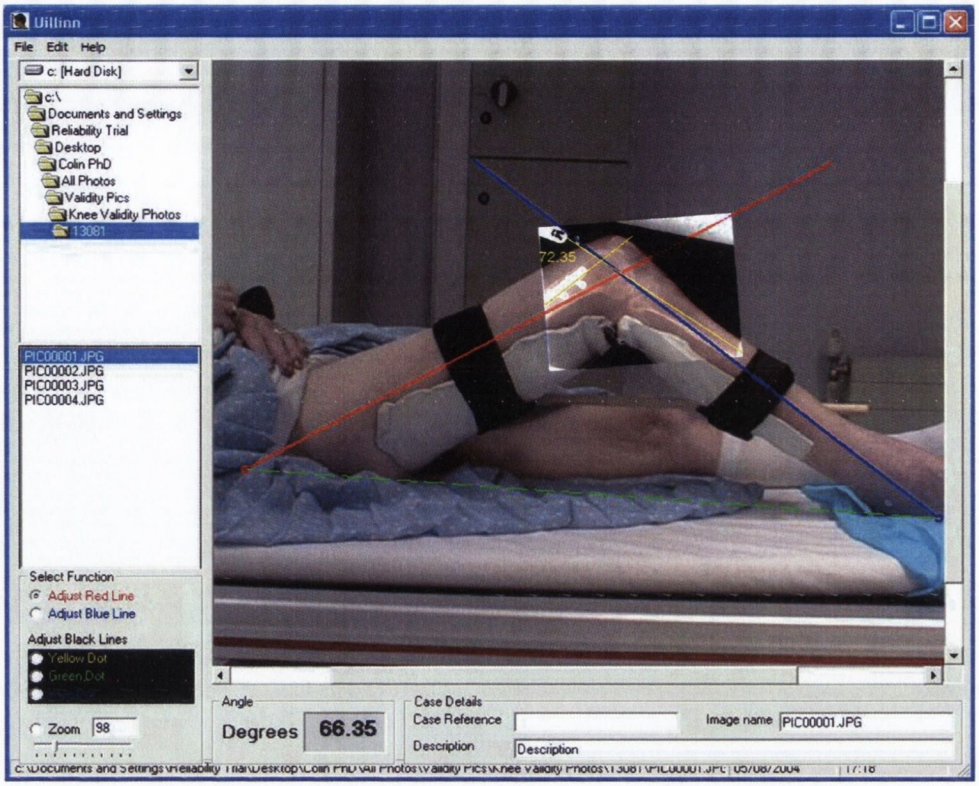


Figure 6.7. Overlay of the PACS angles calculation on the Uillinn Method (Parallel Approach) screen.

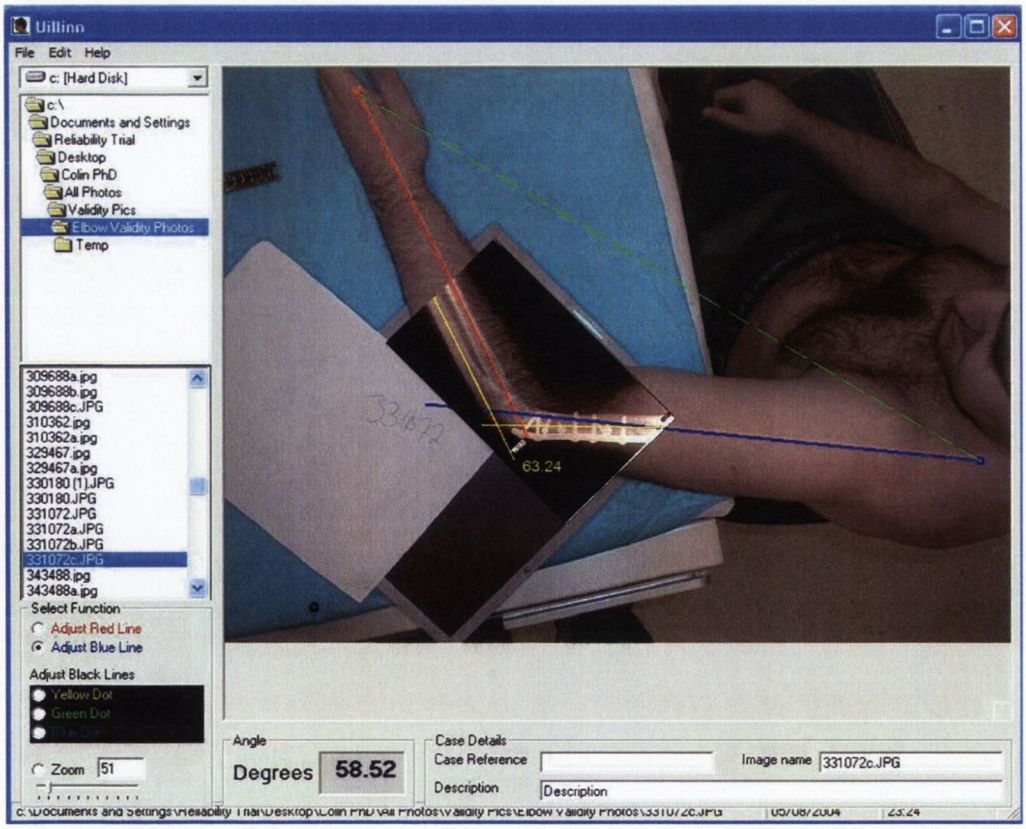


Figure 6.8. Overlay of the PACS angles calculation on the Uillinn Method (Parallel Approach) screen.



Figure 6.9. Overlay of the PACS angles calculation on the Uillinn Method (Mid-Point Approach) screen.

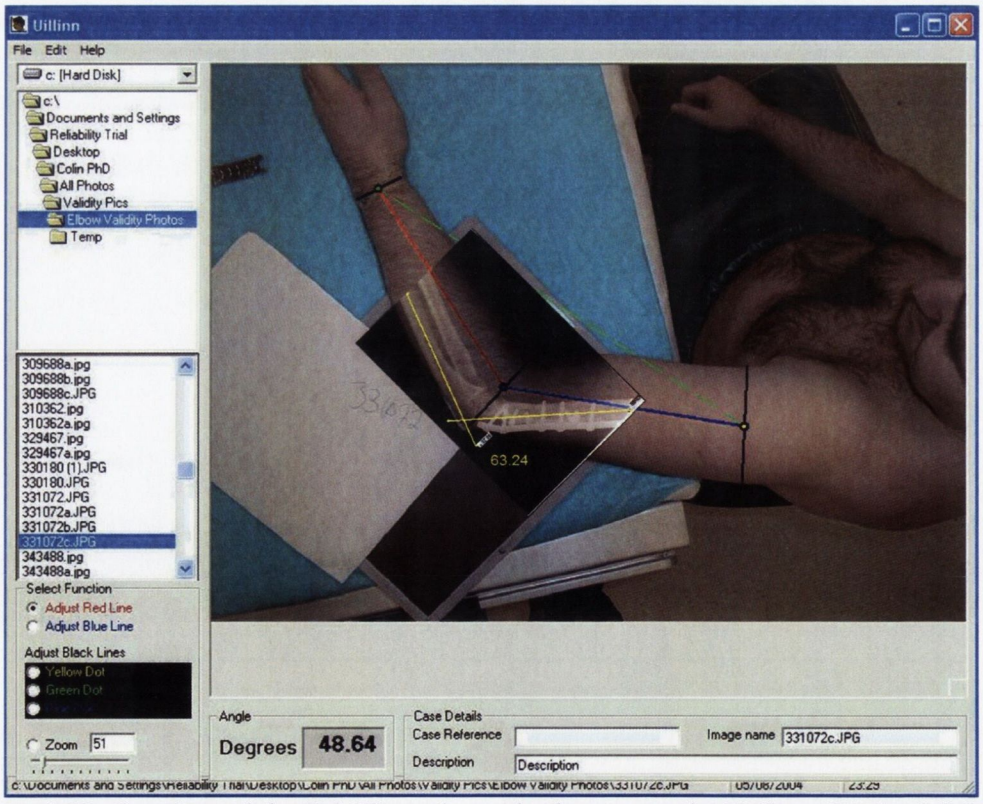


Figure 6.10. Overlay of the PACS angles calculation on the Uillinn Method (Mid-Point Approach) screen.

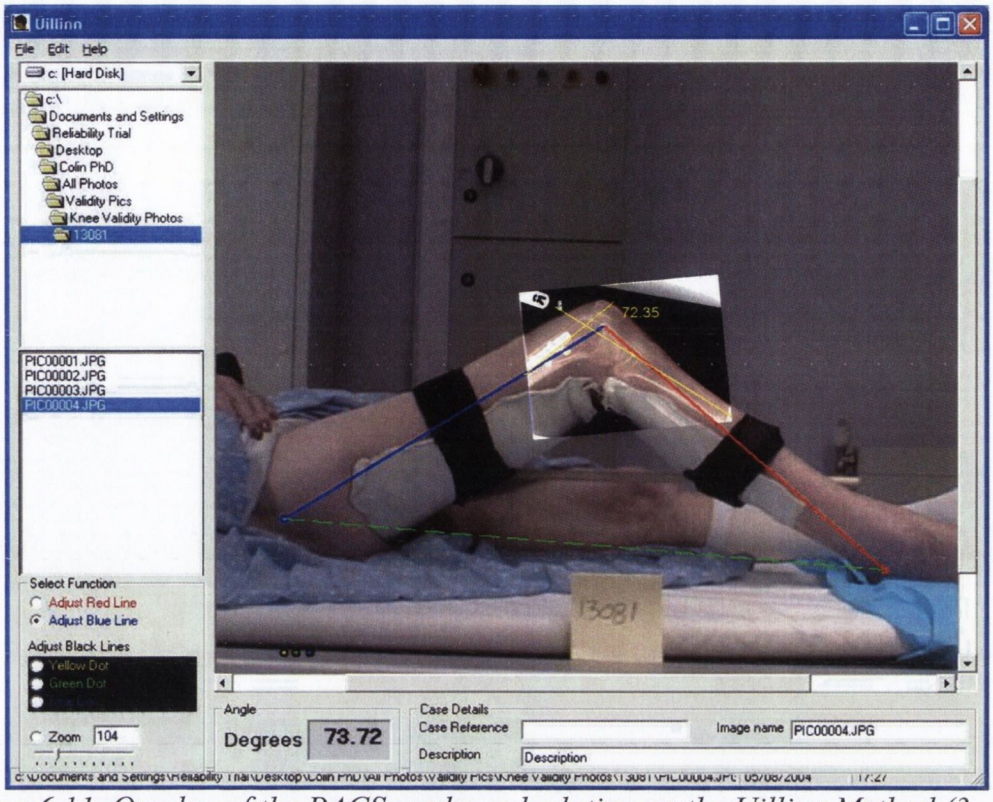


Figure 6.11. Overlay of the PACS angles calculation on the Uillinn Method (3-Marker Approach) screen.

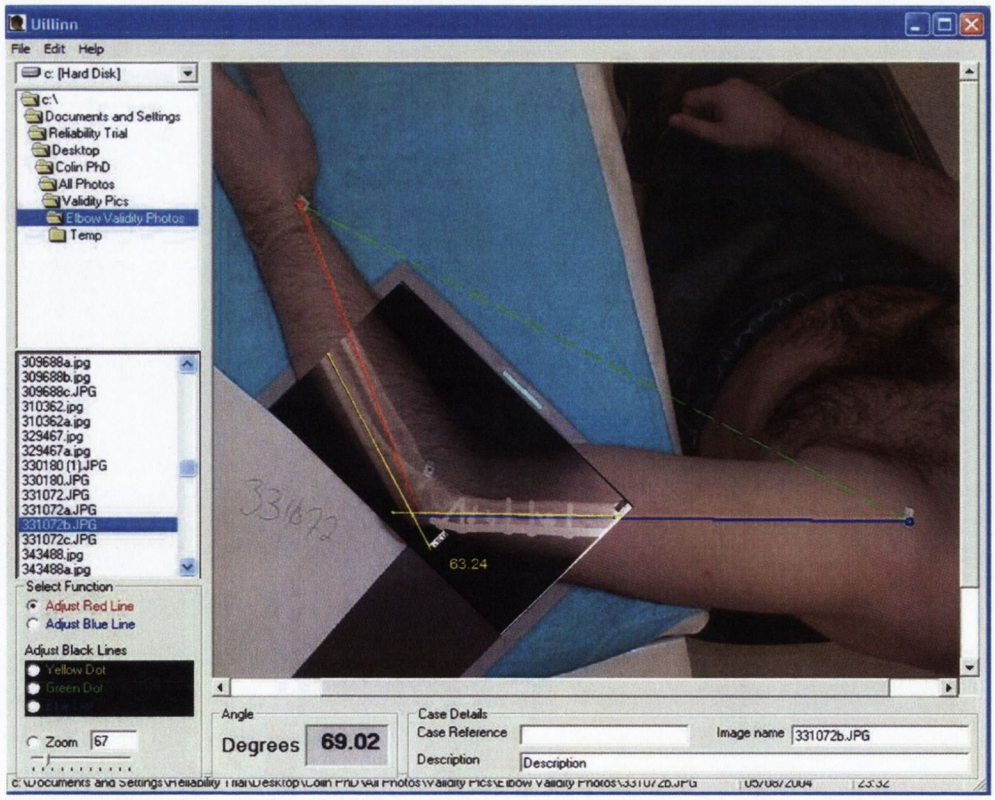


Figure 6.12. Overlay of the PACS angles calculation on the Uillinn Method (3-Marker Approach) screen.



Figure 6.13. Overlay of the PACS angles calculation on the Uillinn Method (4-Marker Approach) screen.

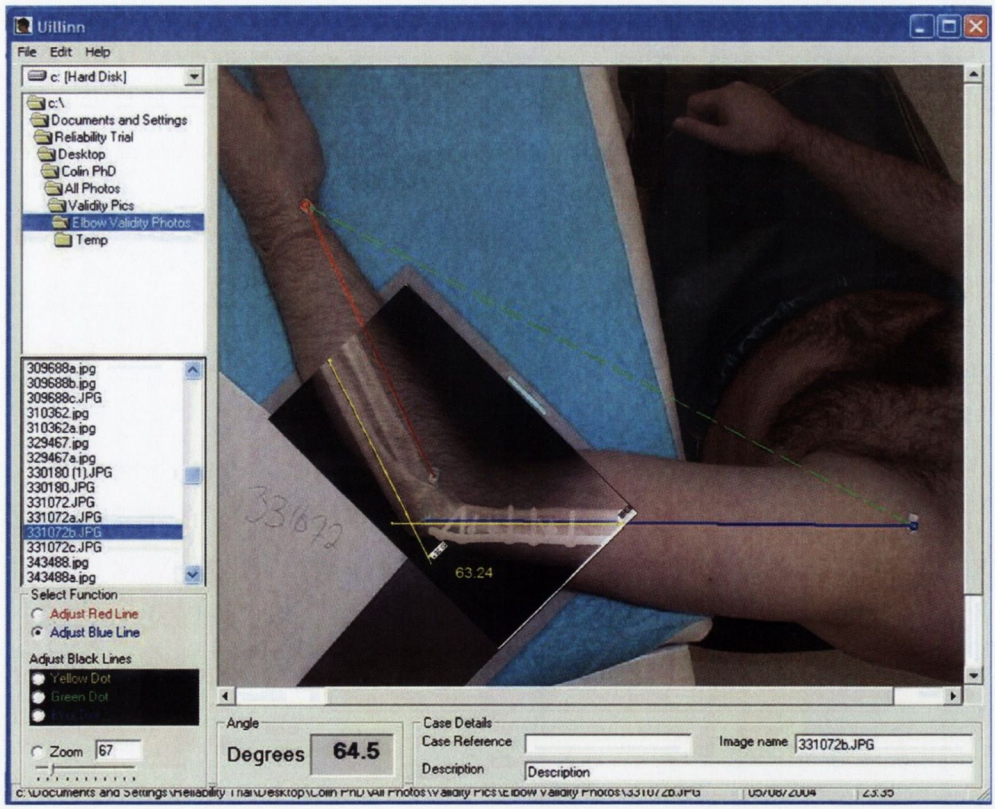
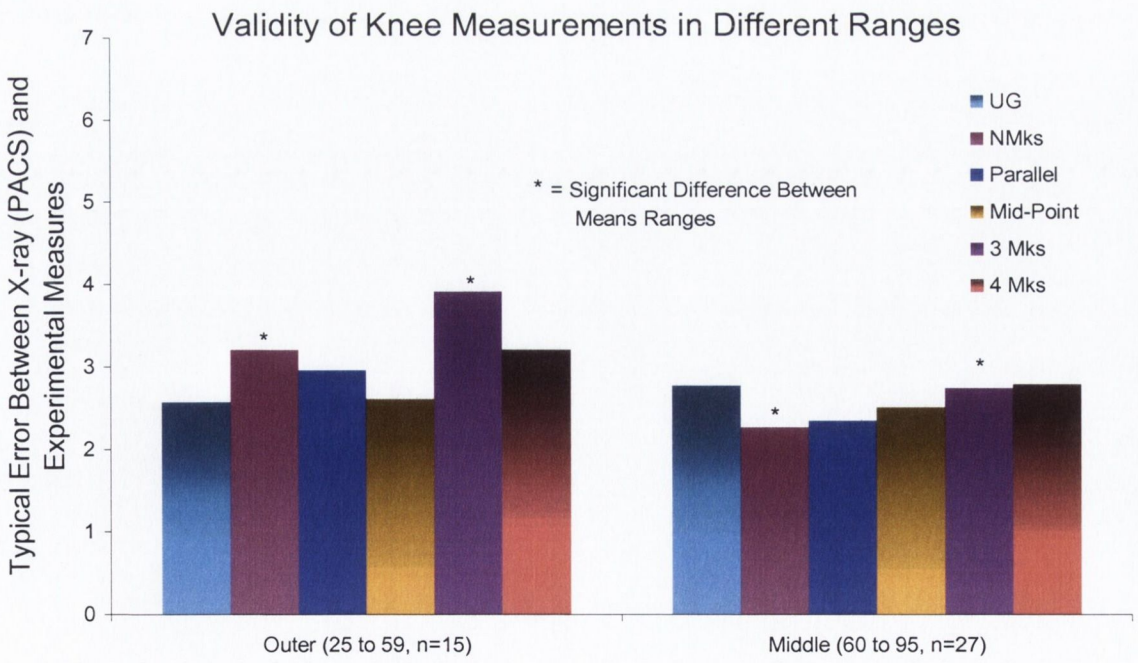


Figure 6.14. Overlay of the PACS angles calculation on the Uillinn Method (4-Marker Approach) screen.

6.5.7. Different Joint Ranges

Data for the entire 140° of elbow and knee flexion/ extension angles (Berryman Reese and Bandy, 2002) were not measured due the restrictions of routine radiographic procedures employed by the study. Therefore, comparisons of the performance of the all the 3 subdivisions of joint position, i.e., inner, mid and outer could not be assessed. Nonetheless, there was a sufficient spread of angles sampled to allow some analysis of change in performance in 2 ranges at each joint. At the knee the data could be divided into categories of approximately outer and mid-ranges. The outer range did not comprise angles less than 25° and was therefore do not comprehensively reflective outer-range. Angles 60° from to 95° were categorised as Mid-range (see Table 6.7).

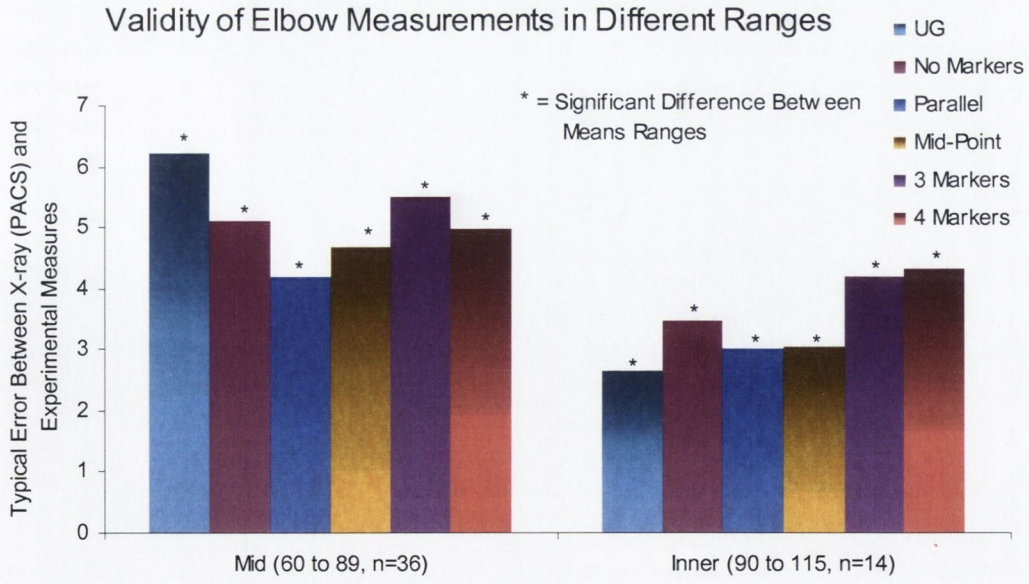
At the elbow the data was divided into mid-range and inner range. The mid-range group was comprised by the angles measures in the range 60° to 89°, while the inner range cohort contain data from 90° to 115° (see Table 6.8). The data was analysed with TE, 95% LOA tests and a two-sample t-test (assumed equal variances) on the differences between the PACS and test data in the different ranges.



Graph 6.15 Typical Error (with t-tests shown) of Agreement with PACS in different ranges for Knee measurements.

The Knee	Outer Range (25 to 59) n= 15					
	UG	*No Markers*	*Parallel*	*Mid-Point*	*3 Markers*	*4 Markers*
Average Difference	-1.51°	-1.44°	2.86°	1.53°	-3.96°	1.04°
Standard Deviation of the Differences	3.64°	4.54°	4.18°	3.69°	5.53°	4.53°
Upper LOA	5.61°	7.45°	11.05°	8.77°	6.89°	9.93°
Lower LOA	-8.65°	-10.34°	-5.34°	-5.7°	-14.8°	-7.84°
Typical Error in Range	2.57°	3.21°	2.95°	2.61°	3.91°	3.2°
The Knee	**Mid Range (60 to 95) n= 27**					
	UG	*No Markers*	*Parallel*	*Mid-Point*	*3 Markers*	*4 Markers*
Average Difference	0.46°	1.06°	3.98°	3.43°	-0.53°	0.81°
Standard Deviation of the Differences	3.91°	3.21°	3.31°	3.55°	3.87°	3.95°
Upper LOA	8.14°	7.36°	10.48°	10.39°	7.06°	8.55°
Lower LOA	-7.21°	-5.24°	-2.52°	-3.52°	-8.12°	-6.92°
Typical Error in Range	2.77°	2.27°	2.34°	2.51°	2.74°	2.79°
Comparison Between Ranges						
Significant Difference Between Outer and Mid Range (p=0.05)	*No*	*Yes*	*No*	*No*	*Yes*	*No*
Difference Between Averages	*1.98°*	*2.5°*	*1.12°*	*1.9°*	*3.43°*	*-0.23°*
Systematic Bias (over 1.6°)	*Yes*	*Yes*	*No*	*Yes*	*Yes*	*No*

Table 6.7. Comparison of the Validity of each tool for Knee measurements in Outer and Mid Range.



Graph 6.16 Typical Error (with t-tests shown) of Agreement with PACS in different ranges for Elbow Measurements.

The Elbow	Mid Range (60 to 89)				n= 36	
	UG	*No Markers*	*Parallel*	*Mid-Point*	*3 Markers*	*4 Markers*
Average Difference	0.2°	-3.31°	2.53°	10.35°	-5°	-3.65°
Standard Deviation of the Differences	8.8°	7.21°	5.93°	6.62°	7.75°	7°
Upper LOA	17.46°	10.82°	14.17°	23.33°	10.18°	10.1°
Lower LOA	-17°	-17.45°	-9.1°	-2.63°	-20.22°	-17.41°
Typical Error in Range	6.22°	5.1°	4.19°	4.68°	5.48°	4.96°
The Elbow	**Inner Range (90 to 115)**				**n= 14**	
	UG	*No Markers*	*Parallel*	*Mid-Point*	*3-Markers*	*4-Markers*
Average Difference	7.33°	3.65°	8.76°	16.03°	1.09°	1.76°
Standard Deviation of the Differences	3.73°	4.9°	4.24°	4.31°	5.93°	6.08°
Upper LOA	14.64°	13.27°	17.08°	24.49°	12.72°	13.69°
Lower LOA	0.02°	-5.95°	0.44°	7.56°	-10.53°	-10.16°
Typical Error in Range	2.63°	3.46°	3°	3.05°	4.19°	4.3°
Comparison Between Ranges						
Significant Difference Between Mid and Inner Range (p=0.05)	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Difference Between Averages	*7.12°*	*6.97°*	*6.22°*	*5.67°*	*6.1°*	*5.42°*
Systematic Bias (over 2.2°)	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*

Table 6.8. Comparison of the Validity of each tool for Elbow measurements in Mid Range and Inner Range

At the knee there were significant differences (p=0.05) between the data in the different ranges for measures made using the No Marker and 3-Marker approaches to the UM. Only the Parallel and 4-Marker approaches did not show a SB between the different ranges. Each of the methods generally measure a relatively larger value for mid range measurements than they did in outer range. TE was relatively static between the two ranges (see Table 6.7; Graph 6.15).

At the elbow, all of the measurement approaches showed a significant difference (p=0.05) in their divergence from the PACS data between the mid and inner ranges. In addition, there were substantial SBs between the two descriptive ranges, with the value of inner range measurements comparatively larger (than the PACS data) than the mid range values. In contrast the TE between the test data and the PACS data would be greater in mid range than in the inner range (see Table 6.8; Graph 6.16).

Analysis of the performance for each of the measurement techniques in the different joint ranges was not a primary goal of this investigation and the

methodology could be criticised on a number of issues. These include some inequalities in cohort numbers, low subject numbers in some groups, loose interpretation of the parameters of each joint range and the failure to gather homogenous data for analysis of all ranges. However, other authors have cited the performance of the UG as being worse in the some ranges than in the others (Enwemeka, 1986; Brosseau, 1997; 2001) and therefore it was felt worthwhile to attempt some analysis of this issue.

6.5.8. Other Research

Four studies were identified that used strong criterion measurements in the form of X-rays, to compare the performance of the UG and other assessment methods of knee flexion angle measurements.

Enwemeka (1986) repeated knee assessments of ten healthy subjects using both X-ray and the UG. The study repeated 6 measurements in a range from 0° to 90° and the conclusions were based on the Pearsons correlation coefficient (PPMCC). Across all ranges the X-ray and UG correlated to an r value of 0.99, with a mean difference of 2.52° ± 6.15°. This current work had a lower mean difference (with standard deviation) values of 0.24° ± 3.9° and a similar r value (0.97).

That study stated that in the outer range (0° to 30°) the mean difference between the X-ray data and the UG was 4.59° (± 4.76 °), which they declared as too high and concluded poor validity for UG measurements in this range. However, as mentioned previously (1.2.1) the standard deviations in this range are actually lower than in their mid range values (30° to 90°, means 0.52° ± 5.58° to 3.81° ±7.2°), this would indicate a lower random error in the outer range than the mid-range making their conclusions controversial. This current study did not consider values below 25° flexion and therefore can not draw direct comparisons against Enwemekas' (1986) outer range data (0° to 30°).

Gogia et al (1987) assessed 30 healthy knees using X-ray and the UG in a range from 0° to 120°. They based their conclusions on correlation coefficients (PPMCC and ICC), they also published the average differences between their X-ray values and the UG data. They had data for measurements by two examiners as well

as simultaneous X-ray values. The correlation coefficients ranged from 0.97 to 0.99, which were similar to the r values for the current study (see Table 6.3, page 276). Their mean difference values of 3° (examiner 1 Vs X-ray) and 1.7° (examiner 2 Vs X-ray) were higher than those seen in this current work and would be classified as holding a SB to the true joint value using the Batterham and George (2000) 2.5% formula. A SB was not seen for UG measurements of the knee in this current trial.

Brosseau et al (1997) investigated the reliability and validity of the UG and the parallelogram goniometer (PG) (see 1.2.1, page 9; Figure 1.2, page 6) using a group of 60 healthy knee joints. Repeated external measurements were made in addition to radiographs. Their results were based of the PPMCC and show that the UG generally scored better than the PG for validity and reliability of knee measurements. They divided their data into two groups, large knee flexion values (103° ± 5.2°) and small knee flexion values, i.e. towards extension (16.2° ± 3°). They found that small knee flexion values had a much lower r value (between 0.33 to 0.48) than larger flexion values (0.73 to 0.78). The average differences between the X-ray data and the UG data was 2.4° for the larger knee flexion values and 6.85° for the smaller flexion values (relative extension). This reflected a substantially larger non-random error in relative extension over that seen for larger knee flexion measurements (relative flexion), a finding not reflected in this current work (see Table 6.7). Both of their r values were much lower than those seen noted in this present results (see Table 6.3).

A follow up study by the same research group was published a number of years later (Brosseau et al, 2001). This time they examined a group of 60 subjects with residual musculoskeletal impairments of the knee. Two examiners both repeated assessments of the subjects in relative flexion (118° ± 23.85°) and in relative extension (4.75° ± 3.75°) with both a UG and a PG. Results for the validity section were again based on PPMCC. On this occasion knee flexion correlated well with the X-ray data for both methods (UG r=0.97-0.98, PG r=0.97-0.98) and relative knee extension measurements showed a drop in performance for both devices (UG r=0.39-0.44, PG r=042-0.51). In contrast to the results from their previous study, on this occasion there was no SB present in the UG data for large knee flexion values

(1.59°) and a smaller non-random error in the outer range measures (3.82°) than those in their previous results. These figures were about double those seen in the current results (see Table 6.7), while the r values for the total data was r= 0.97, similar to their relatively flexed values.

There is consensus regarding poorer measurement performance of the UG (and the PG) at knee extension than its performance for relatively flexed knees (Enwemeka, 1986; Brosseau, 1997; 2001). However, analysis of random error was not conducted in those investigations which limits comparisons to the current study. Nonetheless, the mean difference values (SB) from the current results do not concur with those published in other studies, this study showed no SB in either the larger or smaller knee flexion values. The only study to show no SB in the data was for large values (relatively flexed) by Brosseau et al (2001). That study also showed similar PPMCC values to this current work (see Table 1.6, page 26; Table 6.3, 276).

The only study identified in the literature that attempted to investigate the validity of the UG at the elbow was undertaken by Petherick et al (1988). That study claimed validity for both the fluid filled goniometer (FFG) and the UG based on comparisons between each other. As discussed in Chapter 1, (1.3) it is misleading to consider this type of methodology as a validation study because it does not use a strong criterion measurement and the reader is unable to determine the relationship of either of the tools measurements to the actual bony orientations. Therefore, no comparisons were made between these studies. The lack of data regarding criterion based validity for measurement of the elbow is addressed by this current study. These results show a poorer performance for the UG at the elbow than at the knee. The UG will generally measure values 2.2° greater than X-ray and TE will be in the region of 6°.

6.5.9. Skin Markers

There is controversy regarding the use of adhesive skin markers as assistive means to osteokinematic measurements. Fish and Wingate (1985) state that skin markers are necessary for photographic measurements, and that using skin markers for photography improves upon the reliability of UG measurements. However, the

reliability of replacing their markers or photography without the use of markers was not addressed by those authors.

Reapplying the markers is an issue that Karkouti and Marks (1997) found added to the error associated with photographic based systems, especially over an extended period of time. They believe that if skin markers weren't necessary photographic measurements would be easier. This statement was the basis of the investigation of Aim (4) (see 6.2) which was essentially to compare the use of the 3-Marker approach to the No Marker approach. Both of these approaches to the UM use the same traditionally held anatomical landmarkers points for their measurements (Norkin and White, 2003).

There would be support for the use of adhesive markers if the validity scores (TE, 95% LOA, SB (Yes/No)) for the 3-Marker approach were better than those calculated for the No Marker method, based on the fact that they both use the same reference model.

At the knee the No Marker approach had a TE with the PACS data of 2.75°, a LOA spread of –7.4°, 7.8° and no SB in the data. While the 3-Marker approach had a TE of 3.37° (an additional 0.62°) a LOA spread of –7.6°, 11° (an additional 1.7° spread) and had a SB in its data. In addition there was a significant difference ($p=0.05$) between the 3-Marker data and the PACS measurements, while no such difference was present in the No Marker data.

A similar story was seen at the elbow, the 3-Marker method had an additional TE of 0.3° (5.5°), an additional 0.9° LOA spread (–18.5°, 11.8°) over the No Marker method, the No Marker method had no significant difference with the PACS data ($p=0.05$), while the 3-Marker method had ($p=0.05$).

Therefore the method that allowed the examiner to simply estimate the position of the anatomical reference points from the 2 dimensional photograph, based on the skill of the examiner (No Marker) had consistently improved validity results over the 3-Marker model to photography. The adhesive markers were seen to actually lessen the validity of the calculations rather than improve upon them. This was a surprising result.

Paradoxically, the 4-Markers approach to the UM showed some improvements in error scores over the 3-Marker approach (see Table 6.3, 6.4; page 276, 282) while remaining generally worse than the No Marker approach.

These results can be explained in a number of ways. One possible reason that the No Marker approach performed better than the 3-Marker approach may be due to the fact that identifying bony landmarks by palpation can be difficult. Besier et al (2003) state that accurately placing adhesive markers is the greatest source of error facing dynamic motion analysis. Selfe (1998) found that the intra-rater reliability of placing adhesive markers on the exact area of skin showed a 95% confidence interval of within 1.5 cm. Placing markers on over-weight subjects may be even more difficult (Marks and Karkouti, 1996).

Obesity is a risk factor for joint disease (Aoyagi et al, 2002) therefore it is likely that there was a higher incidence of over-weight people attending for routine joint X-rays and physiotherapy assessment.

This notion was backed up by the subjective experiences of the examiner during the measurement procedures. There were numerous over-weight subjects examined, the precision of the identification and marking procedures was challenging and probably held substantial error. This may be one reason to explain the poor validity results for the 3-Marker method.

Another issue to explain the results may be that the No Marker approach allows the examiner the freedom to align the VMAs to the position that subjectively 'fits' best. Essentially, when placing the VMAs for the No Marker approach the examiner may be positively biased by the longitudinal axis of the limb segment rather than being solely driven by visible skin markers. Therefore the procedure mirrors the Parallel approach to some degree. That approach directly seeks the longitudinal axis of each limb segment, similar to the X-ray measurement procedures, and was shown to have had the lowest random error scores of all the techniques. It is not surprising that the X-ray, the Parallel and the No Marker approaches agreed well as they share similar VMA placement techniques, all being influenced to a greater or lesser degree by the longitudinal axis of the long segments.

A postulated reason why the 4-Marker approach improved upon the 3-Marker approach relates to the fact that human motion is not a perfectly rigid mechanical linkage system and does not employ simple spherical or axial joints (Xudong et al, 2003). Nordin and Frankel (1989) and Norkin and Levangie (1992) agree and state that there is normally glide, pivot and roll between the joint surfaces giving so-called hinge joints a dynamic centre of rotation. This model of joint arthrokinematics implies the paradigm of a fixed 3-point joint model that pivots

about one static definite point has limited validity to osteokinematic science. Therefore an individual segment model that marks the limb proximately and distally, such as the 4-Marker approach, can ignore the joint centre and therefore may offer a better alternative to this type of measurement.

The UG would not be capable of this task because it has a fixed single pivot. This appears to be the reason that Brosseau et al (1997, 2001) investigated the parallelogram goniometer (PG) (see 1.2.1; Figure 1.2). Unfortunately, those authors were not able to show improvements in either validity or reliability for this device, showing that the PG was not able to successfully manage the difficult task of calculating the orientation of the segments independently. Theoretically, the UM can manage this task for static joint angles assessments. The system calculates the angle between the VMAs regardless of their spatial relationship and therefore allows them to be independently aligned to each limb segment. This may account for the generally better results seen for the UM over the UG.

However, in contrast to the view of a dynamic joint centre at the knee Churchill et al (1998) assessed knee joint flexion angle on 15 cadavers and found that the transepicondylar axis closely approximated the true angle of joint incidence. They found that the true flexion axis passed through the centres of the posterior femoral condyles, but there was no significant difference between the angles at each focus point on their sample. The average difference between the angles was 2.9°. However, their results are based on a small number of cadaver specimens which may be a methodological limitation.

The changeable nature of the joint centre is recognised as a real problem in the field of dynamic motion analysis (Xudong et al, 2003; Besier et al, 2003). In order to overcome the need to obtain a location for the theorised axis of rotation, Besier et al (2003) recommend the use of the 'technical coordinate system' model termed the 'CAST' technique. This system clusters three or more markers on each limb segment and allows independent tracking of the segments. The analysis system then constructs the angle of incidence between the segments as the true joint angle, regardless of the precise location of the supposed anatomical axis of rotation. However, they say that technical difficulties still remain with this systems operation. Therefore, currently motion analysis systems normally attempt to identify the anatomical joint centre. Xudong et al (2003) believe that this point is very difficult

to accurately define from surface markers and that a computer driven algorithm is necessary to track it.

In the X-ray assessments observed in this current study the true centre of joint rotation was interpreted by the point of intersection of the PACS VMA lines (see Figure 6.15). The VMAs were aligned with the approximate longitudinal axis of the femur and the tibia, in the same way as previous authors (Brosseau et al, 1997, 2001; see Figure 6.16). In our calculations this point always fell slightly anterior and superior to the lateral epicondyle (see Figure 6.4, 6.5, 6.6, 6.7, 6.8, 6.9, 6.15), as oppose to the posterior aspect as reported by Churchill et al (1998).

Between subjects and throughout different joints angles the centre of rotations were subjectively seen (by the author) to vary somewhat, which seems at least to add to anecdotal evidence to support the concept of a poorly localised centre of joint rotation. This was true for both for different joint angles and different subjects.

In addition to the changeable location axis of rotation, a dynamic relationship between the skin and the bony structures has previously been demonstrated (Sati et al, 1996). This contributes further to the difficulties of trying to track the specific joint centre from one surface marker.

For the reasons of; a poorly definable joint axis, change in soft tissue position relative to the hard tissues during motion and poor reliability and validity of marker placement, an independent segment model with two or more markers per segment, such as allowed by the UM or the CAST technique (Besier et al, 2003) may be consideration as the most valid technique for designers of static or dynamic motion analysis systems.

Static measurement devices should favour a move away from the fixed 3-point model of large joint angle assessment. The evidence from this study in the form of better validity scores for the 4-Marker approach to the UM than those seen for the 3-Marker approach, support this concept. Therefore, if a clinician or researcher were to favour the use of adhesive skin markers for photographic measurements, a segmental approach rather than a 3-point approach should be used.

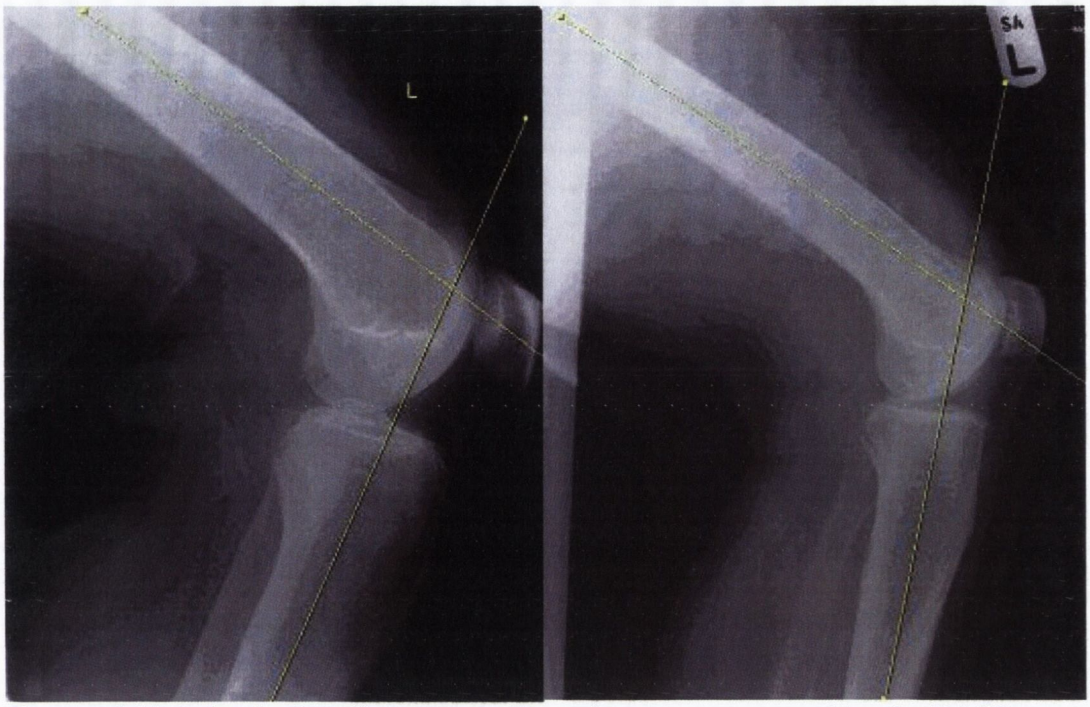


Figure 6.15 Individual Differences in Joint Axis as represented by the PACS measurement lines (yellow).

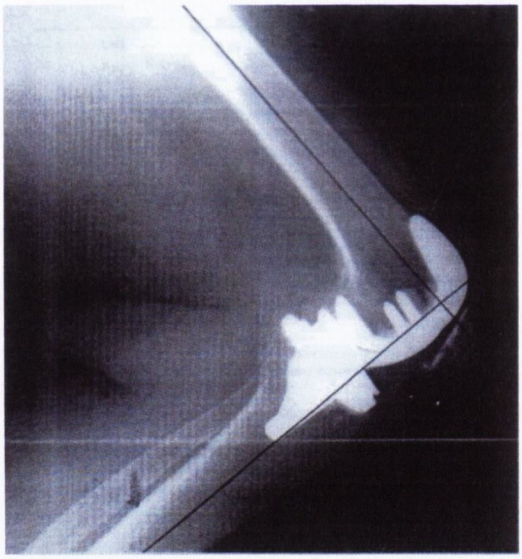


Figure 6.16 Example of X-ray Measurements made by Brosseau et al (2001), lines drawn through the longitudinal bone axes (Black lines).

6.5.10. Limitations

The change in subject position during the knee examinations was necessary, but created the potential to introduce true variation in joint position between the external measurements and the X-ray calculations. In fact one subject was excluded from the study on the grounds that they moved after the external measurement, just prior to the X-ray procedure. It is possible that other minor changes in joint position were present in the experiments, this may have added further random error to the results of both the elbow and knee.

To negate these sources of error in the methodology, it would have been better to examine the subject from the same lateral view and without a change in position between measurements. It is not clear how this would be achieved in a practical manner without any risk to the patient such as the use of unapproved equipment.

Another limitation to the study that arose from the restrictions introduced by the use of the standard radiographic procedures was the limited positions that the joints were assessed in. This meant that the data gathered was in a reasonably narrow range and did not reflect the entire expected challenges that each assessment technique would meet in a clinical setting. It also meant that attempts to answer questions proposed by some authors regarding the variation in performance of external measurements for a wide range of different joint positions were limited (Enwemeka, 1986; Brosseau et al, 1997; 2001).

Some authors investigated the reliability (Williams and Callaghan, 1990; Watkins et al, 1991; Youdas et al, 1993; Burton et al, 1999; Brosseau et al, 2001) and validity (Brosseau et al, 2001) of visual estimation (VE) as a means of joint angle estimation (see 1.2.5). From these studies, there is some controversy regarding the reliability of this method of assessment. With regard to validity, only one study used a criterion measurement against VEs. Brosseau et al (2001) found that VE had a correlation with radiographic measurement from r=0.24 to r=0.94, which was similar to the results for the UG. The limited analysis (PPMCC) and wide ranging results implies that this area remains controversial. Unfortunately this current study did not attempt to investigate VE, which in retrospect seems like a missed opportunity.

One of the findings of this study was that measurement results are likely to change from joint-to-joint. This finding backs up similar findings from previous

studies (Boone and Azen, 1979; Fish and Windgate, 1985; Bovens et al, 1990). It also implies that the validity results for the UG and the UM are only applicable to the elbow and the knee and can not be related to other joints.

6.5.11. Further Work

Further investigation of VE at the elbow, knee and different joints should be considered worthwhile as it is commonly used clinically (Coxford et al 1998). Questions regarding this technique could be combined simultaneously with investigations regarding other external measurement tools such as the UG and the UM at other joints.

Further knowledge regarding the error associated with PACS measurement assistant tools (MATs) should be undertaken in order to establish the capacity of these tools for the provision of criterion data for further external measurement trials.

Image handling software was used to 'over-lay' the X-rays and the photographs in order to display the disparity of orientation between the UM and PACS VMAs may provide an interesting platform for further research regarding the precision of palpation points on the body, such a study may consider identifying points on an obese versus non-obese population. In Figure 6.8, 6.9, 6.13 and 6.14, the position of the bony features can clearly be seen with regard to the position of the adhesive markers. A study could use this type of image handling software to accurately calculate the distance from palpated bony point to the radiographic interpretation of the points.

The data provided by this study may allow further development of the UM or other photographic based systems, and the conception of new approaches to assist the examiner when placing the VMAs of the UM.

Once the recommended photographic procedure for any new method remains the same, new techniques may easily be tested on the existing photographs and against the existing PACS data.

6.5.12. Summary

This set of investigations successfully achieved all of its proposed aims. Concurrent validity was investigated for the UG and 5 different approaches to the

UM, using justifiable and clinically relevant data analysis tools. Both the knee and elbow joints were assessed.

The results are presented in terms of random and non-random error. Non-random error did not reflect the random error scores, which presented somewhat of a paradox to interpretation. The apparent contradiction proposed by the data analysis was whether validity is better in a method that tracks changes seen in the criterion value in a predictable way, despite consistent value errors (either over-estimation or under-estimation), or is it more valid to use a method that closely matches the true criterion data, but in a poorly predictable and inconsistent manner. Another alternative is to elect as the most valid the method that can perform reasonably well with regard to both these forms of error.

At the knee two approaches had a similarly small TE in their association with the PACS data, the Parallel and the Mid-Point methods (2.6°). This result means that both these methods behaved similarly to the PACS calculations for changes in the joint positions. However, both these methods had the highest SB against the true value of the PACS data, they would both generally under-estimate the value by 3.6° and 2.7° respectively. On the other hand, the No Marker approach and the UG had no SB to the PACS data, i.e. they over-estimated and under-estimated the amount equally by a random amount, however, they both had an additional amount of TE with the criterion data.

There does not seem to be any justification for using the 3-Marker approach to the UM at the knee as it possesses high levels of both non-random and random error.

At the elbow, measurement error was generally higher than those seen at the knee for all techniques. This implies that measuring the knee provides more valid data than measuring the elbow.

Only the No Marker approach had low levels of non-random error at this joint. By far the greatest SB was seen at the Mid-Point approach (+ 12°). However, it was not possible to write this approach off as having poor validity as it had the second lowest random error measurements (4.6°). The lowest being the Parallel approach (4.3°).

A balanced view of the validity scores considering random error, non-random error and the performance at both the knee and elbow joint seems to present Mid-Point approach as having the most *versatile* validity scores. This could be chosen on the grounds that it was the only method that had a low non-random error score for both joints and that at the knee there was only an additional 0.15° of TE over the best score (Parallel). At the elbow this method had an additional TE of 0.8° over the best score (Parallel) (see Figure 6.18).

One detracting finding for the No Marker approach was the fact that it was one of only the two methods that had a significant difference between its measurements for mid-range and outer range at the knee (see Figure 6.15). However, while this measurement technique generally under-estimated the measurements in outer range (-1.4°) and over estimated data for the mid-range (1°), neither value exceeded the limits of SB.

The UG (2.75°) out performed the random error scores of the 3-Marker approach (3.4°) at the knee, to have the second worst TE scores. However, it had the worst TE scores at the elbow (5.9°). This supports Fish and Wingate (1985) proposal that a 3-Marker photographic model performs better than UG at the elbow. However, the current study found this statement is not true at the knee.

Changes in joint angles are best detected by the Parallel and Mid-Point approaches, while the No Marker and 4-Marker approaches are more reflective of the true joint angle.

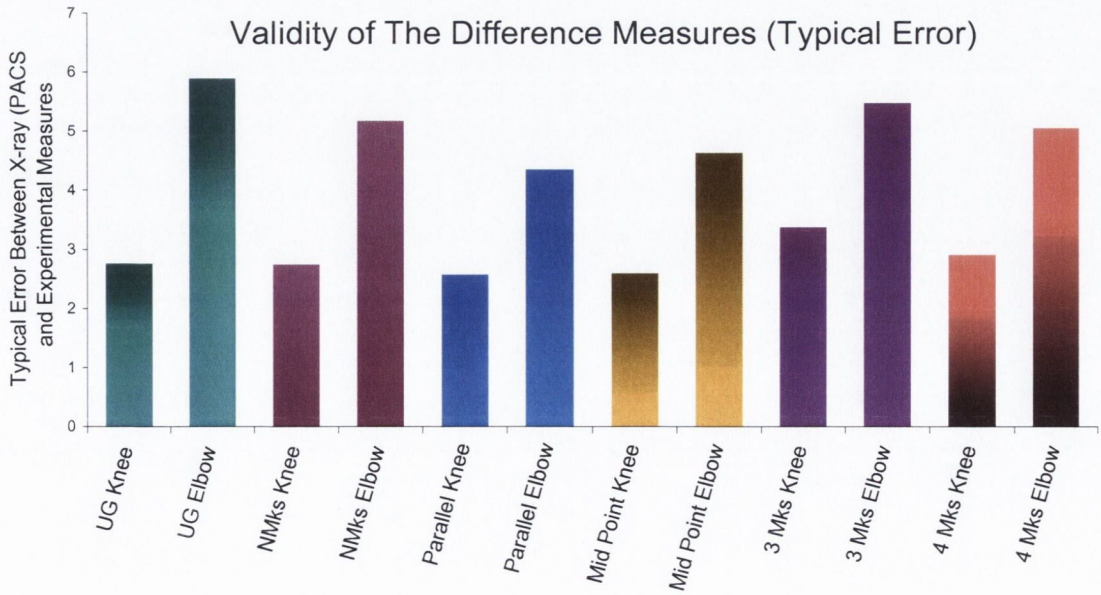


Figure 6.18. Typical Error to the X-ray data. All measurements.

6.6. Conclusion.

Validity of measurement is at the core of a scientific approach to health care and therefore knowledge regarding the validity of measurement tools is paramount for health professionals. It is important that the joints angles recorded by clinicians are reflective of the true joint angle.

This study examined the validity of the UM and the UG for the measurement of sagittal elbow and knee joint angles. Five different approaches to the placement of the virtual measurement arms (VMA) of the UM were investigated.

A rigorous methodology was used to test the measurement techniques on a patient population. Simultaneous measurements obtained from X-rays were the 'gold-standard' criterion against which the precision of each external measurement techniques was assessed. The Picture Archiving and Communications System (PACS) which is a computed radiographic system was used to assess the criterion data (X-ray). As expected this system has some level of reliability error associated with it which must be appreciated when considering the results.

The study backed up previously published views that the UG (and other tools) will perform differently at different joints. With this in mind, this study is unique and important as no other study was identified in the published research that used a criterion based research design to investigate the validity of the UG at the elbow. Results were discussed in terms of random and non-random error, which is the most appropriate and comprehensive method of considering measurement agreement.

The No Marker approach to the UM consistently scored well for both types of error at both joints. The Parallel and Mid-point methods consistently had the lowest random error, but relatively high non-random scores. There was no evidence to support the use of adhesive markers to improve the validity of photographic results.

This investigation considered the criterion validity aspect of the measurement tools performances only. In order to elect the most clinically relevant tool, the results for the reliability of these tools should also be considered. Therefore, these results could be seen as another piece of the jigsaw regarding the performance of our osteokinematic measurement tools, but ultimately should be considered with the reliability results.

Chapter 7

A CLINICAL TRIAL OF THE UILLINN METHOD©.

7.1. Introduction.

The universal goniometer (UG) remains the most popular of the tools available for joint angle measurement (Coxford et al, 1998; Burton et al, 1999). Early designs of this tool have been reported in the French medical literature from the early 1900's and there seems to have been little change in its design over time (Berryman Reese and Bandy, 2002). Despite this device being around for approximately 100 years there are still controversies regarding its performance as a measurement tool. Investigations of validity and reliability of the UG have been described in Chapter 5 and Chapter 6, however, it is clear from the literature review (Chapter 1) that further research into this tool is needed, particularly for its performance at joints other than the knee and elbow.

The experiments described in this work so far have explored whether the Uillinn Method© (UM) has anything to offer over the UG for clinical osteokinematic measurements. The system was intended to help the examiner be more confident in the observations of a change in a patient's joint range over time, as it allows an infinite number of re-examinations of the recorded image. This feature may make it easier to draw comparisons between joint positions from different points in a treatment cycle or rehabilitation programme.

Validity and reliability studies for the UM were described in conjunction with those of the UG (Chapter 5, 6). The results from these investigations were promising and showed improvements over the performance of the UG.

In the literature review some other tools have shown improvements over the UG. One such tool was the NK Hand Assessment Laboratory joint motion module system (NK goniometer) [NK Biotechnical Engineering Company, Minneapolis, Minn] (see Figure 1.6). This tool has a similar design to the UG except that it encompassed a digital read out. This minor functional modification had a substantial improvement on the reliability correlation scores (see Table 1.2, page 22). Despite these improvements the NK goniometer is not often cited in the measurement literature. A small verbal survey of 9 practicing physiotherapists in Ireland in 2004 revealed that none had ever used that device or even heard of it.

The apparent poor uptake of the NK goniometer may have been due to a multitude of reasons, which may include issues such as equipment costs, time constraints, device mobility, a non user friendly design or other practical barriers. It

is not possible to confirm or refute these assumptions as the literature regarding the NK goniometer or other new measurement equipment does not appear to have investigated these issues. This presents a lack of knowledge regarding the functional aspects of integrating new technologies to the clinical environment.

It was felt that to thoroughly investigate the UM it was important to provide physiotherapists with an opportunity to use the system in a clinical setting and then to report their experiences and observations on any barriers encountered. This information could be valuable for the further development of the UM, different osteokinematic measurement tools or other new technologies.

This study had four primary goals. Initially it aimed to set up a trial for the UM in a clinical environment. When this was successfully negotiated a training programme would be provided for the therapists involved. The success of the training programme would be assessed and the appropriate personnel would then be asked to use the system of a patient population over a 3-month period. Ultimately the study aimed to provide qualitative data on the experiences of the therapists who used the system and quantitative data on the nature of the systems usage.

7.2. Specific Aims of the Study

1. To facilitate a 3 month trial of the UM in a hospital physiotherapy department.
2. To provide and assess a training programme for a group of physiotherapists on the operation of the UM.
3. To evaluate the physiotherapists experiences of using the system.
4. To conduct an audit of the number and type of patients assessed during the trial period.

7.2.1. Justification of the Aims.

Aim (1). To facilitate a 3-month trial of the UM in a hospital physiotherapy department.

The UM was conceived as being a joint angle measurement tool that could be used routinely in a clinical setting. However, it was not known whether this application was realistic or how user-friendly the design of the system was. This type of information could only be gathered from a population who had used the system in an appropriate environment.

The most obvious method of acquiring this data was to facilitate a clinical trial of the system by therapists. After a trial period of using the system in a hospital setting it was felt that therapists would be familiar with the systems use. This familiarity would enable the collection of information about the system from this experienced population. The aspiration to collect this type of data justified the requirement for the clinical trial.

A period of 3-months was thought to be adequate for physiotherapists to acquaint themselves with the system prior to being questioned. It was considered that a system of joint angle measurement would not receive equal use throughout all of the different physiotherapy specialties as some areas of practice such as respiratory care do not routinely assess available joint movement. Therefore, only a number of specialties were chosen to be involved in the study.

Aim (2). To provide and assess a training programme for a group of physiotherapists on the operation of the UM.

Current objective joint angle measurement practice mainly uses either the UG (Coxford et al, 1998; Burton et al, 1999) or visual estimations (VE) (Williams and Callaghan, 1990; Coxford et al, 1998). Changing an established practice can be difficult and may require more than just a change in the availability of equipment. There is usually a resistance to change from the routine (Daft, 2001; Curtis and White, 2002).

It was felt that a period of supervised use may help to overcome any reluctance to change routine practice and to consider the UM as a viable option for clinical measurement. A training programme was deemed an appropriate platform for supervised exposure to the system.

More fundamentally a training programme would be necessary to provide instruction in the operation of each of the 3 stages or the UMs procedures. These stages are; the Photographic Procedure, Transferring and Storing the photographic files, and finally placing the virtual measurement arms (VMA) to calculate the angular data. The techniques associated with each of these stages would have to be shown to the therapists in order to use the system successfully.

It was expected that there would naturally be different aptitudes for using such a computer based measurement system, which would present heterogeneity of training requirements in the target population. Therefore, the success of an instruction programme would have to be assessed to ensure it imparted adequate skill to each member of the group.

There was a need to design an examination process to gauge both the knowledge regarding the operation of the UM and also to assess the practical skills needed to use the system. The theoretical issues could be assessed using a series of closed questions, while observed variability in repeated assessments of a joint could be compared against previously established reliability parameters in order to assess practical skills. On successfully passing both the theoretical and practical sessions each therapist could be thought of as safe to use the system on a patient population.

Simply put, the justification for this aim would be to familiarise the involved personnel with the operational procedures of the UM through a training programme.

An examination procedure would ensure good standards of measurement data would be provided by the system.

Aim (3). To evaluate the physiotherapists experiences of using the system.

This aim describes the central purpose of this investigation. After the 3-month trial the therapists would be in a strong position to give an account of how the UM affected their practice, including their perceptions of the advantages and disadvantages of the system. It was decided to gather this data by using a written questionnaire. To ensure that the therapists provided a reflective overview of the system they were asked to keep personal records of experiences. This was done in an effort to ensure that when they came to completing the questionnaire they would be able to fully report their experiences.

Most of the conclusion of this qualitative study would be drawn from the information gathered from the questionnaires.

Aim (4). To conduct an audit of the number and type of patients assessed during the trial period.

In addition to the subjective experiences of the therapists towards the clinical use of the UM, gathering quantitative data would allow the reporting of the average usage of the system amongst the therapists. Essentially this would establish the prevalence of systems use over the 3-month period.

The assessment of the knee and elbow joints only would be recommended during the trial and the audit would attempt to give an account of which joint was most frequently assessed. Such information could be interpreted in conjunction with the qualitative data gathered from the questionnaires.

7.3. Methodology.

The physiotherapy department of AMNCH hospital, Dublin, was chosen as the venue for the 3-month field trial. The specialities that were involved in the study were; adult out-patients, paediatric out-patients, hydrotherapy and the elective orthopaedic department. Approval for the trial was obtained from the department manager and the senior physiotherapists in each of the specialty services.

The patient type that was recommended to be assessed by the UM were those attending with elbow or knee range of motion (ROM) deficits. It was proposed that these patients be photographed and measured on initial assessment, at their pre-discharge assessment and/or whenever the physiotherapist deemed it beneficial, e.g. pre and post a particular treatment technique.

All staff involved in the study were trained in the use of the system over a 3-week period. The training programme initially involved a tutorial given by the author which lasted for less than 30 minutes. The tutorial gave some background to static osteokinematic measurement science and introduced the concept of photographic based measurements especially with regard to the UM. The tutorial was accompanied by a 3-page document that outlined some of the concepts associated with the use of the system (see Appendix M). The system was portrayed as a valid and reliable measurement tool that acted as a virtual goniometer and would provide the therapist with a permanent visual record of the exact joint restriction seen at a particular moment in time.

The therapists were instructed as to how to correctly use the digital camera, transfer the images from the camera to the computer, how to name and store each file appropriately, and also how to display the images using the UM software. They were instructed as to how to manipulate the UMs' virtual measurement arms (VMA) in order to obtain measurement data. Each therapist had their own Windows Folder on a dedicated laptop PC in which to store their pictures.

The Parallel approach to VMA placement (2.3) was elected as the method to be used during the trial. This was chosen because early analysis of the knee reliability data showed that it had the strongest results. Therefore it was this

approach that was taught to the therapists. A number of demonstrations were also conducted during this tutorial and any questions were openly answered.

After the tutorial was completed a separate illustrated document was made available to all the therapists which contained step-by-step instruction on how to use the system (see Appendix N). Over the next 2 weeks the therapists were asked to practice using the system at their convenience.

Each morning throughout the trial the author, who was not involved in taking the measurements, set up the system. This procedure entailed plugging the external Smart-Media drive into a laptop PC which was secured to a desk in the physiotherapy treatment area and starting the UM software. The camera was stored locally in a draw in the Adult Out-Patient department. Throughout the trial the author made himself available for procedural queries or technical support.

During the 3rd week post the initial tutorial each of the therapists were individually assessed as to their proficiency in the use of the system. Their performance was assessed using a scoring sheet (see Appendix O), which was divided into 'Theory' and 'Practical' sections.

The Theory section comprised of a number of 'Yes' or 'No' style questions designed to establish whether the therapists understood the concept of the system and the basic operation of the system. During the Practice section their competence towards using the system was judged. Each participant was asked to demonstrate repeated measurements of the authors' knee and elbow.

All measurements made were blind and were made in the presence of the author. The blinding procedure was achieved by positioning a piece of adhesive paper on the screen where the data results were displayed (see Figure 7.1, page 327). The placement of the VMAs was done twice for both the knee and the elbow. The VMAs position was reset between each measurement. The author then repeated the measurements of the same photographs with the system blinded.

Both intra-rater and inter-rater reliability were gauged with regard to the 95% limits of agreement (LOA) figures that had previously established (Intra-rater reliability 95% LOA Knee = 3.5° (see Table 5.5, page 197), Elbow= 4.5° (see Table 5.6); Inter-rater reliability 95% LOA, Knee= -8°, 10° (9°) (see Table 5.16), Elbow = -10.5°, 7.5° (9°) (see Table 5.17, page 210)). By conducting this mini-reliability test, it was hoped that each user would display an acceptable proficiency in the use of the

system. If they did not achieve an acceptable score it was felt that further training may be required. Only after all of the therapists were assessed and passed as competent were patient assessments permitted.

When the therapists identified a suitable patient they was explained to them that a new type of joint angle measurement involving photography was available and were asked for permission to for its use. Only if they were agreeable was the system used.

During the training each therapist stored their photographs in the Windows Folder particular to them. Therefore, each new photographic file present in their Folder was an indicator of the use given to the system. In order to audit the amount of use the system had over the 3-month period the total number of pictures in each of the therapists Folders were counted. In addition, the number of recorded knees and elbows were counted to ascertain the ratio of use at each joint.

The qualitative data was gathered using a custom made questionnaire. This 7-page document (see Appendix P) contained a total of 31 question, 20 closed and 11 open. It was piloted on 2 therapists who worked on a part-time basis, one in the out-patient department and the other in paediatrics. Both therapists had attended the training session and had not used the system on a patient population. The questionnaire was piloted within 2-weeks of the completion of the training period. Some minor changes in the text of the questions were undertaken in order to improve clarity. At the end of the 3-month period the revised questionnaire was distributed to the 13 remaining physiotherapists.

The questionnaire was designed as an assessment tool to provide information on a number of issues relating to the use of the UM, as experienced by each therapist. For ease of interpretation these issues were categorised under 4 headings; *Training Issues, Operational Issues, Professional Issues and Design Issues*. These headings were chosen on practical grounds rather than on recommendations from other authors and there may have been some cross-over in categorisation of the questions, i.e. question 25 asked 'When you were using the Uillinn Method©, including the training period, did you experience any procedural problems with the system?', this question was categorised under Design Issues as it may have provided

information regarding problems arising from the system design, but it also may have been appropriate to categorise it under Operational Issues or even Training Issues, as it may equally have provided information on these issue. The headings were not included in the questionnaire and the layout of the questions was not always sequential to the headings.

It was felt that some of the questions (question 6 to 23) should only be answered by persons who had used the system clinically. For this reason respondents who reported a failure to use the UM during the 3-month trial were asked to skip these questions, which were a large part of the questionnaire. The author knew and worked with all of the participants which required the questionnaire to be completely anonymous in order to prevent bias.

7.3.1. Training Issues

This set of questions related to the training process for the UM. The therapists were quizzed regarding the adequacy of the training programme, how long before they felt confident in the systems use. The question set also attempted to ascertain whether there was a reduction in their ability to perform the task over time and if they required further training after the 3-month period.

7.3.2. Operational Issues

The questions under this heading related to how the therapists positioned their patients during the assessments, how often they used the system, how long the measurements took and the amount of 'hassle' they felt was associated with the measurement procedure. They were also asked if they felt using the system was worthwhile and if they failed to complete any of their measurements.

7.3.3. Professional Issues

Some of the supposed professional concerns were dealt with in this section. The therapists were asked where they felt measurement error would arise from when using the system. They were asked what aspects of the 'assessment and treatment cycle' (as laid out by the European Core Standards of Physiotherapy Practice (2002), European Region of the World Confederation for Physical Therapy) would the UM offer most benefit. The questionnaire also asked what patient group best suited the UM and what other joints, besides the elbow and knee, could it be used to assess. The final issue in this section provided a platform for the therapists to report concerns such as ethics, information protection and perceived barriers to using the system.

7.3.4. Design Issues

This final section contended with design features of the UM. They were queried about the ease of operation of the system, what type of procedural problems they had encountered and what they thought the major strengths and weaknesses of the system were. The questions also asked the therapists whether storing the pictures was advantageous to their practice. Finally they were invited to rank their level of confidence in using the system and what were the factors that limited their confidence when using the system.

The final question (number 31) allowed the respondents to report any other concerns or comments they had about their experiences of the system.

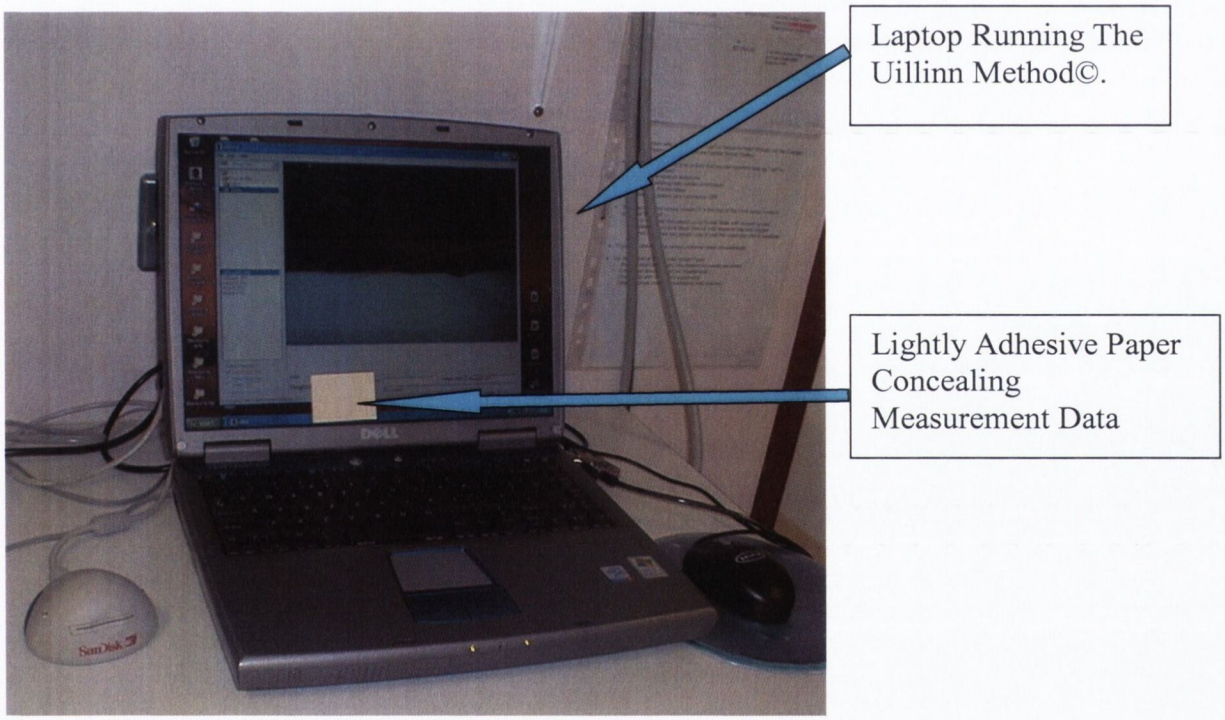


Figure 7.1 Blinding the Uillinn Method©.

7.4. Results

A total of 15 therapists were involved in the UMs training sessions; 2 from the hydrotherapy service, 2 from the paediatric service, 2 from elective orthopaedics and 9 from the adult out-patients service. At the end of the 3-week training period 13 of the therapists were assessed using the scoring system developed for the study (see Appendix O), the remaining 2 therapists were asked to pilot the questionnaire and then took no further part in the trial.

7.4.1. Uillinn Method© Scoring Sheet (Practical Training)

All of the 13 who were scored regarding their understanding of the UM responded positively to all of the questions in both the 'Theory' and 'Practical' sections. Results from this showed that they all displayed an understanding of the concepts and procedures regarding the UM. All of the scores in the practical section of the test fell within the established RC score for the intra-rater reliability (see Table 5.3, 5.9, 7.1, 7.3, page 196, 199) and also within the 95% LOA scores for inter-rater reliability (see Table 5.15, 5.22, 7.2, 7.4, page 206, 219).

Based on these results it was felt that each physiotherapist could effectively use the system clinically and no further training was required.

Knee - Intra-Rater Reliability			RC= 3.3°
Therapist	*1st Assessment*	*2nd Assessment*	*Difference*
1	33.14°	31.51°	-1.63°
2	100.2°	98.78°	-1.42°
3	98.95°	99.27°	0.32°
4	75.1°	74.02°	-1.08°
5	101.2°	99.81°	-1.39°
6	76.65°	76.81°	0.16°
7	60.97°	63.42°	2.45°
8	27.28°	27.37°	0.09°
9	83.26°	83.82°	0.56°
10	30.02°	28.4°	-1.62°
11	83.98°	84.06°	0.08°
12	71.35°	68.19°	-3.16°
13	38.1°	39.2°	1.1°

Table 7.1 Results from the Practical Section of the Training Session. All repeated measurements fall within the Repeatability Coefficient (RC) of 3.3°.

Knee - Inter-Rater Reliability			LOA=-8°, 10°
Therapist	*Examiner 1 (Most Disagreeable)*	*Examiner 2*	*Difference*
1	31.51°	34.16°	2.65°
2	98.78°	99.78°	1°
3	98.95°	99.49°	0.54°
4	74.02°	78.11°	4.09°
5	101.2°	98.12°	-3.08°
6	76.65°	78.39°	1.74°
7	63.42°	61.73°	-1.69°
8	27.28°	30.23°	2.95°
9	83.26°	86.23°	2.97°
10	28.4°	31.07°	2.67°
11	84.06°	81.6°	-2.46°
12	71.35°	69.91°	-1.44°
13	38.1°	41.7°	3.6°

Table 7.2 Results from the Practical Section of the Training Session. All repeated measurements fall within the Limits of Agreement (LOA) of –7.9°, 10°.

Elbow - Intra-Rater Reliability			RC= 4.4°
Therapist	*1st Assessment*	*2nd Assessment*	*Difference*
1	87.46°	89.91°	2.45°
2	60.84°	60.78°	-0.06°
3	65.86°	65.87°	0.01°
4	78.95°	79.27°	0.32°
5	60.6°	63.43°	2.83°
6	55.8°	55.54°	-0.26°
7	59.59°	60.61°	1.02°
8	74.24°	72.16°	-2.08°
9	76.07°	74.51°	-1.56°
10	97.34°	99.9°	2.56°
11	50°	46.5°	-3.5°
12	102.97°	103.16°	0.19°
13	66.08°	66.07°	-0.01°

Table 7.3 Results from the Practical Section of the Training Session. All repeated measurements fall within the Repeatability Coefficient (RC) of 4.5°.

Elbow - Inter-Rater Reliability			LOA= -10°, 7.3°
Therapist	Examiner 1 (Most Disagreeable)	Examiner 2	Difference
1	87.46°	90.32°	2.86°
2	60.78°	61.4°	0.62°
3	65.86°	66.76°	0.9°
4	78.95°	79.49°	0.54°
5	63.43°	58.9°	-4.53°
6	55.54°	61.6°	6.06°
7	59.59°	64.91°	5.33°
8	72.16°	74.7°	2.54°
9	74.51°	75.72°	1.21°
10	99.9°	97.22°	-2.68°
11	46.5°	48.02°	1.52°
12	102.97°	105.3°	2.33°
13	66.07°	68.6°	2.53°

Table 7.4 Results from the Practical Section of the Training Session. All repeated measurements fall within the Limits of Agreement (LOA) of –10°, 7.3°.

7.4.2. Audit of Systems Usage.

The system was used a total of 64 times over the 3-month period and assessed a total of 30 individual patients. One therapist each from the paediatric, adult out-patient and hydrotherapy services did not give the system any use over the trial period, therefore only 10 out of 13 therapists used the system during the trial.

One of the therapists (out-patients) only used the system for knee assessments and three others (1 paediatrics & 2 adult out-patients) only used the system for assessments of the elbow (see Table 7.5; Graph 7.1). Three therapists only used the system twice during the entire period. The average use for the entire group (n=13) was 5 times each. This related to one episode of use per day for the entire group.

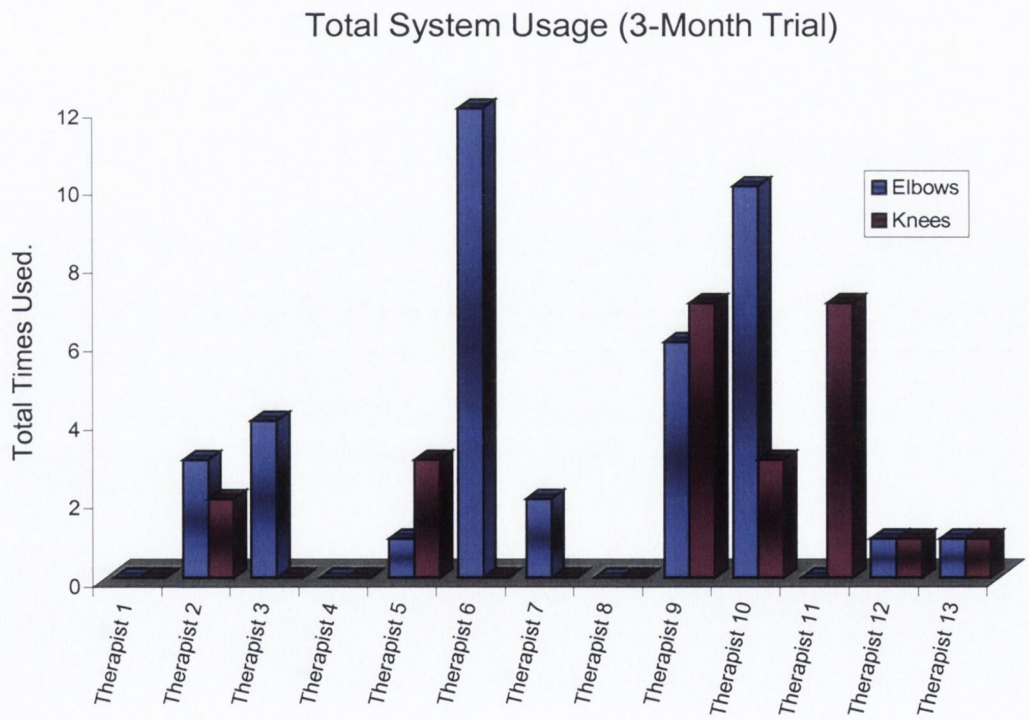
7.4.3. Questionnaire Responses.

At the end of the trial period, questionnaires were distributed to 13 therapists. However only 11 were completed and returned within the allotted 2-week period. Due to the anonymity of the questionnaire it was not possible to track down the 2 missing documents, despite repeated requests to all the therapists involved in the study for their return. This presented an 84% response rate from the small study group. Of the 11 returned questionnaires 4 therapists reported that they had not used

the system during the trial period. This finding was at odds with the audit reported above which found only 3 therapists had not used the system.

		Elbow	Knee	No. of different patients
Hydrotherapy	Therapist 1	-	-	-
Hydrotherapy	Therapist 2	3	2	2
Paediatrics	Therapist 3	4	-	3
Elective Orthopaedics	Therapist 4	-	-	-
Elective Orthopaedics	Therapist 5	1	3	2
Adult Out-patients	Therapist 6	12	-	2
Adult Out-patients	Therapist 7	2	-	2
Adult Out-patients	Therapist 8	-	-	-
Adult Out-patients	Therapist 9	6	7	6
Adult Out-patients	Therapist 10	10	3	7
Adult Out-patients	Therapist 11	-	7	2
Adult Out-patients	Therapist 12	1	1	2
Adult Out-patients	Therapist 13	1	1	2

Table 7.5 Total Usage of the Uillinn Method© of the 3-Month trial.



Graph 7.1 Total System usage by each therapist.

Those therapists who did not use the system were asked to skip a total of 17 questions (6 to 23). Therefore, only 7 therapists out of the original 13 were

questioned on matters that related directly to *experiences* of using the system clinically.

7.4.4. Training Issues

All respondents felt that the training regime of one half hour training session followed by a practice period with an accompanying instruction manual was adequate. The majority (5/ 11) felt using the system twice was enough to be confident in its use. Only 1 therapist needed to use the system more than 3 times (5 times) to be confident with its use. This data seemed to imply that training in the use of the system was relatively straightforward and the therapists found it was reasonably easy to use.

How Many Times Needed To Use Uillinn Method Before Confident? (n=13)

Once	2
Twice	5
Three Times	5
Four Times	-
Five Times	1
More Than Five Times	-

Table 7.6. Training for the Uillinn Method©.

7.4.5. Operational Issues

The approximate number of times each therapist felt they used the system per appropriate patients was 2 (median) (mean 2.2; range 1 to 4). Despite the elbow being the most frequently measured joint, the majority reported the knee as the easiest joint to assess. Two (2 /6) therapists felt there was no difference in ease of use between the joints.

The median time taken to use the system was 2 minutes (average 2.8 mins; range <1mins to >5min). Three of the seven (3 /7) reported that this was too much time, a further 3 (3 /7) felt it wasn't too much time. The remaining therapist reported that it would be too much time for routine patients, but it wouldn't be too much time for chronic or problem patients. On a 4-point lickert scale which ordered the amount of 'hassle' involved in using the system and ranged from 'None' to 'A lot' (question 12), the two mid-point options of 'minimal' and 'moderate' hassle were chosen equally by 6 of the 7 respondents, the remaining therapist reported no hassle when

using the system. All but one therapist reported the extra effort involved was beneficial and worthwhile.

All of the photographs taken during the trial, but one, were analysed using the UM software. The incomplete assessment was reportedly due to an interruption post photography, and this break caused the therapist to forget to complete the measurement.

The patient position favoured for the knee assessments was either fully supine or long sitting. At the elbow the majority (4 /6) photographed their patients in sitting while the remainder had their patients stand during the procedure.

7.4.6. Professional Issues

When asked about the potential sources of error for the UM 2 therapists (2/ 7) reported that they had no concerns on this issue. The remainder (5/7) reported concerns either with accurately placing the virtual measurement arms (VMA) with regard to patient position during the assessments, one therapist felt that guidelines for this latter issue were necessary.

Only one therapist (1 /7) felt that the system should not be adopted to the wider physiotherapy community as they felt some experience was necessary before it should be used. The most commonly proposed clinical areas that may benefit from the system were assessments of pre- and post total knee replacement (TKR) and joint manipulation under anaesthetic (MUA). Other areas suggested for potential use included research, paediatrics, orthopaedics (especially rehabilitation post joint fractures & management of osteoarthritis) and neurology.

It was suggested that other joints that may be suitable for assessment by the UM are the ankle (4 /11), the wrist (3 /11), the shoulder (2 /11), the hip (2 /11) and the metacarpophalangeal (MCP) joints (1 /11).

When asked which area of the assessment and treatment cycle' (European Core Standards of Physiotherapy Practice (2002)) would most benefit from the development of the UM, the majority of therapists (5/7) felt that the system offered most to the initial assessment of the patient. The least beneficial aspect of this cycle was in the area of communication and discharge (or transfer of care) (4/7).

The therapists were asked about the barriers to using the UM in a clinical setting. Those four therapists (4 /11) that claimed did not use the system over the trial period all responded that limited access to the equipment and time constraints had prevented them using it. A further 4 (4 /11) respondents (who had used the system) felt that time constraints were the major barrier. Another commonly cited issue was maintenance and access to the equipment (3 /11). One of the therapists felt the consent to photography was a barrier for the system. This issue was mentioned by 2 other (2/ 11) therapists as an ethical concern towards using the system, the remainder had no other professional (ethical) concerns over using the system.

7.4.7. Design Issues

The initial question asked with regard to the general design of the system related to photography in general and whether there was an advantage in storing photographs of the joint position from a particular assessment. Three out of eight (3 /8) respondents reported no theoretical advantage all agreeing that the measurement data would be recorded in the physiotherapy notes and therefore there would be no need to re-assess old photographs. However, the majority (5 /8) felt that re-assessing photographs on the same day as a new assessment might improve measurement accuracy. They also agreed that the photographs may be used as a biofeedback link for the patient or help when reviewing a case in order to write a discharge summary (1 /8).

A four-point lickert scale asked the therapists to grade how easy the system was to use, the options for this question ranged from extremely difficult to extremely easy. The majority (9 /11) replied that they found the system 'easy' to use, while the remaining therapists (2 /11) found it 'extremely easy' to use. Despite these reports of ease of use the majority (8 /11) only reported 'moderate confidence' with the system. Two reported that they were 'extremely confident' in its use and one therapist (1 /11), who had previously reported the system as 'easy' to use stated they had only 'minimal confidence' in their ability to operate the system.

The main limitation to their confidence appeared to be lack of practice with the system (5 /8) or unsure about accurate placement of the VMAs (2 /8). One therapist (1 /8) reported a general lack of confidence with information technology (IT).

Just over half (6/11) felt that the system could be changed to make it easier to use. The changes suggested included a prompting feature in case a step was forgotten, integrating the transfer, storage and renaming of files from the camera to UM software, improving the zoom feature and making the system generally quicker including a suggestion of developing a measurement system on the camera instead of the computer, i.e. a hand-held system.

There were only 4 problems experienced during the trial. Two of these related to file storage (both were resolved without assistance), another problem occurred when the battery in the camera ran out and finally one therapist tried to store the calculated angle data on the photograph itself, a feature not currently developed.

The main theorised weakness to the system was that it was time consuming to use (8/11). There was also some concern about the need for equipment (2/11) and the potential cost involved (1/11).

Seven different perceived strengths were reported for the system. The most frequently cited of these was that the system provided objective data (6/11), the next most popular strengths were the ability to compare current joint positions with previous photographs (4/11) and a perceived improvement in reliability over the UG (4/11). A user friendly design (2/11), the visual component of the system (2/11) and biofeedback for the patient (2/11) were also quoted as strengths. The final proposed strength was that the system offers the therapist the ability to store multiple assessments on the camera to be assessed at any subsequent time (1/11).

7.5. Discussion.

The UM had been tested for validity and reliability in a previous chapters (see Chapter 6 and Chapter 5) but the question remained whether or not the system was practical for the clinical environment for which it was conceived. The ultimate objective of this study was to determine how often physiotherapists used the system and if they found the system of any benefit to them in their daily work.

This study was primarily designed as a qualitative investigation that would allow therapists to express their experiences and observations regarding the clinical use of the UM. Quantitative data regarding the actual usage the system over the 3-month trial period was provided by a system audit.

An additional benefit of the study was the provision of supporting evidence for the previously calculated reliability data (both the intra-rater and inter-rater) (see Chapter 5). This was achieved by determining the agreement between repeated tests for all of the 13 therapists as part of the training process. All this data fell well within the predetermined limits of reliability.

The four specific aims of the trial were accomplished (7.2). The 3-month trial in a hospital setting was successfully accomplished. Fifteen qualified physiotherapists were trained in the systems use, 13 of these were tested in their ability to use the system clinically, while the remaining 2 were used to pilot the questionnaire. At the conclusion of the trial the revised questionnaire was distributed and an 84% response rate was achieved. The information gained provided valuable data on the clinical use of the UM. The systems usage was also successfully audited.

All of the therapists reported that the training protocol was effective. This involved a 30-minute *in-service* lecture followed by an opportunity to practice using the system for a 2-week period. The Parallel Approach to placing the VMAs (2.3) was chosen as the method to be used during the trial.

Results from the end of training scoring sheet (see Appendix O) showed that all 13 therapists (100% of those tested) could safely use the system. The training regime was further vindicated by responses from the questionnaire, and 10 out of the 11 therapists who responded to that question felt using the system 3-times or less was adequate to be able to use it effectively. The practical aspect of the testing session asked each therapist to repeat measurement of the authors elbow and knee,

including photography of the limb. The author who invigilated over the test sessions, also assessed the recorded joint angle. All of these measurements were blind (see Figure 7.1, page 327). The highest single variation in the repeated measurements by the same examiner (intra-rater) was 3.5° for elbow assessments (see Table 7.1). This was within the RC score of 4.5° and indicated that the system was being used appropriately. The highest variation for similar measures at the knee was 3.16°, also within the established RC of 3.5° (see Table 7.1).

Agreement between angles calculated by the author and each of the examiners also provided evidence that the system was being used correctly. The predicted variation for measurements at both joints was a spread of approximately 9° (95% LOA) (see Table 5.15, 5.22). The highest variation observed at the elbow and knee was 6° and 4.1° respectively, both well within this limit (see Table 7.2, 7.4).

Four out of the eleven therapists reported that they did not use the system during trial. This report was at odds with the audit of system usage, which showed that only 3 out of the total 13 therapists didn't use the system. The obvious meaning of this comparison was that an additional therapist who reported not using the system actually had used it, but only once.

However, a number of other explanations for this mismatch could be proposed. From the audit of photographs in the individual therapists Windows folders, 3 therapists only used the system twice during the trial. It is possible that one of these therapists forgot they used the system or felt using it just twice was not adequate to report themselves as an experienced user. It is also possible that a therapist mistakenly placed their picture in another therapists Windows folder and thereby skewing the results of the system audit. Indeed to back up this theory 2 therapists reported problems when incorrectly storing files during the trial.

Whatever the reason for the mismatch the data showed that 23% to 30% of therapists who had the system available to them failed to use it at all over the 3-month period. Of the remaining percentage (70% to 77%) their use of the system ranged from twice (approximately once every six weeks) to thirteen times (approximately once per week) over the trial (see Table 7.5).

Based on the number of hours worked by physiotherapists in the adult out-patient service, there are at least 65 half hour sessions per working week. Over these

sessions therapists may see patients with a wide range of disorders. It was therefore impossible to directly correlate the use of the system with different patient groups from the data gathered in this trial, i.e. incidence of UM use to patients with elbow, knee or other pathologies, and it is not possible to objectively say whether the systems usage was high or low relative to appropriate patient groups. This posses a relevant but undetermined issue for the study and is a question that should be investigated under further work.

Eleven of the thirteen therapists responded to the questionnaire, a further 4 therapists were instructed to skip questions that required practical experience with the system in order to answered them appropriately. Therefore, only 7 out of the possible 13 therapists responded to all of the questions.

One therapist reported having to practice using the UM 5 times before they were 'confident' in its use, subsequently they reported not using it at all over the term. All of the therapists who claimed to have used the system during the trial needed only to use the system three times or less to be 'confident' in its use during the training period. However, the level of confidence seemed only to be 'moderate' for the majority of the therapists (8/11). The main confidence inhibitor for the majority of therapists was related lack of practice with the system. One respondent admitted to lacking confidence with IT in general.

There was unanimity in reporting the system as either 'easy' or 'extremely easy' to use. Nonetheless 6 out of 11 suggested some changes for the system. These included the need for specific guidelines to placing the VMAs, integrating the file transfer process and generally making the system quicker.

Five therapists felt that the method of placing the VMAs was one of their main concerns as a source of error. This notion was interpreted as a criticism relating to the Parallel Approach (2.3) of VMA placement as oppose to the UM in general. No other approach to placing the VMAs had been discussed with any of the therapists and their opinions on this issue were not directly sought. Specific guidelines for VMA placement were suggested as a way to improve this issue. Nonetheless, the majority of therapists (6/ 11) intuitively felt that a major strength of the UM was improved levels of objectivity and reliability over the UG.

The average time taken to calculate the measurements was reported as over 2 minutes per assessments. This amount of time was reported as the major weakness of the system and a barrier to its use. This finding was in-line with a survey by Kay et al (1998) of attitudes towards outcome measurements, their physiotherapists reported that the main barriers to use was also a lack of time. This is in agreement with an earlier opinion by Williams and Callaghan (1990) who stated that it is important that measurement tools are quick, easy to use and cheap.

Indeed the cost of equipment was another concern mentioned by a number of therapists (2 /11). One idea offered as a means of improving the speed of calculation was to develop a hand-held version of the UM. Which may be an option for future developments, especially with the increasingly popularity and availability of integrated mobile-phone cameras technology.

All of the therapists (7/7) felt that the system would be of most benefit to orthopaedic joint disorders such as an outcome measure for; rehabilitation post fracture, TKR and assessment pre- and post joint MUA. Other potential uses were reported as a research tool or assessment of paediatric and neurology patients. Whatever area the system was to be used in, the majority felt that the system offered most to the initial assessment stage of the patient care.

The concept of photography in measurement had a mixed response. Two (2 /8) therapists raised concerns about patient consent issues. Others reported the visual component of the system was one of its major strengths. The ability to assess photographs from different sessions at the same time was reported as an advantage by 5 therapists (5/ 8). Indeed in the literature there is some evidence that joint assessments made on the same day offer improved reliability over measurements made over separated sessions (Marks and Karkouti, 1996). Another benefit offered by photographs was reported as biofeedback and motivation for patients.

The audit of system activity established the frequency of use but not its incidence of use, i.e., the total amount of use was calculated but this could not be compared to the total amount of potential use (incidence). The incidence of use would have provided interesting data and may have been better able to reflect the percentage uptake of the system by the therapists. Furthermore, such a study may have been able to make comparisons between incidence of use of the UM and

incidence of use of the UG. A Medline search (August, 2004) did not reveal any information on the incidence of use of the UG or any other type of measurement tool.

The prevalence of use of the UM over the trial varied from therapist to therapist. However, in general it was felt that the average use was poor and probably related to the time constraints associated with the busy clinical environment.

The study was largely of a qualitative design. This was chosen because most of the research literature recommends this type of investigation when trying to establish personal beliefs, lived experiences and interactions with the environment (Strauss and Corbin, 1990; Pope and Mays, 1998; LoBiondo-Wood and Haber, 2002). However, in some cases qualitative research has been criticised for lacking rigor (Green and Britten, 1998). This study tried to avoid this failing by reporting as much detail regarding the methodology as possible. A technique recommended by Pope and Mays (1998) to improve rigor of such trials.

The study design successfully provided information regarding the practicalities of using the system clinically by a small sample group. Only 7 out of a possible 13 therapists were entitled to answer questions regarding experiences of using the system. However, no attempt was made to provide inferences towards how a broad spectrum of physiotherapists might experience the UM. For this reason descriptive statistics only were thought to be adequate to analyse the data.

7.6. Conclusion.

Issues of validity and reliability are vital for health care measurements. However, the measurement tools must be practical and accessible. This study monitored the usage of the UM measurement system over a 3-month period and allowed physiotherapists to report their experiences of using the system through a written questionnaire. The study was largely qualitative and did not use inferential statistics.

Training in the use of the system was straightforward for most people and required a short training lecture and three or less practice attempts to be familiar with its operation. The average time taken to use the system was 2 minutes which most felt was too long to be used on a regular basis and was a barrier to its use.

When the therapists finished the training regime they were scored on their knowledge of the system and skill using it. The skill test involved repeated measurements made using the system. This aspect of the test provided a means to check and justify the predetermined reliability agreement data.

The audit of use showed that it was used on average only twice a week by a total of 13 therapists, 4 of which failed to use the system at all over the trial. It is suggested that this represents a low level uptake of the system, although the study design did not gather data to corroborate this theory.

The elbow was more frequently assessed than the knee. Most therapists felt that the UM has a potential clinical use as an outcome measure for intervention with an orthopaedic patient population.

The UM was used infrequently by the majority of therapists in the clinical setting. Nonetheless, it was proposed by the majority of respondents that the system provided a means to accurately check old joint angles against those from a current assessment, and that the visual component could be used as a biofeedback mechanism for patients. It was also proposed that the system might have a valuable role as a research tool.

Chapter 8

DISCUSSION & CONCLUSION

8.1. Introduction

This study explored issues relating to the practice of static joint angle measurement in the clinical environment. The ability to calculate these measurements accurately is of central importance to health care providers. The data can be used to assist in the diagnostic process and act as an outcome measure, which is a vital issue to provide an evidence base for physiotherapy intervention.

The study initially assessed the current knowledge regarding this type of measurement, identified knowledge deficits, and then progressed to examine the performance of the universal goniometer (UG) and a photographic approach titled the Uillinn Method© (UM). The latter method is a new system that was devised as part of this thesis and uses digital photography and a PC software program to form a virtual goniometer with which to assess the joint angle. The relative performance of 5 different methods of applying the UM to the human model were investigated.

There was a need to clearly establish the validity and reliability for both the UG and each of the different approaches to the UM. The knee and elbow were used as the subject matter for the investigations. Qualitative data was also gathered for the use of the UM in a clinical setting.

This final chapter was written in order to provide an overview of the relevant issues discussed in the previous chapters and to consider the experimental findings relative to each other, to re-state the new knowledge presented by this research and to consider the implications of the thesis.

8.2. Over-view of the Literature

8.2.1. The Literature Review

There are two main bodies of movement analysis, static joint angle assessment and dynamic motion analysis. The latter term typically considers measurement of joint angles during observations of function. The main focus of this study dealt with the former term, static joint assessments. This type of analysis usually assesses the end-of-range (EOR) joint positions.

It is widely accepted that these measurements are difficult (Stokdijk et al, 1999) and it has been suggested that no available technique provides infallible data without some amount error (Lantz et al, 1999).

The majority of studies reviewed had flawed methodologies. The main limitations were the type of data analysis conducted. Rothstein (2001) proposes that reliability (and validity) is not an all or nothing phenomena and cannot be simply be accepted or rejected, therefore studies that focused on significance testing are of limited relevance.

There had been a trend in the older published research to focus the expression of reliability or validity results in terms of correlation coefficients. However, this type analysis does not adequately address the issues involved and does not present the clinician with data that is easily interpretable to the clinical environment. Instead, validity and reliability must be gauged using analysis of agreement and a statement of the expected error associated with the measurement (Altman, 1999; George et al, 2000; Batterham and George, 2000).

Other methodological factors such as low experimental numbers and criticisms of experimental procedures have also limited the strength of some of the research. Encouragingly, more recent studies have shown a trend for more robust study design and data analysis.

Despite some methodological shortcomings in the literature, the review was able to establish some issues that have general accord. The first of which was that intra-rater reliability is generally stronger than inter-rater reliability. This finding implies that most examiners carry some amount of individual interpretation to their

measurement practice. This finding provides evidence that examiners contribute noticeably to overall measurement error.

Most authors agree with this statement and concur that further sources of error arise form the measurement instrument itself, and from the subject being measured. This latter source is termed biological error. Therefore, it is important that we recognise the 3 source of measurement error as examiner error, instrument error and biological error.

Another common finding was that measurement protocols improve on the performance of the measurement tools. This underlines the need for guidelines to using any measurement tool. Norkin and White (1995; 2003) have provided these for the UG. However, such guidelines appeared to be absent for photographic based measurement systems.

There is consensus that measuring different joints and even different quantities of joint movement, i.e. different ranges, will provide variations in the performance of measurement tools. This issue reflects the extent of the challenges facing external joint measurement.

The wide and confusing array of results for validity and reliability can be seen in Tables 1.2 to 1.6 (pages 22 to 26). This implied a knowledge deficit surrounding the exact amount of reliability and validity error associated with the UG. In particular gold-standard criterion validation studies for this tool were sparse and no study had attempted to validate this tool at the elbow.

Photographic systems are not a new measurement technique, yet there has been limited research into their validity and reliability. An associated issue was the reliance of photographic systems on adhesive markers. These have been extensively used to highlight anatomical landmarks in the photographs. Limited reliability and validity research had been carried out on this practice at the knee, with less than perfect results. No study could be identified that had examined the necessity of adhesive markers for photographic measurement.

In summary, there were a sizeable number of reliability studies for the UG but only a limited number of criterion validity research. Published results for both of

these issues were confusing and controversial and there was a need for further research. Specifically there was a necessity for clear and clinically interpretable data regarding the expected error associated with the commonly used UG.

Photography was another method that required further investigation, particularly if new digital systems were to be regarded as a viable option for static joint angle assessments.

8.3. Development and Application of the Uillinn Method©.

From the literature review it was identified the there was a need to improve confidence and develop alternative approaches to clinical joint angle measurements. The invention of digital photography presented new possibilities to this practice. In 1997, Karkouti and Marks stated that if a photographic measurement system could deliver an instantaneous picture it may be useful as a clinical measurement tool. The Uillinn Method© (UM) was developed in accordance with this belief and was able to explore and take advantage of the new digital technology. After a lengthy development period the latest adaptation of this system is UM, version 2.4.

During the development process it became clear that one of the main practical challenges associated with the systems use was how best to apply it to measuring the human model. In answer to this issue 5 different approaches were developed (2.3, see page 102). Two of these used skin markers and the remaining 3 did not. It was not clear which of these was the most appropriate for clinical use.

8.3.1.The Rationale for Research

The validity and reliability of both the UG and photographic based measurement systems remained unclear. The newly developed UM was based on photography and therefore employed poorly understood technology. These two issues alone provided a compelling rationale for further research.

A more fundament issue for the UM system was whether or not the newly developed method of calculation was accurate, and how the photographic procedures could affect measurement error. Therefore, there was a need to test these issues prior to assessing its performance on the human model.

A further issue with the clinical use of the UM was whether or not the physiotherapists, for whom it was designed, would actually use the system. A clinical trial of the system would provide data on the amount of usage the system may receive, and if the system was user-friendly or if design changes were required. Such a qualitative trial would provide the experienced users of the system with a platform to comment on its value to their clinical practice.

8.4. Overview of the Experiments.

A series of experiments were designed that would conclusively determine the error associated with the UG and the UM for measurements of the knee and elbow. In order to prevent the type of short-comings seen in the previous literature, these quantitative investigations used relatively large subject numbers, blinded procedures and to use a range of different data analysis tools. The analysis tools chosen were ones that have been recommended by numerous authors (Bland and Altman, 1986; Altman, 1999; Daly and Bourke, 2000; George et al, 2000; Batterham and George, 2000; Hopkins 2000a, 2000b), and the conclusions were therefore primarily based on the analysis of random and non-random error.

Non-random error related to whether a systematic bias (SB) was present between different measurements of the same entity. With regard to reliability a SB between different examiners would indicate that the measurements of one of the examiners was in general larger than another examiner. For validity a SB would mean that the gold-standard measure was consistently recording larger (or smaller) data than the experimental measurements.

Random error was calculated using the Limits of Agreement (LOA), the Repeatability Coefficient (RC) and the Typical Error (TE) analyses. The LOA test provided boundaries of expected agreement between the repeated scores. That is, for 95% of repeated measurements the agreement between the tests would be within the calculated limits.

When there is no SB in the data the RC was used to express the same results. This type of analysis presents the therapist with some confidence as to the possible error in their measurements and is expressed in a clinically relevant way. A low RC (or LOA) implies a high precision for the measurement tool (Daly and Bourke, 2000).

The TE score related to the usual expected error between the repeated measurements. It does not account for SB. This data allows the clinician to determine the amount of change that should routinely be ignored, i.e. if two different therapists measure a patients knee with a UG pre- and post treatment and the patients flexion angle improves by 4.5° (UG TE= 4.5°: see Table 5.13, see page

202), the therapists could expect that no actual change had occurred but rather the different measurement data was due to expected error.

8.4.1. The Uillinn Method© Calculation System: A Calibration Question.

This initial set of experiments were undertaken to investigate how accurate the method of angle calculation was and whether specific photographic procedures were needed to maximise the certainty of the data.

All experiments were conducted in a non-clinical environment. The results found that the angular calculations would agree with a mathematical control by less than 0.82° for 95% of occasions and that TE would be as low as 0.3° (see Table 4.6; Graph 4.8, page 149).

However, if the angle was rotated by more than 5° from parallel to the cameras lens, or parallax were present in the photograph, then the true nature of the angle would be misrepresented and error would be increased. At 30° rotation the RC would be as high as 15° and TE would be approximately 5.5° (see Table 4.12; Graph 4.14, page 161).

This study clearly displayed that the system could successfully calculate angular data. It also highlighted the need for strict photographic procedures that controlled rotation and parallax. These procedures were drafted and presented in a set of user guidelines (see 4.4.7, page 164). At the conclusion of that chapter (Chapter 4) there was a sound understanding of instrument error for the UM in the laboratory environment.

8.4.2. Reliability of the Universal Goniometer and the Uillinn Method©.

Reliability is a fundamental indicator of the performance of clinical measurement tools (Batterham and George, 2000). This type of research must consider two basic concepts; intra-rater and inter-rate reliability. The reliability experiments in this work (Chapter 5) considered both of these issues for the UG and for all of the different approaches to the UM.

There were 3 different examiners involved in the study, each had minor variations in their individual performances. The reliability results (both intra and inter) at both joints were comparatively high for the Mid-Point, 3-Marker and 4-

Marker approaches, with the former method being slightly stronger. That method (Mid-point) found intra-rater reliability for the knee to typically be 1.3° (RC= 3.5°) and 2° at the elbow (RC= 3.8°) (see Table 5.26, page 227). Between examiners the TE for that approach would be 2.2° (RC = 7°) at the knee and 2.3° (RC = 5°) at the elbow (see Table 5.33, page 238).

These results were much stronger than those observed for the UG. Intra-rater reliability results for the UG were; TE= 2.4° (RC= 6.8°) at the knee and TE= 3.2° (RC= 8.7°) at the elbow. Between examiners TE (multiple) at the knee was 4° (RC=13°) and 7.2° (RC= 11°) at the elbow (see Table 5.26, page 277). At the elbow there was also a large consistent error between examiners, which in one case was observed as being over 6°.

To demonstrate what these figures imply to measurement practice when the UG is used, some examples are listed below, the equivalent scores for the UM are listed in brackets;

If the measurements were repeated by the same examiner using a UG an actual change in joint position could be suspected after approximately 2.5° (UM, 1.3°) at the knee and 3.2° (UM, 2.3°) at the elbow. Confidence (95%) in change could be stated for the UG when knee measurements varied by 7° (UM, 5°) and when elbow scores differed by 9° (UM, 3.5°).

If different examiners were involved, using the UG, they could suspect a true change in the knee angle had occurred when readings differed by 4.5° (UM, 2°) and 7.2° (UM, 2.1°) at the elbow. They could not declare with 95% confidence that there was a position change until a difference of over 14° (UM, 7°) was seen at the knee and 16.5° (UM, 5.4°) at the elbow.

These results indicate the limitation of the UG. While it might be considered that its performance is acceptable for single examiner measurements, it is difficult to justify this tool for analyses between examiners, particularly for assessments of the elbow. This fact has negative implications for multicentre trials that use the UG as an outcome measure.

8.4.3. Concurrent Validity of Universal Goniometry and the Uillinn Method©.

Each of the methods were compared to angles calculated from simultaneous X-rays. This gold-standard data was obtained using the picture archive and communications system (PACS) system (6.3.4, see 271). The results showed that there was a SB between the true joint angles and the external measurement techniques for most approaches. The amount of SB varied substantially. Only the No Marker method had no SB with the X-ray data at both joints. Ironically, that approach was the only method that had a SB for inter-rater reliability (knee measurements).

At the knee the largest SB was seen for the Parallel approach (+3.6°), while at the elbow the Mid-point approach had the largest SB (+11.9°) (see Table 6.3, 6.4, see page 276, 282). The magnitude of SB did not reflect how closely each method tracked the changes in X-ray data.

Despite their SB both the Parallel and Mid-point approaches had the lowest random error (TE at knee = 2.6° (Mid-Point), 2.6° (Parallel); TE at elbow = 4.6° (Mid-Point), 4.4°(Parallel)) and the narrowest LOA spread at both joints (see Graph 6.3, 6.4, 6.9, 6.10, page 278 to 286). The SB between the UM (Mid-Point & Parallel) and the X-ray data implied generally larger values for the latter measurement, despite changes in one being closely tracked by the other.

The poorest validity scores were the 3-Marker approach at the knee and the UG at the elbow, these had TE scores of 3.4° and 5.9° respectively. The poor results for the 3-Marker approach implied that the position of the adhesive markers did not closely reflect the X-rays' joint angle.

As both the Parallel and Mid-point approaches best mapped the changes in the X-ray data, they are put forward as the approaches that had the strongest validity and best clinical value.

8.4.4. Using the Uillinn Method© in the Clinic Environment

The UM (Parallel) was trialed by a group of 13 physiotherapists over a 3-month period in a clinical environment. All therapists were tutored in how to use the system, and all showed that they had adequate command of the system prior to using it in their clinical practice.

The feedback from the therapists was mixed. There was a wide variation in the amount of use each therapist gave the system, but generally its uptake was poor. The main barrier to its routine use appeared to relate to its operational time (approximately 2 minutes). Most felt that this was too long.

However, most agreed that it could be used routinely on selected patient groups. These included patients for manipulation under anaesthetic (MUA), patients under going joint replacement rehabilitation and patients with other joint pathologies, such as osteoarthritis (OA).

8.5. Discussion.

8.5.1 The Best Quantitative Approach to Static Joint Angle Measurement.

Validity and reliability experiments had been conducted separately and the election of the best method over-all had to considered the results of both of these trials. Batterham and George (2000) state that it is justifiable to count reliability rather than validity as a more fundamental issue to clinical measurement, because if a measurement tool is unable to reproduce data to a reasonable degree, validity levels will be subject to change over repeated measurements. This notion was considered and the reliability data was the first criteria used to judge the different approaches.

However, it was recognised that reliability does not directly correlate to validity. It is possible to have a reliable measurement that has poor validity, but due to a prerequisite of measurement consistency, it is less likely to have a measure with high validity scores that is unreliable.

This prevented total focus being placed on reliability and called for a comprehensive view of the test data. The election of the best approach considered both issues, with only the initial emphasis on reliability.

The 3-Marker, 4-Marker and Mid-Point approaches all had equitable and high reliability scores. Of these 3 most reliable techniques, the Mid-Point approach had the best validity score based on the random error and narrowest LOA spread with the X-ray data, notwithstanding a SB.

The 3-Marker and 4-Marker approaches achieved strong reliability but relatively poor consistency with the gold-standard data. Therefore, for the 5 different approaches to the UM, the Mid-Point approach scored highest on both the validity and reliability tests and is recommended as the optimal approach to the UM. This system cut error by up to a third in some cases when compared to the standard UG.

The lack of emphasis on non-random error (or SB) when judging which approach had the most important validity score was justified for two reasons; firstly, the primary function of all of these measurement methods is to detect alterations in

the movement of a joint, either with regard to normal values or during repeated assessments over time. This task is best achieved by an approach that displays a low random error, which implies it will closely reflect when actual changes in the joint angle occur.

This ability maximises the sensitivity and specificity of a measurement, which can be defined as; the former term referring to the ability to correctly identify when actual changes occur (a true positive), and the latter term referring to the ability to correctly identify when no change in the angle has occurred (a true negative).

These qualities are important for both clinicians and researchers. For clinicians there will be more assured and confident decision making and a better potential to determine a best evidence approach. For researchers these qualities will help to control both type I (false positive) and type II (false negative) errors in a study's conclusions, thus providing for a better evidence base.

Another reason to discount the non-random error in the data was due to the fact that all but one method had some level of SB in their results. That method (No Marker) had relatively high random error scores making it difficult to defend as the optimal approach.

It may be argued that the SB seen with the Mid-Point approach may be important to some disciplines. However, where the main purpose of the assessments is to detect changes in available movement, a moderate SB seems irrelevant. It is argued that the main strengths of a clinically relevant measurement tool should be the ability to track true alterations in a joint position. This was best achieved by the Mid-Point approach.

8.5.2. Measurement Error

The experience of this study supports the view that precise measurement of human joint motion is challenging. The difficulties with external joint measurements arise from a number of issues amongst which are difficulties associated with the pliability of soft tissue and the presents of a dynamic joint centre. The inherently complex and uniquely individual nature of each person adds to the difficulties of standardised measurements (Hageman and Sorensen, 1991; Bland and Altman,

1996). Psychological, physiological and environmental factors make repeated assessments of body systems difficult to predict. All of these issues can broadly be grouped together as the difficulties associated with biological variation.

Examiner errors are those errors associated with the skill of the assessor. Conducting joint angle measurements on a patient population can be difficult and require knowledge of anatomy and pathology in order to accomplish the assessment with consideration to the patient. Different examiners will inevitably use slight variations in their approach to the measurement. This will culminate in some degree of disagreement between examiners.

Different devices may be more adept at measuring different joints and it may be that some tools are completely inappropriate for selected measurement tasks. When a measurement tool cannot adequately accomplish its job the miscalculation is classified as instrument error.

The total measurement error under these 3 headings (biological, examiner and instrument) has been described as measurement noise (Hopkins, 2000a). Some amount of noise seems inevitable during osteokinematic measurement. Minimising this noise should be ever present in the minds of professionals involved in clinical measurement.

8.5.3. Value of the Research.

The UM was presented as a novel approach to joint angle measurement based on new technologies, in the form digital photography and PC based software. From the literature limited research was identified for the performance of photographic systems. This current study appears to have been the first validity and reliability investigation for photography that used large cohorts, included a patient population (validity) and used a strong criterion.

The study assessed UG simultaneous the UM and could therefore draw comparisons between the two approaches to measurement. It was found that a photographic system had the potential to be substantially more valid and reliable than the standard UG once certain guidelines were followed.

Despite photography being used extensively in the past no guidelines towards optimal clinical procedures were found. This study identified sources of error for this type of system and presented strategies to minimise this error. These strategies were outlined in the Guidelines for Using the UM (4.4.7, page 163). It is felt that the guidelines drafted in this study could be applied to any single source (2-D) static joint angle assessment method, such as the internet based goniometer (IBG) (Russell et al, 2003) or possibly even radiographic measurements (Gray et al, 1991).

Overall, it was felt that system uptake was relatively poor despite therapists intuitively believing that it was a more reliable system than the UG. This begs the question as to how important is the precision of static joint angle measurements to practicing physiotherapists. The evidence from this study suggested by the relatively poor system uptake, is that precision is not the highest priority for them. The most important issue appears to be speed of assessment. Therefore, developers of clinical measurement tools should build this feature into the system design.

The main appeal to clinicians for a measurement tool appears to be speed of application, this is certainly a feature of the UG and may imply that despite the obvious improvements gained by the UM, the UG may be a popular choice for clinicians into the immediate future.

Finding a balance between speed of use versus accuracy of measurement is an issue that is sure to challenge both the developers and users of outcome measures into the future. As the demand grows for an evidence base and outcome measures become even more important, the issue of measurement accuracy is sure to assume a greater weighting in finding this balance.

The speed of the UM (digital photography) system is substantially faster than a conventional photographic measurement system. However, as mentioned the operational time of the UM was the main barrier to its routine clinical use (see 7.4.7, 334). The method that was trialed during the study was the Parallel approach that does not use adhesive markers. The additional time required to place such markers would inevitably present further barriers to using a photographic system in regular

clinical practice, thus providing a practical reason to discard their usefulness in this environment.

Furthermore, based on all the qualitative and quantitative data the use of adhesive skin markers for photographic based static sagittal joint angle measurements of the knee and elbow were not justifiable and are not endorsed by this author for either research or clinical practice. The main limitations to their use arose from uncontrollable skin and adipose movement relative to the underlying bony structures and the appropriateness of the location site of the markers.

8.5.4. The Uillinn Method©: Implications for the Clinical Environment.

The evidence provided by this study for the improved levels of error with the UM were limited to the knee and elbow. However, it is likely these improvements in error would be carried to other joints as well. It may well be that the UM could improve on the validity and reliability of all the orthopaedic joint angles in the body.

In particular, transverse rotations are more difficult to measure than sagittal plane joint angles (Walker et al, 1984; Greene and Wolf, 1989; Bovens et al, 1990; Hayes et al, 2001). This may be due to difficulties attempting to locate anatomical reference points that are not adjacent to the joint. For example, rotation of the hip is commonly assessed using the rotational displacement of the shank of the flexed knee, which is relatively removed from the hip. It may be that an instantaneous photographic system could have a role to play in measuring movement in this plane due to its 2-dimensional (2-D) nature which negates depth.

It is also possible that the UM or other photographic systems may offer alternatives to the assessment of different types of important somatic angles, other than just joint range of motion. The assessment of scoliosis has been informally suggested as one such angle. Currently, patients with scoliosis require an X-ray to gauge the amount of lateral curvature in the spine, it is possible that a series of angles calculated from external photographs could be developed to assess the progression of this type of pathology. Such a system would present a means to reduce the amount of X-ray exposure to this young population.

Further research would be required to establish these proposals. Nonetheless, it is felt that the UM has future implications for general somatic joint and body

angles measurements rather than being limited to sagittal views of the knee and elbow.

The most commonly suggested use of the UM proposed from the surveyed therapists related to use with orthopaedic patient groups. Particularly, with patients undergoing joint manipulation under anaesthetic (MUA), total knee replacement (THR) or post joint related fractures. These are patient groups that would have been commonly seen by the therapists surveyed and therefore were sound aspirations based on experience.

Another area mentioned by the therapists was as a motivational tool for a patient through visual biofeedback. The theory is that when a patient can visualise achievements in joint range recovery they are more likely to comply with the demands of a rehabilitation programme.

The visualised proof of available joint range may also be beneficial to areas outside the health care profession. Litigation is an area where hardcopy evidence is highly valuable, thus a photographic measurement system may have a role to play in this field.

The Tardieu Scale is becoming an increasingly popular method to assess spasticity in the paediatric population (Morris, 2002; Shircore et al, 2004). This scale has been put forward as an outcome score for treatment (Wallen et al, 2004) and as an indicator for Botulinum toxin intervention (Boyd and Graham, 1999). It is said to be a more reflective and reliable way to assess spasticity than other methods such as the Ashworh scale (Morris, 2002; Fosang et al, 2003). However, it has been found that the reliability of the method could be improved upon (Fosang et al, 2003; Mackey, 2004). Assessing the angles of spastic response appears to be difficult using the standard UG which limits the scale. Shircore et al (2004) have used electrogoniometers to improve the assessment of this angle. It is possible that digital photography and the UM may help to improve its reliability.

With the increasing popularity and general shift towards the use of information technology (IT) there will inevitably be further integration of health care information in the digital environment. Such plans are already afoot through various management strategies including the proposed introduction of systems such

as the electronic patient record (EPR) which is designed to collate all of the clinical information regarding a particular patient in one computer file that is available to the multi-disciplinary team. Digital measurement techniques such as the UM may be well suited to integration with such a system.

There may be positive financial benefits to the advent of the UM. The ability to assess joint angles separate from the patient may make it easier and cheaper for specialist clinicians to be consulted regarding a movement disorder. For example, orthopaedic consultants may be able to obtain important information regarding diagnosis and management options from the digital images. Especially if they can manipulate the images and use software based measurement tools. Russell et al (2003) had similar aspirations with their IBG. In that case they proposed a simple way to assess joint angles over large distances, negating the need to travel for repeat assessments.

Probably the most important feature offered by the UM relates to its use of digital photography. This implies an inherent ability to retain and access stored images of joint angles which can be measured and re-measured by both the same and different examiners repeatedly without reference to time. Therefore the system is different from tools like the UG which can only perform measurements at one specific moment.

If different examiners are involved in the measurement of a joint angle, a relatively high level of error is expected, this is due to the established fact that inter-rater error is always worse than intra-rater error. However, the UM will allow subsequent examiners to re-measure original images of the joint angle (initially assessed prior to their involvement) and directly compare that data to the current observed angle. Therefore, measurement error is dramatically cut in terms of random and non-random error, from the high inter-rater levels to the much lower intra-rater level. This implies the potential for profound improvements in measurement practice, which will be irrespective of the number of examiners involved or location of the assessment. This must be considered to have enormous positive implications for both clinicians and researchers.

8.5.5. The Uillinn Method©: Implications for Research.

Research is certainly one area of health care science that could benefit from the technology used by the UM. For example, a recent study by Oginni et al (2004) assessed over 2000 Nigerian children in an attempt to determine pathological coronal knee angles as an indicator of rickets. This large study used either conventional photography or standard UG to determine the knee angles. This process may have been time consuming and expensive for such a large population. The current study has shown that there would also have been substantial heterogeneity between the levels of error for their different measurements techniques. The study would almost certainly have been made easier, quicker, cheaper and more accurate if a digital photography system such as the UM were used.

The reason that Oginni et al (2004) used such large numbers may have been to improve the power of their research. A statistical power analysis must consider levels of measurement error, thus when there is less error in the data the study is more powerful. Typically, a researcher will use a power analysis when designing a study to calculate the size of the cohort that is required to yield statistically significant results. The fact that the UM offers less error than the UG, implies that the researcher will require a smaller cohort in order to achieve the same statistical power (the specifics of this reduction will vary slightly depending on the demands of the study), and will clearly have positive implications for the feasibility of future research.

If a research project is designed to randomise untested or controversial treatments in different patient groups, it should be determined that the less numbers exposed to such experimental treatment the better. The implications of low measurement error in this context, will allow for lower patient numbers and therefore have positive implications for research ethics.

A demand for lower cohort numbers may also allow investigation in naturally small patient groups. For example it may be possible to statistically compare different treatments regimes for rare disorders, such as patients with bone tumours which are uncommon.

The financial restraints of a research project may sometimes lead to a restriction in he cohort number or time period of the experiments. The lower error displayed by the UM should also be of benefit to such considerations.

The fact that only one examiner is required to conducted the assessments remotely from the patient implies that multicentre studies are a much more reasonable possibility. Previously, such studies may have been difficult due to high levels of expected measurement error. However, the UM offers researchers a means to maintain the relatively low levels of intra-rater reliability error, irrespective of how many centres or examiners are involved in the project.

Another positive feature for researchers provided by the photographic based system is the ability to recheck data. This can be very useful when unusual data entries are seen. The researcher should easily be able to check any outlying data from the photographs and either accept or reject them with confidence.

Aside from the UM measurement system, some of the image processing techniques carried out for this study may provide valuable educational and communicative systems. The X-ray 'over-lay' pictures in particular are capable of giving an easily comprehensible view of the structure of joints (see Figure 6.5 to 6.14, page 296 to 300). In addition to education, this technological feature proved to be useful for quality assurance in the reliability trial and was used to judge unwarranted change between the repeat assessments (see Figure 5.10 to 5.13, page 250 to 252). Further exploration of these techniques and their uses may present other useful advantages to clinicians or researchers.

It may be said that the high levels of error associated with the main outcome measure used by the physiotherapy profession, i.e. the UG, has had a negative effect on the supporting evidence for different approaches to intervention. Controversy remains regarding a large number of commonly used techniques and modalities. The poor sensitivity of outcome measures, high associated error, along with underpowered trials have resulted in a number of false negative and false positive findings which has only added confusion to the professional knowledge base. The invention of the UM with scientifically established levels of error must be seen as a hugely beneficial development to the research community and the physiotherapy profession as a whole. It is hoped that the system should make research more

accessible to both clinicians and researchers and thus provide a means to improve the standard and effectiveness of health care provision.

8.6. Conclusion

This thesis provided new knowledge regarding the expected error associated with the universal goniometer (UG) and photographic measurement systems. A new approach to joint angle measurement using a virtual goniometer, the Uillinn Method© (UM), was designed, tested and pioneered as part of the work.

The calculation system for this new method was found to be accurate by better than 0.3°. When applied to the human model there was a substantial improvement in error compared to the UG.

Different approaches of applying the UM to both the elbow and knee were investigated, the best of which was the Mid-Point approach. This approach does not require adhesive landmarks to be attached to the subject's limb and there was limited evidence to support the use of such markers for photographic based measurement systems.

The UM was trialed in a clinical setting over a period of 3-months. The systems' usage over this time was found to be moderate at best. The main barrier to frequent use appeared to be the additional operational time. Nonetheless, the ability to record a definite image of a patient's movement restriction and monitor it over time was reported as a great advantage of the system. This feature allows direct comparisons between different sessions that can be used to negate the relatively high levels inter-rater reliability error.

The study supported a view by Lantz et al (1999) that the measurement of human joint angles are difficult and the no practical measurement method is likely to be without some error. Minimising this error is the challenge that routinely faces physiotherapists and researchers.

Outcome measures are vitally important to the development of an evidence base for physiotherapy practice. The development of precision measurement tools are an integral part of this evolving process. The effects of both random and non-random error for the UM have been described. It is hoped that this new system, which has been shown to have strong validity and reliability scores, will be available to clinicians in order for them to improve the accuracy of their measurements.

The data provided by this study will enable the clinician or researcher to have confidence when deciding when actual changes in available joint motion have occurred when the measurements are made by either the UG or the new UM.

Appendices

Appendix A

Unpaired T-test comparing Free-Hand Vs Tripod Mounted Photography

t-Test (Assuming Equal Variances) Unpaired Comparison for Means		
	Tripod	Free-Hand
Mean	0.0671	0.108
S.E.M.	0.0592	0.0488
S.D.	0.458	0.378
Variance	0.21	0.143
Sum	4.028	6.485
N	60	60
Sum(x^2)	12.69	9.139
Sum(x)^2/N	0.27	0.7
Correction Factor	0.921026408	
Df	118	
Expected Difference	0	
Common Variance	0.1765	
t(cal)	-0.533226323	N.S. (P>0.05) Two-sided
P(t<=t(cal)) Two-sided	0.594	
t(0.05) Two-sided	1.98	
Lower Conf. Limit of Difference	-0.110971722	
Upper Conf. Limit of Difference	0.192771722	

Appendix B.

Calculations for the Cosine Rule.

Calculations of the cosine rule.

$a^2 = b^2 + c^2 - 2 \times b \times c \times \cos(A)$.

a	=	74.98
b	=	75.02
c	=	75.02

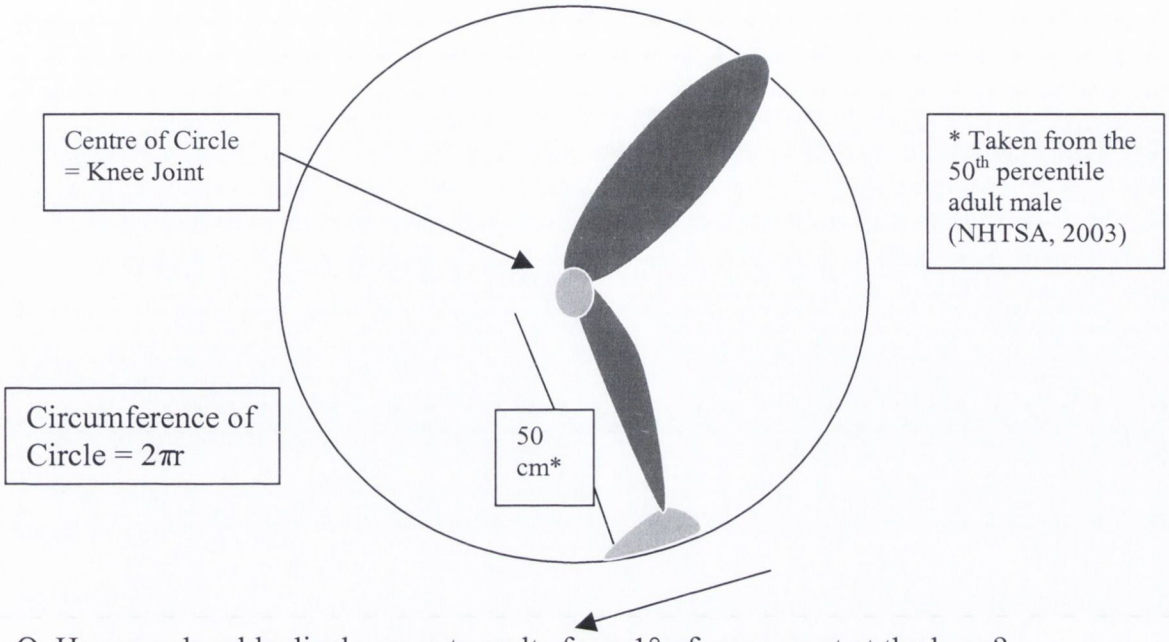
$(74.98)^2 = (75.02)^2 + (75.02)^2 - 2 \times 75.02 \times 75.02 \times \cos(a)$

5622.0004	=	5628.0004 + 5628.0004 – 11256.0008 x cos (a)
cos (A)	=	5622.0004 / (5628.0004 + 5628.0004 – 11256.0008)
A	=	59.96.
B	=	60.017
C	=	60.017

To work out the impact of the 0.02mm limitation on the CD calculations the angular difference between an equilateral triangle of exactly 75mm per side were compared to a triangle with one side 75.02mm, another side 75mm and the final side 74.98mm. Calculations of 1° arc of movement at Knee.

Appendix B (Cont).

Knee to Ankle length for the 50 percentile male (NHTSA, 2003) = 49.5cm



Q. How much ankle displacement results from 1° of movement at the knee?

⇨ Full Arc of 50cm Radius = 2 x π x 50

⇨ Circumference = 314.28cm

360° of circle	=	314.28cm circumference
1° of circle	=	314.28/360
	=	0.873 cm

A. 1° of movement at knee relates to 0.873cm movement at ankle.

Q. How much ankle displacement results from 2° of movement at the knee?

1° of circle	=	0.873 cm
2° of circle	=	0.873cm x 2
	=	1.746cm

A. 1° of movement at knee relates to 1.746cm movement at ankle.

Q. How much ankle displacement results from 5° of movement at the knee?

1° of circle	=	0.873 cm
5° of circle	=	0.873cm x 5
	=	4.365 cm

A. 1° of movement at knee relates to 4.365cm movement at ankle

Q. What is the angular difference results from 2cm of skin movement at the knee?

0.873cm	=	1°
2/0.873cm	=	x/1°
2cm	=	2.3°

A. 2cm of skin movement at the knee results in 2.3° change in angle.

Appendix C

Introduction to the Uillinn Method©.

In all areas of health care management a scientific approach by clinicians is essential in order to achieve optimal health outcomes. A scientific approach demands valid and reliable measurement tools that can deliver objective data upon which sound facts can be established. This implies that clinicians require objective measurement tools to assess human health, function and performance. Physical medicine is no exception in this need for dependable measurement tools.

One very important type of measurement for physical medicine is the calculation of joint range of motion (ROM), i.e. osteokinematic measurement. The *American Academy of Orthopaedic Surgeons* (AAOS) (1994) state that joint ROM is directly related to human function and therefore it is an essential component of a patient assessment.

Universal Goniometry (UG) is probably the most commonly used method of objectively assessing available joint ROM. However, research has shown that the validity and reliability of this device is limited. The Uillinn Method© (UM) was devised as an alternative approach to joint ROM measurement. It was hoped that the UM may improve upon the performance of the UG and possibly offer some useful advantages over it and other measurement methods.

The Uillinn Method©.

The UM uses personal computer (PC) software to create a 'virtual' on-screen goniometer. The system has been designed to report angular changes of as low as 0.01°. While, it has been shown that the entire UM procedure from photography to the angle calculation on the computer will remain within 0.5° of perfect accuracy for 95% of measurements (Dunlevy, 2002). However, it should be noted that certain photographic procedures must be observed if this level of accuracy is to be maintained.

Research into the potential accuracy, clinical validity and clinical reliability of the UM is on going. To date clinical validity and reliability has been established for measurements of sagittal motion at the knee. The findings of that research has established that the UM is limited in agreement with measurement made using X-ray

by 3.6° for 95% of measurements (95% LOA), with a high Pearson product moment correlation coefficient (PPMCC) (r=0.99). The investigation also established that the universal goniometer (UG) only achieved a 95% LOA of 7.6° (r=0.97) with X-ray data. These findings indicate that the validity of UM improves upon conventional UG by as much as 4°. Research into the validity for the UM for elbow measurements is not yet complete but initial results indicate outcomes similar to those seen at the knee.

The intra-rater reliability for the UM for assessment of the knee was shown to have a 95% LOA of 4.5° (r=0.99), while inter-rater reliability achieved a 95% LOA of 6.5° (ICC=0.99). It is difficult to compare these results to the performance of the UG, as reliability tests for the UG using the 95% limits of agreement (95% LOA) analysis have not been carried out. However, comparisons of correlation coefficients taken from previous studies imply that the UM offers an improvement in reliability (see Appendix 1).

The Photographic component of the Uillinn Method©.

During the actual taking of photographs for the UM, it is vital that certain procedures are observed. Following these 3 simple guidelines will ensure optimal results;

1) The camera lens should be centred on the approximate centre of the angle that is to be assessed. Specifically, the viewfinder crosshairs should be placed just below the approximate apex of the joint angle. This will help to avoid parallax error, the pillow effect or other causes of photographic misrepresentation.

2) An effort should be made to keep the incident angle of the lens as low as possible, to achieve optimal results this angle should be no more than 5°. To do this the photographer should stand in front of, and perpendicular to, the angle that is to be photographed. The photographer should check the alignment of the target angle and ensure that the camera is held approximately parallel to it. If the photographer is not certain of the orientation of the target angle with regard to their point of view, subsequent assessment of the image with the UM can not be trusted to be less than 1°.

3) The limb must be exposed adequately during photography.

The Computer based component of the Uillinn Method©.

Once the photographic component of the UM is complete, the image can be transferred to the PC using the Smart card reader (see Fig 1)

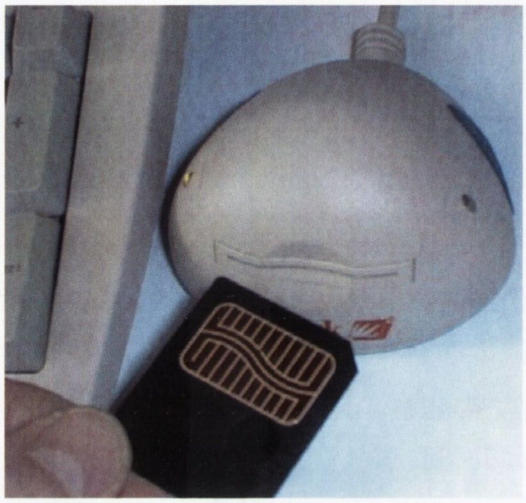


Fig 1. The Smart Media Disk and Smart Media Disk Drive.

Once the photographic file is transferred to the PC and placed in the relevant therapists folder. The file should then be renamed to include the patients initials and the date on which the photograph was taken. If more than one photograph of the same patient is taken on the same day, then additional roman numerals should be used to identify the file. The Uillinn software can be used to display the photographic image.

The image can be zoomed in upon and positioned so the relevant limb segment fills the screen. Once the picture is properly positioned in the screen, the measurement arms can be manipulated to allow calculation of the joint angle. Each measurement arm should be placed so it is approximately parallel to the longitudinal axis of the limb segment above and below the joint. For optimal results the measurement arm should be made longer than the limb segment (see fig 2).



Screen Shot of the Uillinn Method©.

Appendix D

How to use the Uillinn Method©. (Overview)

Introduction.

There are 3 basic steps that must be completed when using the Uillinn Method©, these are;

1) The Photographic Procedure.
2) Transferring and Storing the photographic file.
3) Using the Uillinn Software.

The Photographic Procedure of the Uillinn Method©.

During the photographic component of the Uillinn Method© (UM), it is vital that certain procedures are observed. Following these 3 simple guidelines will minimise measurement error originating from the photograph;

1) The limb must be exposed adequately during photography.
2) The camera lens should be centred on the approximate centre of the angle that is to be assessed. Specifically, the viewfinder crosshairs should be placed just inside the apex of the joint angle.
3) The camera should be held perpendicular to the target joint, 5° of rotation between the camera and the joint will not add any error to the measurement, however, if 10° of rotation is present during the photographic procedure an additional 1.5° of error will be present in the measurement.

   To control this adequately, the photographer should check the camera's alignment with regard to the target angle and ensure that the camera is held approximately perpendicular to it.

Transferring and Storing the Photograph to the Laptop.

When the photograph is taken the image must be transferred to the laptop and stored appropriately in each individual therapists folder. To complete this task;

1) Remove the Smart Media card from the camera.
2) Place the Smart Media card face down into the 'SanDisk' drive.
3) Double click on the "Shortcut to SanDisk ImageMate (E)" (see Fig 1.).
4) Then double click on the "IMOLYM" folder.

5) Right click on your photograph and choose the RENAME option (2nd from bottom).
6) Type the initials of your patient followed by the date. If you have taken more than one photograph of the same patient on that day indicate this with roman numerals. When this is done hit return.

   Example;

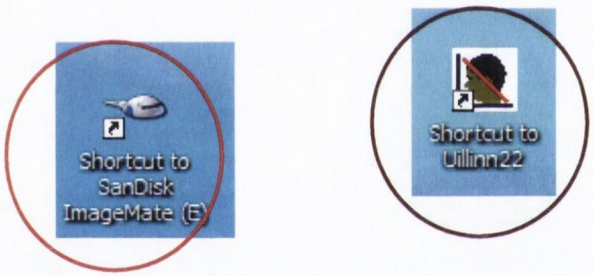
   John Murphy photographed on the *14 April 2003* = JM14Apr03

   2nd photograph of *John Murphy* same day = JM14Apr03ii
7) Drag this file up to YOUR folder on the computers Desktop.

Fig. 1. Shortcuts on the Desktop.

Using the Uillinn Software.

When the photograph has been renamed and transferred to the laptop the Uillinn software can be used to calculate the joint angle.



1. Double click on the Shortcut to Uillinn 2.3 (fig. 1).
2. Double click on Your folder, within the Uillinn window.
3. Double click on the appropriate file (preview available by single click).
4. Zoom in and Adjust image as appropriate (see Fig 2).
5. The RED measurement line will automatically be active. Use the RIGHT and LEFT mouse buttons to manipulate this line. LEFT for the Apex of the angle, RIGHT for the base of the angle (Circular end). The line should be parallel to the longitudinal axis of the limb segment (see Fig. 3).
6. Click on the "Adjust Blue Line" button at the bottom left hand corner of the window.
7. Now manipulate this line so that it is parallel to the longitudinal axis of the other limb segment.
8. The dotted Green line will reflect the base of the triangle that you've just drawn, the apex is the joint angle. Read the angle from the "Angle Deg" box.


Right
Click
Left
Click
Right
Click
Left Click

Appendix E

The five different approaches to the Uillinn Method©

The 3 Marker Approach (3Mks)

For this approach adhesive landmarkers should be used. At the Knee markers should be placed at; the prominence of the greater trochanter, the inferior tip of the lateral epicondyle, and the inferior tip of the lateral malleolus (see Figure 1). At the elbow the points are; the antro-lateral tip of the acromion process, the inferior tip of the lateral humeral epicondyle, and the distal aspect of the radial styloid (see Figure 2).

The photograph must be taken according to the establish guidelines and the photographs transferred to the computer (please see accompanying sheet). The VMAs can be aligned to the adhesive markers.

Three adhesive skin markers are placed over anatomical landmarks. This system allows clear reference points for the angular measurement from the photograph.

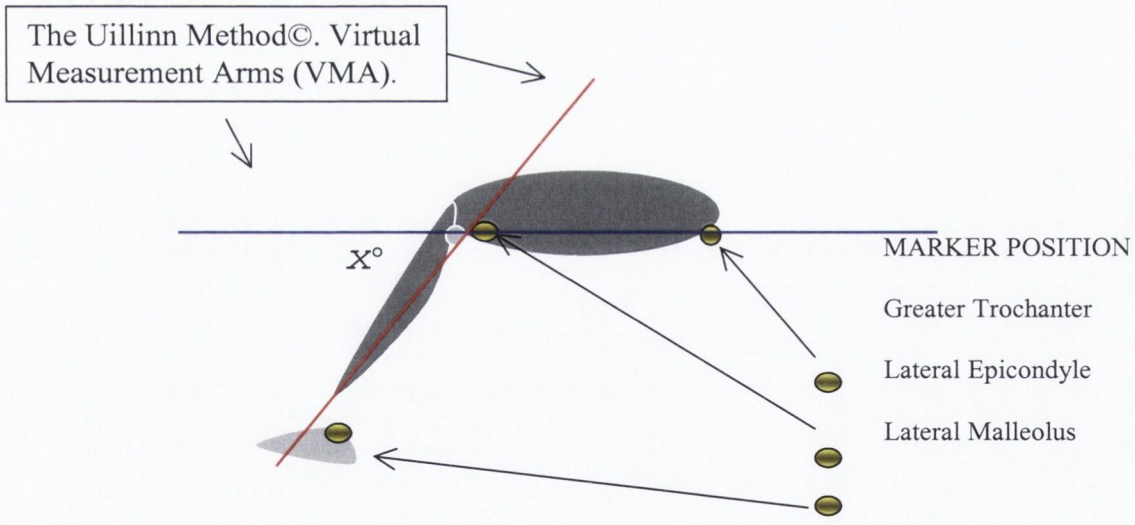


Figure 1. The '3 Marker Approach' to Applying the Uillinn Method©.

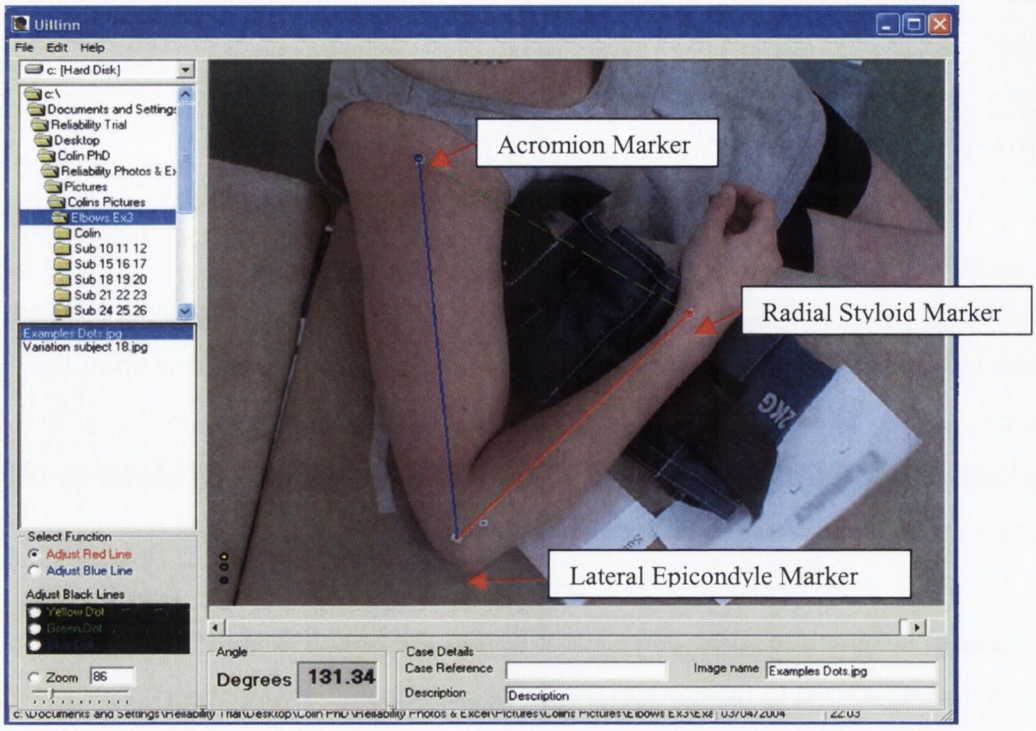


Figure 2. The '3 Marker Approach' at the Elbow.

The 4 Marker Approach (4Mks)

This approach also uses adhesive markers. The points are the same as listed above except there is an additional marker on each limb. At the Knee markers should be placed at; the prominence of the greater trochanter, the inferior tip of the lateral epicondyle, the superior tip of the head of the fibula and the inferior tip of the lateral malleolus (see Figure 3). At the Elbow the markers should be placed on; the antro-lateral tip of the acromion process, the inferior tip of the lateral humeral epicondyle, the prominence on the radial head and the distal aspect of the radial styloid (see Figure 4). The remaining procedures are as previously outlined.

Four adhesive skin markers are placed over anatomical landmarks. This system allows clear reference points for the angular measurement from the photograph.

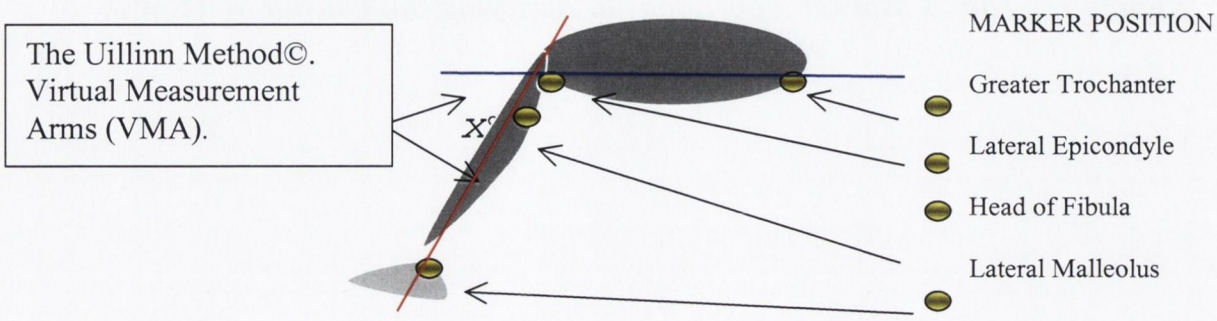


Figure 3. The '4 Marker Approach' to the Uillinn Method©.

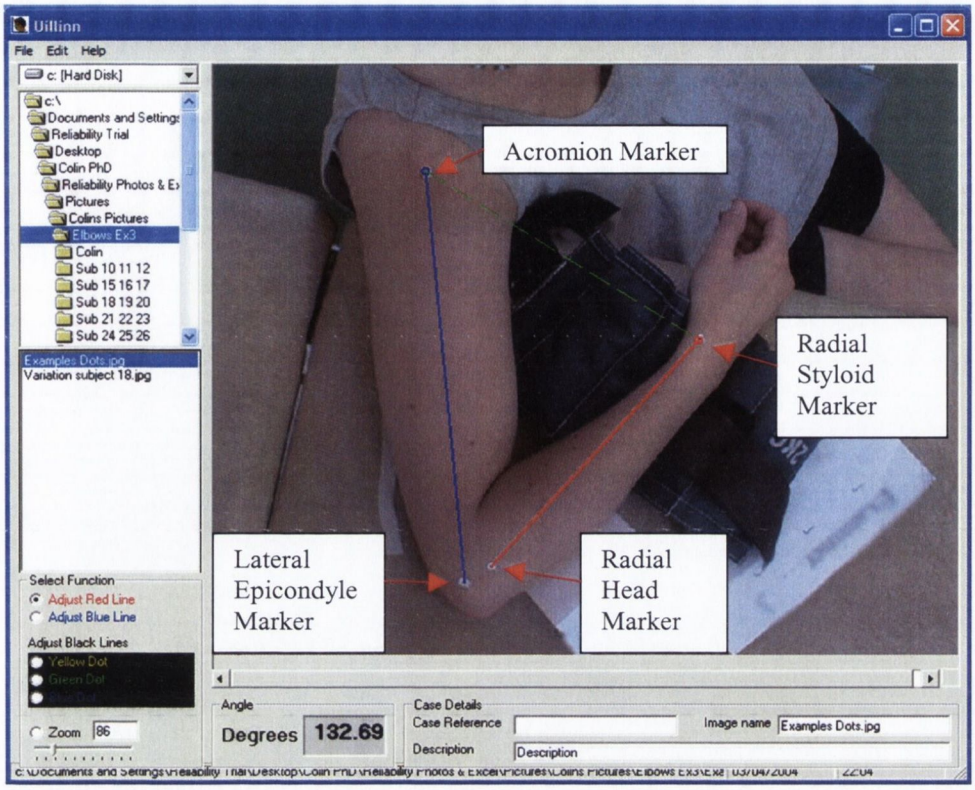


Figure 4. The '4 Marker Approach' at the Elbow.

The No Marker Approach (NMks)

This approach is identical to the 3-Marker approach with the exception of allowing the examiner to use their anatomical knowledge to locate the 3 reference points on the limb (see Figure 5, 6). This point location should be done only after the photograph is transferred to the PC.

The Measurement arms of the Uillinn Software are aligned to be approximately at the Greater Trochanter, the Lateral Epicondyle and the Lateral Malleolus.

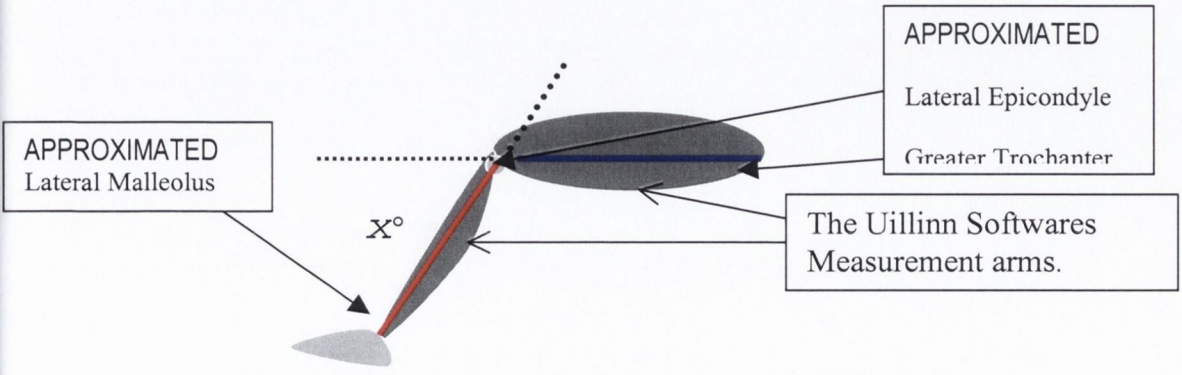


Figure 5. The 'No Marker Approach' to the Uillinn Method©.

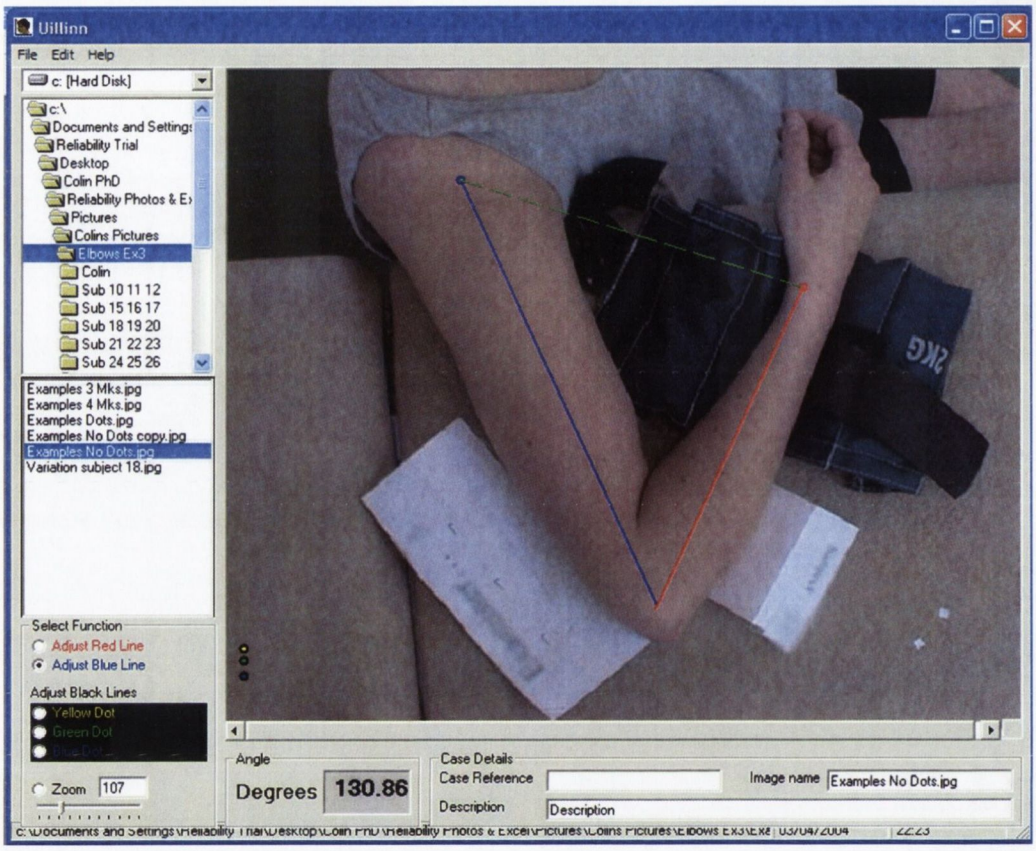


Figure 6. The 'No Marker Approach' at the Elbow.

The Parallel Approach (Parallel)

Photographic methods are standard. When using this approach the examiner should align the VMAs to the approximate longitudinal axis of the segments above and below the joint, using the orientation of the underlying bone as a guide. No specific reference points are recommended (see Figure 7)

The Measurement arms of the Uillinn Software are aligned to be parallel to the limb segments above and below the knee. No specific reference points are used. The anterior surface may be used as a guide.

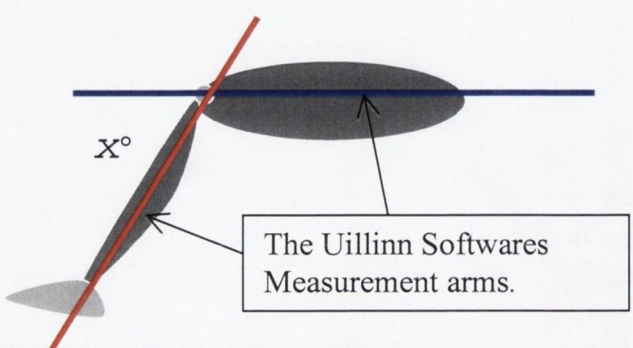


Figure 7. The 'Parallel' Approach to the Uillinn Method©.

The Midline Approach (Mid-Point)

This approach allows the examiner to use 3 assistive guide-lines each had a coloured 'Dot' associated with it. The guidelines are manipulated in the same way as the VMAs. When these lines are adjusted the dot will remain in the exact mid-point of that guide-line. Therefore the examiner can be precise about the location of the mid-point of the limb segment and use this to position the VMAs.

The procedure recommended for using this 'Mid-point' approach is; at the elbow to place one of the mid-point lines across the proximal upper arm roughly at the level of the insertion of the deltoid muscle, another mid-point line is placed across the wrist, and the last mid-point line is placed about the elbow joint between the most posterior aspect of the olecranon and the edge of the cubital fossa, the VMAs can then be aligned directly to the mid-point dots (see Figure 8).

At the knee the suggested method of application is that the first mid-point line should be placed on the proximal thigh segment just distal to the approximate level of the femoral triangle, the most distal mid-point line should be placed at the level of the lateral malleolus, the final line should be positioned about the knee joint itself between the most anterior prominence of the patella and the edge of the popliteal fossa, again the VMAs would then be aligned to the dots (see Figure 9).

When a subject is measured with their joints approaching end range flexion the outline of the joint fossa will be indistinguishable making it difficult to accurately place the mid-point guide at the joint. This problem may be more evident in subjects with excess adipose tissue. In these cases the line of the skin fold would be used instead of the edge of the joint fossa (see Figure 9).

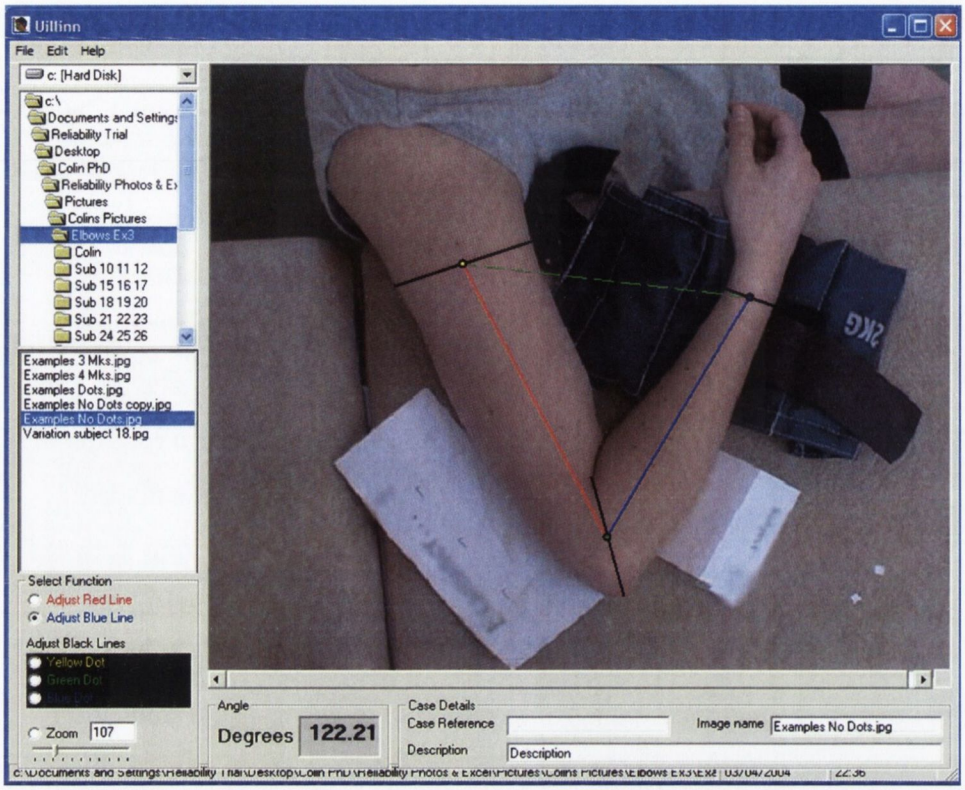


Figure 8. The 'Midline Approach' at the Elbow.

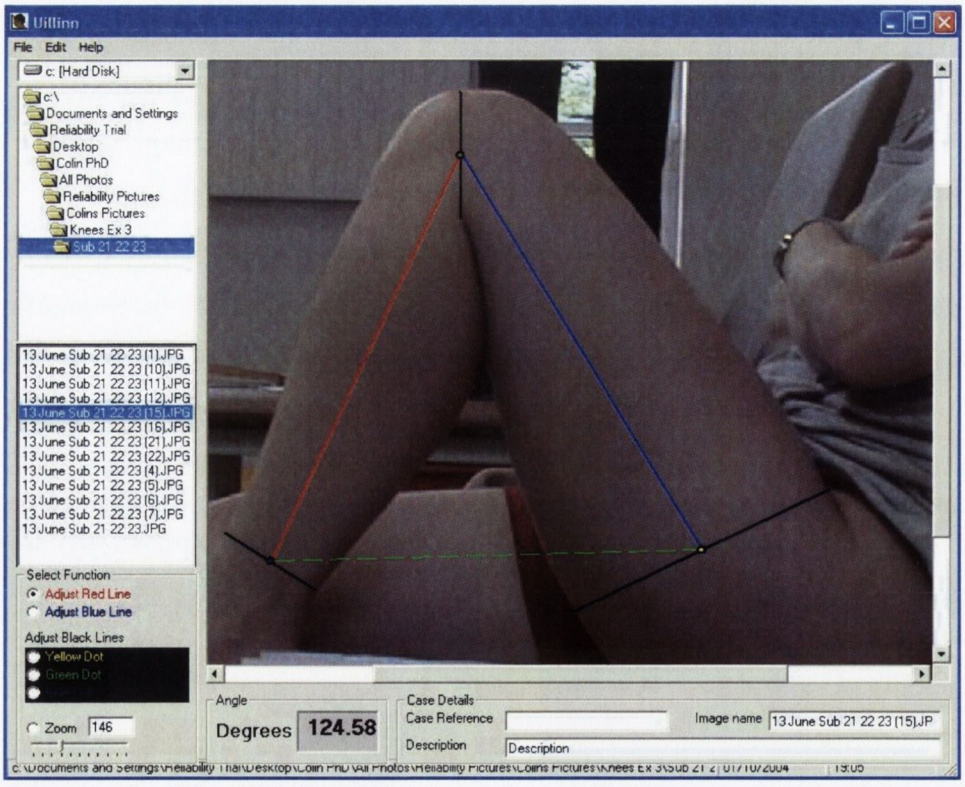


Figure 9. The 'Midline Approach' at the Knee.

Appendix E

Measurement Data For Examiner 1

	Examiner 1				UG Data	
Subject	1st Knee	2nd Knee			1st Elbow	2nd Elbow
1	-3	-2				
3	1	0				
4	9	8		4	76	75
5	11	9		5	45	44
6	13	14		6	128	123
7	18	15		7	76	72
8	22	25		8	31	35
9	26	25		9	140	138
10	28	29		10	115	116
11	29	28		11	88	86
12	30	34		12	49	45
13	30	29		13	62	63
14	30	32		14	20	23
15	33	36		15	35	39
16	34	34		16	75	74
17	39	40		17	128	127
18	54	55		18	122	115
19	61	65		19	84	82
20	62	61		20	43	45
21	73	75		21	120	117
22	81	85		22	75	78
23	83	83		23	44	43
24	84	82		24	136	134
25	85	84		25	77	75
26	91	92		26	28	28
27	91	92		27	112	110
28	95	91		28	26	27
29	95	96		29	72	78
30	96	96		30	129	130
31	99	100		31	80	77
32	100	97		32	25	26
33	100	99		33	120	121
34	102	104		34	22	23
35	105	106		35	108	110
36	108	105		36	69	70
37	121	117		37	37	39
38	125	122		38	132	132
39	126	121		39	68	67
40	131	133		40	25	24
41	132	132		41	90	93
42	134	130		42	24	22
43	134	135		43	113	112
44	134	135		44	122	123
45	135	131		45	77	77

46	135	134		46	27	23
47	135	135		47	110	110
48	136	135		48	80	83
49	136	133		49	44	43
50	145	144		50	120	122
51	150	149		51	49	52
52	151	150		52	80	78

No Markers		Examiner	1	
Subject	1st Knee	2nd Knee	1st Elbow	2nd Elbow
1	82.42	81.84		
3	146.32	147.27		
4	27.13	25.73	84.58	84.57
5	97.07	98.6	51.68	47.99
6	61.01	61.94	132.92	129.06
7	102.59	103.04	77.05	77.02
8	63.77	63.61	36.4	37.48
9	138.68	140.12	146.79	144.76
10	134.84	133.81	122.61	123.17
11	88.87	88.83	90.4	90.21
12	49.3	50.82	56.77	53.91
13	58.93	58.64	74.99	71.96
14	12.93	12.44	36.75	36.82
15	13.42	13.36	45.59	44.69
16	76.6	78.8	82.88	84.52
17	153.9	153.68	126.67	130.32
18	135.03	136.07	132.71	128.25
19	81.5	81.9	89.81	89.78
20	15.20	16.32	49.72	48.5
21	119.83	120.43	126.52	122.91
22	79.54	80.22	81.23	86.91
23	32.39	32.64	49.93	47.67
24	129.86	130.31	147.99	146.17
25	86.91	88.18	87.96	84.02
26	30.86	32.15	35.06	35.36
27	136.07	133.26	115.01	114.47
28	29.51	27.29	38.33	40.89
29	97.67	97.7	81.1	84.61
30	136.33	136.05	135.30	137.02
31	99.07	100.45	84.66	84.57
32	-3.39	-3.88	35.19	33.67
33	136.81	136.93	127.04	127.56
34	26.21	28.68	32.93	36.88
35	129.73	129.03	113.19	117.18
36	97.74	99.93	74.87	72.64
37	15.79	16.97	52	53.09
38	141.55	140.78	141.99	139.16
39	89.57	87.97	72.56	71.44
40	34.98	33.87	33	29.5

41	105.59	106.89	99.92	97.75
42	26.93	25.71	31.18	31.54
43	139.61	140.81	122.47	120.09
44	136.94	135.72	126.94	127.7
45	106.11	105.2	85.05	84.2
46	28.99	29.66	34.11	32.43
47	139.11	140.33	117.43	116.77
48	92.93	92.21	85.66	86.57
49	23.6	24.96	50.35	50.61
50	134.8	136.99	124.5	125.31
51	4.45	6.07	63.59	62.24
52	25.16	26.5	86.29	87.35

Parallel		Examiner	1	
Subject	1st Knee	2nd Knee	1st Elbow	2nd Elbow
1	82.04	81.93		
3	148.26	148.12		
4	26.31	25.44	81.74	82.42
5	97.34	97.35	44.01	42.07
6	61.9	61.85	127.13	124.33
7	103.34	103.81	72.32	69.19
8	64.58	64.8	33.44	32.35
9	140.47	141.92	143.68	143.64
10	136.83	133.54	119.89	118.25
11	88.43	87.17	89.09	90.68
12	50.13	51.3	49.62	46.19
13	58.71	58.63	66.36	66.66
14	11.57	11.82	33.96	34.57
15	13.44	13.37	41.84	42.07
16	77.22	77.6	82.78	81.92
17	153.88	154.07	126.86	127.5
18	136.98	140.79	121.59	122.27
19	81.28	80.76	84.18	85.09
20	15.58	15.79	44.65	45.4
21	120.32	123.7	122.42	122.54
22	78.71	78.32	77.59	77.4
23	31.56	31.93	46.54	46.52
24	132.74	130.91	145.53	146.35
25	86.99	88.78	83.18	77.42
26	30.87	31.77	33.16	33.8
27	137.65	136.11	115.74	113.57
28	28.76	28.7	37.16	37.52
29	98.53	98.46	76.07	78.91
30	138.27	137.43	138.04	139.01
31	99.67	99.54	85.38	83.48
32	-3.66	-3.68	35.56	33.55
33	137.89	136.85	128.9	128.97
34	26.89	27.71	26.89	28.01
35	129.76	129.78	112.48	117.07
36	100.19	99.71	73.6	73.62

37	16.4	15.54	43.82	46.1
38	142.44	141.47	139.48	141.6
39	88.43	89.7	68.52	66.58
40	34.57	34.58	28.88	28.27
41	105.82	106.18	99.98	95.08
42	26.31	25.31	28.26	28.1
43	140.46	142.04	122.74	121.39
44	136.56	135.34	127.17	124.3
45	104.8	103.2	84.59	82.37
46	26.61	29.46	28.42	29.7
47	140.86	141.87	114.55	115.85
48	92.21	93.63	84.35	85.3
49	24.59	23.72	48.16	50.76
50	136.81	136.97	119.35	126.94
51	6.28	5.98	64.55	62.82
52	26.21	25.03	83.16	82.71

Mid-Point		Examiner	1	
Subject	1st Knee	2nd Knee	1st Elbow	2nd Elbow
1	80.89	81.05		
3	147.57	146.72		
4	25.1	25.11	68.41	67.65
5	96.34	97.65	38.58	37.15
6	61.51	59.74	123.05	121.45
7	104.67	103.32	62.66	59.16
8	63.62	63.56	27.95	30.06
9	139.94	140.68	133.69	132.09
10	138.05	136.34	111.11	110.28
11	88.27	89.01	79.46	78.02
12	49.94	50.84	40.81	40.99
13	56.24	56.25	58.33	58.91
14	10.11	10.75	23.83	24.95
15	13.9	14.83	33.39	33.4
16	74.94	76.76	67.64	67.58
17	152.55	152.59	122.4	122.34
18	138.22	138.91	115.73	111.15
19	84.33	82.63	73.01	73.96
20	12.03	16.8	36.53	36.57
21	122.5	124.92	111.72	108.97
22	75.63	77.87	64.62	69.07
23	34.08	35.11	37.66	36.2
24	128.91	128.78	134.8	133.62
25	86.69	85.44	68.89	66.68
26	29.8	32.52	24.44	24.55
27	135.48	137.12	99.32	100.08
28	30.96	30.78	24.87	23.64
29	99.89	99.22	64.25	66.9
30	137.65	138.12	125.99	123.83
31	99.64	99.44	71.48	70.04
32	4.66	3.84	24.35	22.63

33	139.36	139.65	119.44	117.69
34	29.51	34	20.16	22.59
35	129.3	130.71	103.6	103.7
36	99.3	100.12	62.85	61.04
37	16.28	15.28	33.12	35.34
38	142.57	141.29	127.34	125.85
39	90.41	90.12	57.29	55.77
40	38.99	39.07	18.87	16.47
41	106.55	105.64	83.09	85.59
42	31.46	30.38	19.36	20.01
43	141.76	140.77	104.33	102.48
44	137.25	138.11	113.78	112.92
45	106.5	106.23	74.53	73.08
46	30.65	30.49	19.46	15.69
47	142.46	142.31	103.64	104.35
48	98.25	97.45	74.85	74.88
49	25.51	24.92	36.76	38.35
50	138.4	137.45	112.5	115.9
51	2.28	2.48	49.35	47.32
52	31.11	32.61	71.35	71

3 Markers		Examiner	1	
Subject	1st Knee	2nd Knee	1st Elbow	2nd Elbow
1	81.25	80.86		
3	144.29	143.97		
4	26.77	25.75	80.82	80.5
5	96.13	94.84	52.56	51.13
6	62.6	60.66	132.31	131.41
7	101.67	102.26	73.92	75.53
8	63.11	61.96	39.73	40.05
9	140.05	138.47	145.99	144.83
10	132.77	133.11	122.84	121.41
11	91.2	89.42	90.52	87.59
12	53.51	54.91	54.37	52.92
13	57.79	57.98	69.43	65.96
14	15.63	12.1	41.88	36.35
15	13.23	13.36	46.13	47.42
16	76.34	78.67	86.58	84.37
17	150.78	148.46	129.76	133.1
18	133.37	131.69	127.87	123.81
19	81.26	81.6	86.6	84.55
20	15.98	18.41	49.36	48.32
21	119.53	120.42	122.03	118.7
22	78.2	78.94	80.34	81.44
23	36.44	35.47	51.37	46.82
24	130.63	129.38	142.24	145.02
25	84.93	88.51	88.4	82.26
26	32.84	28.64	35.09	33.96
27	133.75	134.78	111.84	113.11
28	28.45	28.98	38.77	36.05

29	98.6	98.39	76.97	81.6
30	135.55	134.46	136.26	134.17
31	99.85	98.29	84.08	82.55
32	-0.879	-1.81	35.95	33.82
33	134.44	134.46	128.79	124.24
34	28.29	29.59	28.54	32.79
35	126.35	124.82	116.46	118.57
36	100.91	103.04	68.86	70.33
37	16.08	16.31	46.86	51.6
38	136.57	136.65	137.42	133.92
39	90.03	87.5	69.88	66.17
40	35.5	36.5	30.83	33.29
41	105.8	104.2	94.99	92.39
42	26.86	25.85	28.65	26.85
43	139.02	139.61	116.92	115.92
44	136.78	134.85	123.09	125.08
45	104.94	106.8	81.93	80.38
46	27.96	28.61	33.76	29.7
47	137.46	136.51	112.12	111.49
48	94.02	93.8	81.63	84.04
49	25.44	24.96	48.87	49.08
50	135.33	135.33	119.77	124.9
51	6.23	6.47	61.19	60.15
52	24.46	25.58	83.98	83.72

4 Markers		Examiner	1	
Subject	1st Knee	2nd Knee	1st Elbow	2nd Elbow
1	81.41	81.73		
3	147.22	147.02		
4	25.35	25.03	80.58	81.32
5	96.76	95.74	53.37	53.51
6	62.5	60.46	133.25	132.55
7	102.84	103.83	76.39	77.56
8	63.97	62.57	40.64	40.55
9	144.46	141.25	150.8	149.97
10	135.67	135.33	124.91	123.1
11	91.45	90.67	90.57	89.26
12	51.48	53.18	56.26	54.29
13	56.83	57.48	69.65	65.78
14	11.87	9.19	42.43	37.63
15	10.64	11.71	44.87	45.96
16	76.7	78.9	89.8	86.13
17	154.7	153.85	134.39	135.13
18	136.78	134.48	130.37	126.91
19	80.3	80.83	86.07	86.46
20	15.54	16	51.05	47.2
21	122.43	121.57	124.81	121.05
22	78.77	80.63	83.89	84.89
23	34.9	34.69	51.83	49.72
24	131.18	130.29	147.47	150.02

25	86.36	88.22	87.83	83.58
26	29.25	25.33	36.93	31.31
27	135.09	136.08	112.64	112.9
28	26.61	26.86	40.04	35.41
29	98.59	98.6	77.19	80.41
30	137.28	136.9	139.81	138.42
31	99.89	98.29	83.86	85.58
32	-2.95	-4.5	34.35	32.61
33	136.8	135.85	129.91	127.64
34	27.97	28.55	27.35	33.24
35	126.71	125.78	117.37	117.67
36	99.77	101.3	71.65	71.52
37	14.74	15.54	46.91	52.18
38	137.1	139.8	140.35	136.25
39	89.68	88.33	69.87	69.13
40	33.73	34.3	31.27	33.42
41	108.44	106.29	97.39	93.71
42	24.4	22.21	28.95	27.25
43	140.03	141.99	118.96	116.94
44	140.48	138.14	123.7	126.69
45	105.43	103.5	82.93	82.11
46	26.1	27.26	34.93	30.69
47	141.26	139.29	113.28	112.78
48	94.31	94.08	84.92	86.89
49	22.54	22.38	50	51.11
50	136.35	137.03	120.64	125.83
51	4.38	4.12	62.3	60.75
52	21.46	22.95	87.23	86

Appendix E (cont.)

Measurement Data For Examiner 2

UG Data		Examiner	2	
Subject	1st Knee	2nd Knee	1st Elbow	2nd Elbow
1	80	83		
3	140	142		
4	35	33	83	79
5	100	115	54	45
6	73	69	128	126
7	102	102	77	71
8	70	67	38	37
9	141	140	142	141
10	134	135	120	120
11	88	89	91	91
12	59	59	50	52
13	67	66	70	69
14	20	18	36	37
15	13	18	30	31
16	85	84	75	80
17	150	150	127	130
18	132	135	115	118
19	93	90	84	89
20	15	18	42	47
21	121	119	119	120
22	82	85	82	80
23	41	43	44	47
24	131	130	143	141
25	90	84	90	80
26	30	35	33	28
27	134	136	112	113
28	33	30	38	35
29	100	100	81	80
30	136	135	135	133
31	82	86	101	101
32	4	4	31	30
33	135	135	127	125
34	28	33	23	29
35	124	127	107	110
36	78	74	71	69
37	11	13	36	40
38	136	138	135	133
39	93	95	68	65
40	43	39	26	24
41	108	107	93	97
42	27	30	25	23
43	138	138	116	114
44	137	138	124	122

45	110	113	94	100
46	38	34	27	26
47	139	139	113	114
48	96	94	82	82
49	26	30	42	45
50	137	134	57	55
51	2	1	56	53
52	27	30	80	86

No Markers		Examiner	2	
Subject	1st Knee	2nd Knee	1st Elbow	2nd Elbow
1	87.61	86.48		
3	145.26	145.42		
4	33.98	33.71	84.44	85.56
5	100.53	97.65	49.79	51.47
6	66.12	59.36	131.16	133.26
7	108.43	106.87	78.88	81.55
8	64.86	65.29	44.89	38.99
9	143.9	144.9	148.14	146.36
10	137.01	138.36	122.02	125.46
11	88.75	90.63	91.78	93.5
12	52.8	50	54.89	52.54
13	62.2	59.98	76.49	78.22
14	8.66	8.4	39.86	36.38
15	17.2	13.98	43.91	47.95
16	83.77	83.28	83.97	86.01
17	152.8	153.63	132.82	133.84
18	141.87	141.34	129.7	126.18
19	85.2	83.09	89.61	93.16
20	15.7	13.27	46.59	46.33
21	123.61	122.36	126.49	127.29
22	83.65	81.99	83.18	85.46
23	36.64	33.97	48.17	52.23
24	128.94	130.95	142.73	145.57
25	90.26	91.12	90.62	86.13
26	30.29	29.11	38.14	37.86
27	136.63	140.66	114.58	116.23
28	27.21	27.26	35.22	40.37
29	102.47	101.49	84.42	88.08
30	143.89	141.51	136.28	137.73
31	100.71	101.48	84.53	88
32	-1.59	-0.12	35.35	37.66
33	140.43	139.62	129.53	127.64
34	29.55	28.42	35.34	40.32
35	127.51	130.22	114.09	118.14
36	100.53	99.96	78.99	72.87
37	14.74	17.35	45.78	43.13
38	141.6	142.3	141.83	142.34
39	90.32	89.7	70.52	69.77

40	36.98	39.36	32.11	31.97
41	109.59	111.59	99.08	101.63
42	30.45	27.72	34.11	28.94
43	142.03	145.51	124.3	120.48
44	141.28	141.83	126.49	123.26
45	110.15	108.6	83.61	85.9
46	33.66	32.09	35.13	29.51
47	147.51	148.33	118.63	123.84
48	97.47	97.63	87.62	90.53
49	27.71	29	51.24	51.1
50	143.36	141.87	131.53	136.37
51	10.91	12.32	67.25	65.69
52	30.96	32.48	88.74	87.23

Parallel		Examiner	2	
Subject	1st Knee	2nd Knee	1st Elbow	2nd Elbow
1	80.45	81.82		
3	145.08	147.57		
4	27.81	28.91	80.98	76.09
5	96.75	96.51	42.34	40.94
6	60.17	61.71	128.57	128.78
7	105.30	102.49	75.03	75.36
8	65.06	65.38	38.14	34.77
9	145.49	145.11	146.06	142.79
10	136.45	135.22	113.99	116.87
11	86.96	88.45	89.8	92.33
12	48.81	51.41	48.93	49.7
13	57.62	58.75	70.78	68.07
14	7.37	6.89	33.05	32.75
15	11.44	11.55	42.05	41.22
16	78.4	74.77	79.4	82.83
17	155.26	153.63	128.77	130.35
18	134.55	131.69	120.43	119.47
19	79.76	79.64	86.53	87.33
20	12.56	9.1	44	44.79
21	124.14	123.23	121.39	120.44
22	74.9	76.41	79.34	80.49
23	32.49	30.43	44.79	46.97
24	132.2	133.68	142.53	141.01
25	84.74	85.87	84.31	78.87
26	27.13	28.87	32.86	30.5
27	136.83	138.84	112.85	112.97
28	22.94	21.58	33.09	31.72
29	96.77	96.53	81.15	81.25
30	142.8	142.38	132.03	132.17
31	99.01	97.63	83.08	80.77
32	-4.39	-5.17	31.09	33.89
33	139.5	140.23	123.51	126.99
34	25.01	24.81	25.89	32.97
35	128.99	126.35	112.32	114.06

36	104.36	101.84	71.73	69.88
37	10.81	11.24	35.61	39.16
38	139.59	141.28	137.11	138.78
39	87.8	88.3	67.61	67.67
40	33.78	33.92	25.30	24.38
41	106.99	107.56	94.44	97.66
42	21.6	21.75	27.66	25.04
43	144.8	143.95	122.08	122.5
44	141.52	141.12	122.56	121.89
45	109.15	108.68	82.39	82.23
46	29.51	30.13	27.67	25.92
47	144.78	143.63	110.74	115.89
48	97.37	95.62	83.42	81.68
49	24.86	24.45	40.81	44.16
50	140.37	139.01	126.41	131.4
51	1.75	-1.03	55.87	55.85
52	25.94	25.31	79.28	78.87

Mid-Point		Examiner	2	
Subject	1st Knee	2nd Knee	1st Elbow	2nd Elbow
1	83.38	81.11		
3	141.45	145.61		
4	28.55	30.39	70.9	69.79
5	96.06	96.54	38.05	38.08
6	63.03	65.02	122.27	121.66
7	98.87	100.31	63.84	62.02
8	60.07	63.28	35.24	30.64
9	143.35	143.23	132.24	131.4
10	136.44	135.22	110.29	112.76
11	86.7	89.15	80.45	82.15
12	55.72	49.09	41.52	41.48
13	55.73	57.12	59.98	59.33
14	12.46	16.88	25.14	25.68
15	17.07	16.61	32.16	33.52
16	77.47	78.18	72.47	70.91
17	151.19	151.73	121.07	122.61
18	139.51	138.92	113.7	112.84
19	83.82	83.29	80.36	79.3
20	17.63	17.92	38.41	36.09
21	122.71	123.59	111.45	113.18
22	79.29	79.11	66.14	67.64
23	39.57	37.15	35.12	38.64
24	128.27	128.27	132.31	131.68
25	85.72	87.89	72.52	67.42
26	31.72	30.88	23.5	23.54
27	135.37	135.66	98.56	99.22
28	32.03	32.95	22.37	22.65
29	99.94	99.01	66.92	67.88
30	124.57	124.12	124.22	126.11
31	98.98	98.45	70.54	69.37

32	2.08	3.71	23.02	23.19
33	137.64	136.3	118.26	119.26
34	32.64	32.68	16.16	24.24
35	129.78	128.09	105.02	109.01
36	96.16	94.44	62.37	60.32
37	17.1	16.73	31	31.3
38	138.67	139.66	129.31	129.38
39	88.43	88.97	58.95	56.22
40	38.54	38.6	19.44	19.36
41	105.88	104.38	85.87	84.24
42	31.31	34	19.69	20.71
43	139.39	139.34	108.28	105.54
44	138.55	136.9	113.18	116.71
45	107.6	107.66	71.46	71.55
46	31.96	31.95	17.25	21.81
47	141.38	141.43	105.4	104.35
48	98.87	98.66	73.34	72.83
49	25.52	26.19	35.23	37.56
50	134.95	133.07	119.08	117.3
51	2.47	2.41	47.43	46.05
52	29.31	28.73	71.14	68.39

3 Markers		Examiner	2	
Subject	1st Knee	2nd Knee	1st Elbow	2nd Elbow
1	84.5	82.78		
3	145.73	144.88		
4	27.01	26.08	82.49	80.18
5	97.83	97.78	45.59	46.93
6	62.09	62.97	128.3	127.84
7	103.17	102.5	75.91	77.58
8	65.56	64.56	42.09	37.53
9	142.17	140.22	145.82	142.69
10	133.82	135.68	118.73	121.3
11	88.78	89.54	90.9	91.28
12	51.72	52.27	52.96	54.18
13	61.47	57.9	68.09	71.33
14	14.79	15.61	37.15	38.2
15	14.33	14.09	44.72	42.66
16	80.57	82.71	82.61	81.92
17	151.5	151.17	125.51	129
18	136.73	134.84	126.85	121.58
19	81.8	80.57	83.48	88.13
20	16.32	13.04	52.4	46.49
21	119.84	121.1	118.92	118.69
22	77.82	80.9	77.94	82.24
23	36.76	34.26	45.98	48
24	131.06	130.09	139.25	140.06
25	87.02	89.1	79.59	74.29
26	28.76	29.7	33.18	34.24
27	135.29	136.16	112.02	111.01

28	31.18	33.2	35.44	34.67
29	101.11	98.52	84.07	82.19
30	137.42	135.17	130.56	129.55
31	99.72	97.65	81.07	76.85
32	-3.5	-2.32	34.77	36.07
33	138.58	138.24	129.57	126.78
34	27.42	25.98	29.11	38.09
35	127.1	128.34	115.17	118.13
36	102.54	103.15	78.09	76.15
37	14.59	13.97	42.33	41
38	140.31	140.76	132.96	133.51
39	90.89	92.37	68.73	68.32
40	36.34	36.45	31.82	33.43
41	103.8	105.34	94.66	95.5
42	25.01	25.51	31.44	28.19
43	140.09	140.04	115.2	113.86
44	137.64	141.12	124.9	124.72
45	109.81	106.01	82.93	81.23
46	31.47	32.5	34.75	31.24
47	140.62	142.79	115.07	116.22
48	93.23	95.57	85.36	84.37
49	25.85	25.69	44.68	47.29
50	137.64	137	120.86	125.03
51	5.13	4.88	60.2	60.06
52	27.65	28.28	83.98	83.71

4 Markers		Examiner	2	
Subject	1st Knee	2nd Knee	1st Elbow	2nd Elbow
1	83.29	82.09		
3	148.67	143.91		
4	25.01	25.38	84.43	82.53
5	99.29	99.44	47.95	48.81
6	61.17	62.11	131.93	130.85
7	100.89	99.96	81.04	80.73
8	64.23	64.37	43.05	37.88
9	143.75	142.05	151.16	145.69
10	135.13	136.95	121.96	124.02
11	88.37	89.73	93.59	92.9
12	48.96	50.54	54.03	56.15
13	60.01	55.45	69.87	73.72
14	11.46	12.08	38.89	39.05
15	15.06	14.81	46.5	44.18
16	80.88	77.68	84.48	83.24
17	152.27	152.42	127	133.2
18	138.14	137.55	129.85	125.1
19	81.99	79.9	83.99	91.74
20	12.79	10.72	53.08	48.72
21	121.23	123.37	121.41	121.23
22	77.55	81.21	81.32	85.37
23	34.08	32.55	46.08	50.64

24	129.86	130.59	142.84	143.92
25	86.36	89.73	82.74	76.02
26	30.35	31.54	34.32	35.59
27	135.86	137.72	113.99	112.23
28	28.54	32.48	36.73	35.67
29	103.33	98.38	82.8	85.39
30	139.65	137.02	134.72	134.81
31	100.78	97.49	82.37	80.33
32	-7.03	-4.87	36.65	37.45
33	137.53	138.67	130.11	133.24
34	29.75	28.34	30.61	40.16
35	127.01	126.59	118.93	120.42
36	102.32	103.61	81.92	78.68
37	11.12	11.04	43.16	42.3
38	142.11	142.08	137.36	138.03
39	91.78	92.23	71.33	70.43
40	33.63	32.86	32.6	33.3
41	104.08	106.33	97.03	98.9
42	22.26	21.85	32.48	28.76
43	142.67	143.43	120.02	115.7
44	140.56	143.64	127.03	126.89
45	109.83	105.43	83.97	83.87
46	29.56	29.63	36.17	32.24
47	143.56	144.15	117.54	116.77
48	93.31	95.88	87.62	85.75
49	23.22	23.37	45.85	49.62
50	138	136.53	127.9	122.46
51	2.37	2.06	61.33	60.69
52	24.99	26.16	84.24	83.77

Appendix E (cont.)

Measurement Data For Examiner 3

UG Data		Examiner	3	
Subject	1st Knee	2nd Knee	1st Elbow	2nd Elbow
1	83	84		
3	145	141		
4	27	31	75	69
5	95	99	44	39
6	66	64	125	122
7	106	103	69	67
8	69	67	33	33
9	139	142	135	136
10	127	130	111	116
11	89	91	83	89
12	58	57	43	45
13	61	62	67	62
14	7	15	28	24
15	10	12	35	36
16	82	80	83	69
17	149	146	124	124
18	131	130	121	113
19	82	84	79	80
20	14	14	39	35
21	116	120	115	113
22	78	83	80	78
23	38	40	34	39
24	124	122	135	137
25	82	82	71	72
26	28	28	25	20
27	134	135	103	106
28	34	24	26	18
29	95	93	65	73
30	136	133	125	128
31	98	99	78	85
32	-3	-1	32	27
33	134	135	121	117
34	33	30	17	24
35	122	122	106	107
36	98	100	64	62
37	11	15	37	40
38	138	136	130	129
39	87	92	67	74
40	39	36	20	17
41	104	104	92	91
42	28	29	19	22
43	134	134	110	112
44	135	133	112	114

45	110	100	79	78
46	34	34	18	20
47	140	140	104	105
48	95	95	82	85
49	29	29	36	36
50	140	136	123	124
51	4	7	55	51
52	32	28	83	81

No Markers		Examiner	3	
Subject	1st Knee	2nd Knee	1st Elbow	2nd Elbow
1	82.16	80.99		
3	143.58	144.11		
4	29.9	26.12	82.5	84.7
5	98.54	97.69	46.37	43.69
6	61.46	65.59	125.92	129.25
7	105.38	104.09	75.4	73.69
8	66.13	65	34.28	37.54
9	145.63	143.6	145.84	149.94
10	137.85	137.85	124.28	125.15
11	87.73	84.91	91.29	88.21
12	51.83	48.69	51.64	46.97
13	55.76	57.16	70.6	72.43
14	8.5	7.13	34.43	36.67
15	13.18	13.66	40.11	46.75
16	78.28	78.27	80.54	76.69
17	154.11	156.49	133.13	132.26
18	136.39	130.95	124.61	129.5
19	81.7	76.8	90.61	89.65
20	14.09	17.83	46.6	45.57
21	117.92	119.01	120	120.19
22	81.37	80.33	79.24	83.47
23	34.03	31.43	43.81	46.96
24	127.31	130.62	146.74	146.99
25	92.1	86.97	81.49	79.98
26	28.09	28.57	35.28	34.18
27	133.99	136.89	115.29	112.95
28	25.31	27.99	36.56	36.07
29	98.39	100.47	68.78	74.36
30	140.35	138.89	136.58	136.15
31	101.85	99.45	82.49	83.68
32	-2.65	-3.33	30.13	30.97
33	143.24	139.08	126.17	127.8
34	29.58	29.05	32.86	30.82
35	128.62	128.9	111.6	112.09
36	99.56	98.89	72.67	74.3
37	17.85	14.52	44.81	45.45
38	139.3	139.53	135.57	134.73
39	87.93	87.43	74.29	72.6
40	35.66	33.55	29.56	33.14

41	106.78	108.03	96.91	93.75
42	24.75	27.89	26.77	25.68
43	140.62	142.82	116.89	113.63
44	137.54	140.83	125.94	123.47
45	106.98	109.33	82.36	81.81
46	32.62	34.24	32.03	31.88
47	141.93	143.76	114.63	113.48
48	91.62	92.98	84.94	86.68
49	24.66	25.67	47.44	48.34
50	140.06	135.58	121.39	126.05
51	5.89	3.5	60.51	57.87
52	26.06	25.33	80.97	84.09

Parallel		Examiner	3	
Subject	1st Knee	2nd Knee	1st Elbow	2nd Elbow
1	80.99	83.43		
3	144.9	147.99		
4	25.20	25.38	84.1	81.4
5	97.1	94.57	40.52	43.5
6	62.96	60.9	127.39	125.2
7	105.39	104.48	70.85	72.65
8	67.34	68.13	37.78	38.28
9	143.43	145	142.24	141.84
10	136.41	134.9	115.93	115.89
11	86.54	86.35	89.32	86.98
12	47.57	47.09	51.85	47.5
13	59.06	57.13	68.5	69.13
14	10.3	13.64	29.65	31.28
15	16.89	17.05	37.01	37.07
16	76.76	74.68	79.57	77.98
17	151.87	152.73	130.02	129.34
18	135.24	133.67	117.33	109.95
19	82.34	80.62	80.72	84.8
20	18.65	23.93	40.82	41.75
21	117.08	117.38	121.88	121.92
22	75.74	78.58	82.02	79.56
23	37.82	39.45	40.73	43.12
24	131.57	132.67	142.52	143.62
25	91.87	91.36	75.36	72.51
26	29.66	28.38	35.49	34.84
27	136.65	135.05	108.1	106.66
28	29.46	29.41	28.46	30.55
29	98.49	98.6	74.09	74.41
30	131.79	130.95	142.65	145.20
31	101	99.73	79	74.99
32	-5.28	-5.36	30.22	25.86
33	142.21	142.6	122.82	122.24
34	31.04	32.14	29.24	27.43
35	125.5	123.39	109.33	107.55
36	102.75	101.68	64.18	68.49

37	19.81	21.85	37.29	37.76
38	138.77	138.36	138.67	139.04
39	87	86.49	67.62	65.23
40	38.4	38.39	27.82	25.19
41	108.32	105.93	92.12	90.66
42	20.53	21.99	28.26	28.48
43	141.71	141.29	112.34	115.14
44	142.95	142.47	117.71	115.94
45	108.62	107.51	76.39	80.58
46	23.9	23.6	35.97	36.44
47	143.04	143.92	106.46	104.83
48	94.46	94.76	76.57	77.66
49	30.28	31.34	41.54	42.68
50	140.49	143.15	112.45	120.53
51	5.22	6.64	52.76	52.28
52	29.63	27.93	78.02	78.24

Mid-Point		Examiner	3	
Subject	1st Knee	2nd Knee	1st Elbow	2nd Elbow
1	79.82	80.1		
3	147.45	147.451		
4	22.73	22.34	71.88	73.01
5	97.9	98.74	38.28	39.27
6	61.41	63.78	121.02	122.21
7	105.04	104.49	63.69	66.51
8	66.44	66.96	31.26	32.61
9	144.04	145.06	134.18	132.21
10	136.93	137.18	110.51	111.42
11	87.53	87.12	80.47	80.02
12	54.34	53.58	40.21	40.31
13	63.35	60.64	59.8	58.22
14	10.97	11.81	23.91	26.27
15	15.87	16.08	29.72	31.6
16	80.48	79.23	67.13	65.87
17	151.32	151.47	125.47	122.31
18	138.57	139.55	116.46	105.34
19	83.45	82.09	77.37	75.42
20	15.61	16.77	36.47	37.81
21	124.93	123.52	112.74	111.38
22	80.85	80.56	66.68	71.07
23	35.51	37.37	35.2	38.12
24	129.69	128.12	134.95	134.22
25	87.4	87.23	66.63	67.88
26	32.24	31.22	23.32	22.58
27	136.82	136.92	100.38	99.61
28	28.97	30.54	22.87	24.37
29	100.02	99.97	66.59	61.63
30	137.91	138.11	127.85	127.07
31	100.93	102.39	70.23	71.66
32	0.38	1.1	23.05	21.16

33	139.63	139.13	116.9	119.81
34	34.17	32.71	20.01	11.74
35	127.41	129.88	105.55	105.01
36	99.9	100.03	61.11	59.26
37	16.82	18.41	30.61	32.07
38	140.75	141.23	127.14	125.06
39	91.22	91.02	61.86	58.4
40	37.65	37.43	21.93	22.08
41	106.53	108.06	82.36	83.18
42	30.1	30.52	17.19	18.46
43	142.37	141.47	104.33	104.72
44	138.94	138.22	113.66	115.45
45	106.69	107.15	71.18	71.77
46	35.12	33.88	20.04	19.49
47	142.63	142.34	100.51	104.2
48	96.65	97.8	72.52	74.87
49	26.59	25.9	37.33	38.02
50	138.77	137.97	107.03	108.89
51	3.08	3.28	48.64	46.73
52	26.46	26.04	72.52	73.45

3 Markers		Examiner	3	
Subject	1st Knee	2nd Knee	1st Elbow	2nd Elbow
1	81.96	81.69		
3	134.03	145.3		
4	28.51	27.75	82.19	79.38
5	98.92	96.68	43.26	47.51
6	62.3	64.14	127.1	130.88
7	104.29	104.38	75.49	74.68
8	68.38	68.21	39.16	36.98
9	144.86	142.48	139.55	139.63
10	133.88	136.02	123.91	121.9
11	91.53	89.48	90.89	89.63
12	53.13	53.32	51	49.13
13	59.72	62.54	69.58	66.05
14	11.99	13.7	38.61	38.25
15	13.34	13.09	43.57	42.37
16	81.17	80.64	82.13	81.05
17	148.27	147.36	129.56	131.39
18	136.06	134.74	124.98	118.3
19	80.13	83.21	92.97	84.57
20	21.74	14.39	48.42	47.96
21	121.68	122.2	121.79	121.03
22	80.9	82.41	81.95	84.85
23	37.32	35.48	46.07	47.54
24	131.79	132.51	143.07	144.31
25	88.79	89.5	78	77.08
26	29.74	30.07	32.77	34.23
27	136.03	134.7	111.21	112.94
28	32.86	29.42	36.9	40.92

29	100.74	98.81	79	83.67
30	138.68	137.15	135.03	136.14
31	102.77	101.25	82.66	84.53
32	-0.94	-5.7	31.64	32.84
33	136.98	137.65	123.25	121.56
34	27.76	27.68	29.18	36.49
35	125.31	127.85	112.82	118.11
36	103.03	99.13	74	75.64
37	16.21	15.78	46.32	46.62
38	141.89	140	136.86	134.46
39	91.95	94.11	75.28	71.67
40	38.38	37.39	32	32.59
41	105.82	106.84	96.64	95.03
42	25.38	25.76	28.91	26.82
43	139	138.66	115.4	112.66
44	141.7	141.37	127.15	125.36
45	109.44	108.19	82	81.24
46	32	31.35	31.57	33
47	139.66	141.59	116.9	114.62
48	95.55	96.42	82.48	84.72
49	27.79	26.68	45.14	50.76
50	136.87	135.5	121.63	123.58
51	6.33	6.93	59.35	55.12
52	25.51	25.55	85.17	83.41

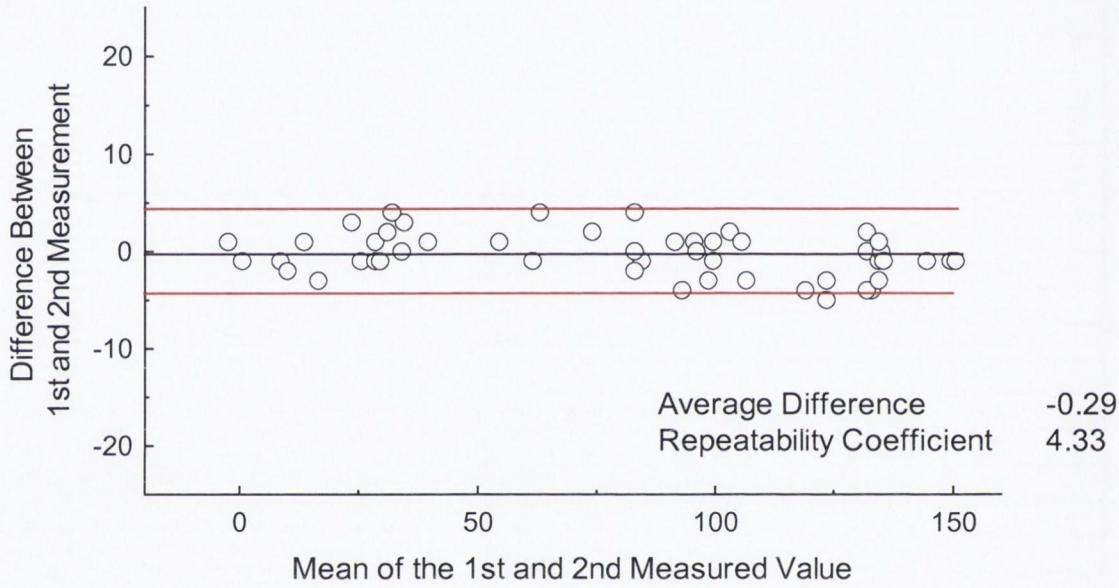
4 Markers		Examiner	3	
Subject	1st Knee	2nd Knee	1st Elbow	2nd Elbow
1	81.12	81.28		
3	136.65	144.79		
4	29.78	26.73	82.89	81.36
5	100.8	96.98	41.97	47.76
6	61.21	62.62	127.81	130.09
7	105.23	106.11	76.9	75.86
8	69.05	67.37	37.51	37.06
9	144.58	142.91	143.56	142.67
10	135.62	137.01	123.98	122.37
11	91.6	88.59	91.67	89.61
12	50.31	52.14	52.17	49.57
13	58.58	60.92	70.94	65.45
14	9.14	10.77	36.21	38.54
15	10.23	10.26	42.16	41.5
16	79.49	78.87	81.27	79.72
17	152.11	150.81	131.8	132.56
18	136.92	136.44	126.24	119.79
19	80.1	83.06	92.91	85.85
20	18.34	11.23	49.62	49.01
21	122.62	124.12	121.17	122.41
22	81.8	82.34	83.37	86.82
23	35.86	34.3	46.93	48.2
24	131.43	132.06	145.83	146.94

25	88.55	89.88	77.76	77.41
26	27.32	27.19	32.54	32.93
27	136.08	133.81	112.05	114.6
28	28.62	26.94	37.38	42.61
29	101.3	100.18	80.63	84.63
30	139.38	139.96	136.34	137.31
31	101.33	99.41	83.16	85.67
32	-4.39	-7.22	31.9	34.24
33	138.28	137.86	125.01	125.45
34	26.56	26.5	29.44	35.94
35	123.97	127.21	115.27	117.72
36	100.46	98.3	74.76	75.87
37	13.75	12.85	47.47	47.96
38	141.7	140.74	137.95	136.94
39	91.84	93.55	76.4	73.02
40	35.95	34.31	32.88	32.66
41	105.82	108.57	98.93	96.57
42	22.71	21.56	29.8	28.73
43	141.88	142.93	116.8	114.76
44	141.89	142.16	125.56	124.83
45	107.56	108.09	82.8	81.24
46	29	28.08	32.37	34.02
47	140.07	142.83	116.08	116.15
48	95.78	97.22	83.42	86.33
49	23.84	23.88	45.19	52.03
50	137.16	136.34	123.04	123.84
51	3.03	3.78	60.5	56.67
52	23.83	23.81	86.45	84.19

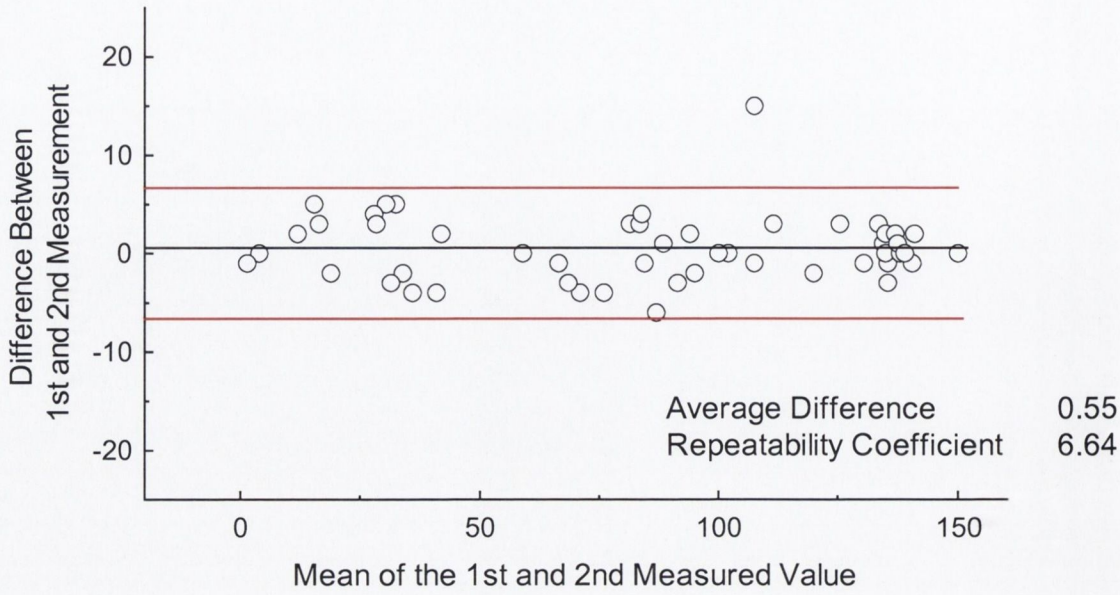
Appendix F

Scatterplots for the Intra-rater reliability of the different measurement techniques.

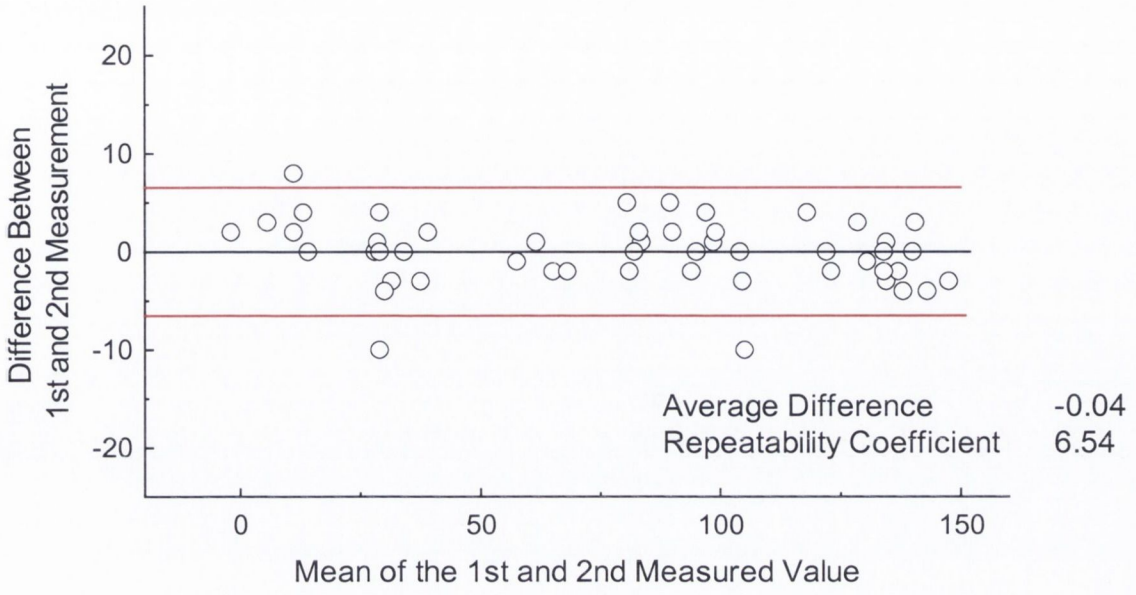
Intra-rater Reliability of the Knee Angle Measurments.
Universal Goniometer. Examiner 1 (n=51)



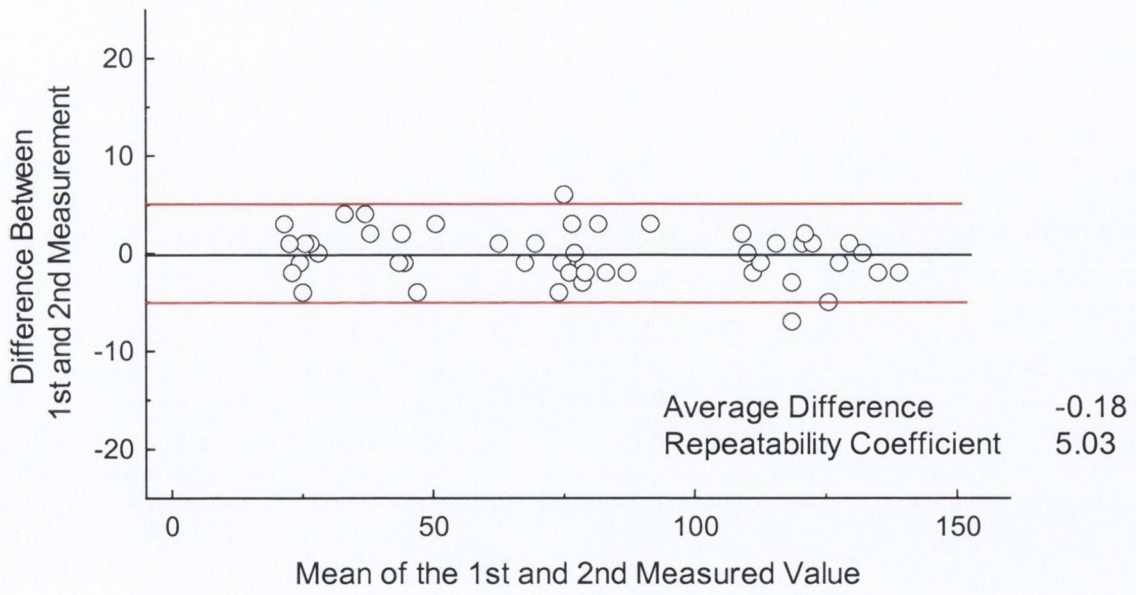
Intra-rater Reliability of the Knee Angle Measurments.
Universal Goniometer. Examiner 2 (n=51)



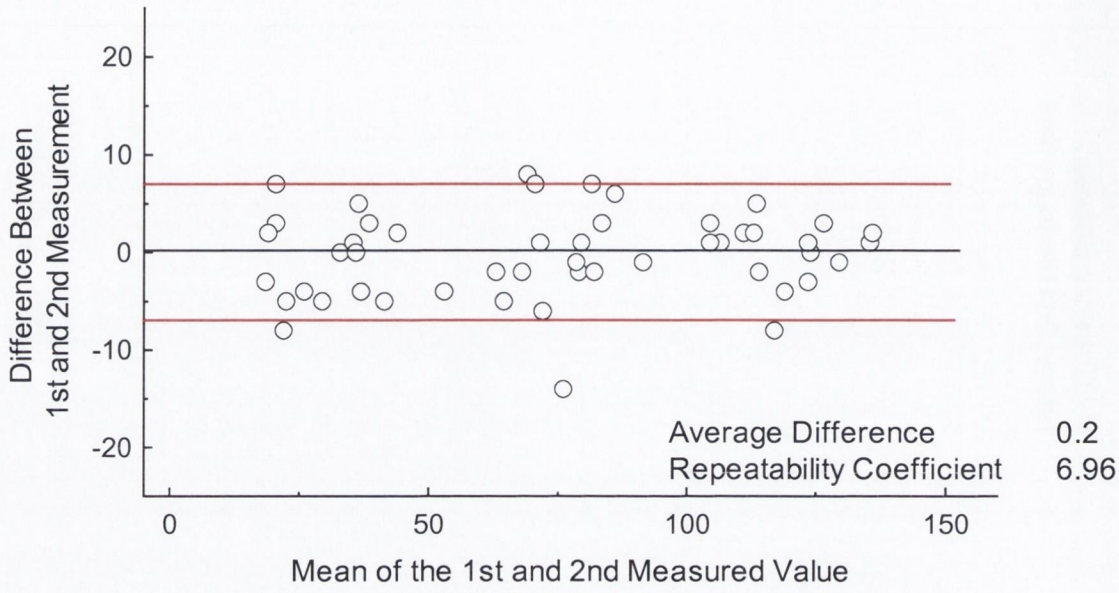
Intra-rater Reliability of the Knee Angle Measurments.
Universal Goniometer. Examiner 3 (n=51)



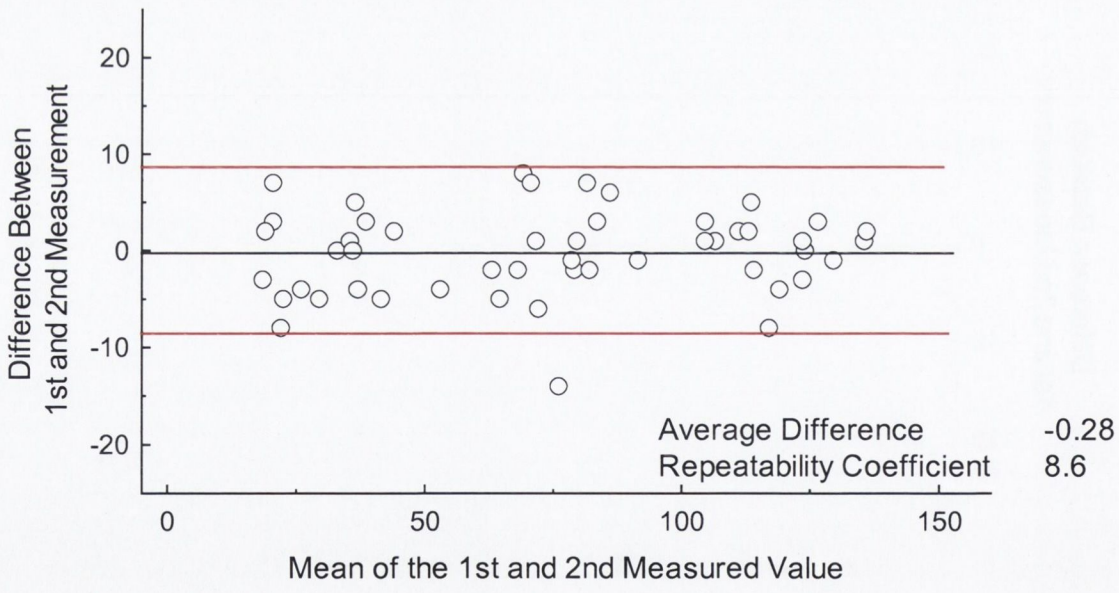
Intra-rater Reliability of the Elbow Angle Measurments.
Universal Goniometer. Examiner 1 (n=49)



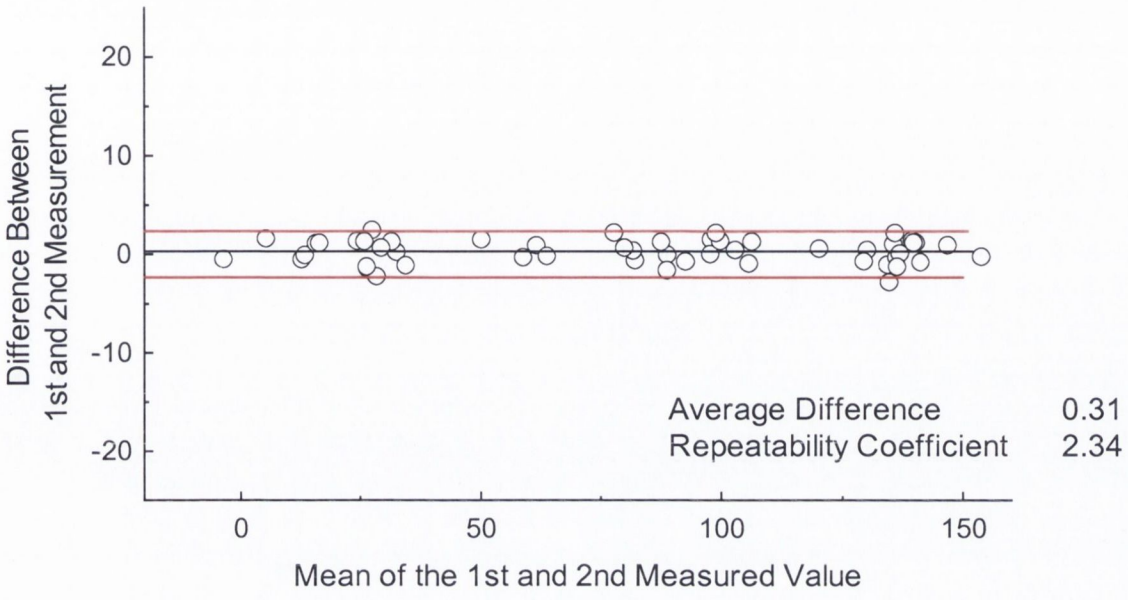
Intra-rater Reliability of the Elbow Angle Measurments.
Universal Goniometer. Examiner 2 (n=49)



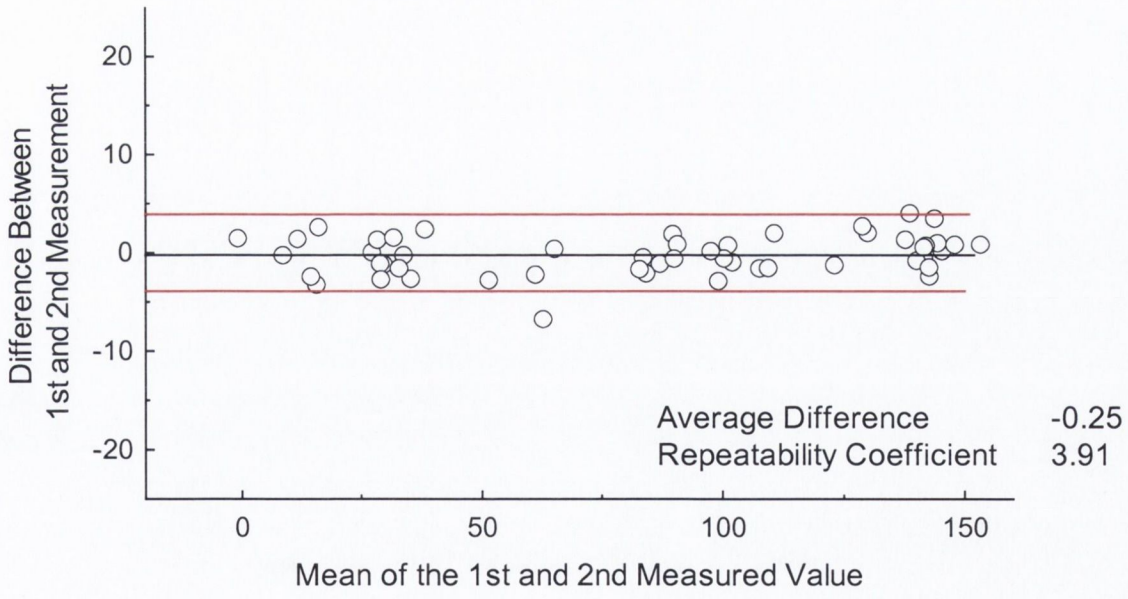
Intra-rater Reliability of the Elbow Angle Measurments.
Universal Goniometer. Examiner 3 (n=49)



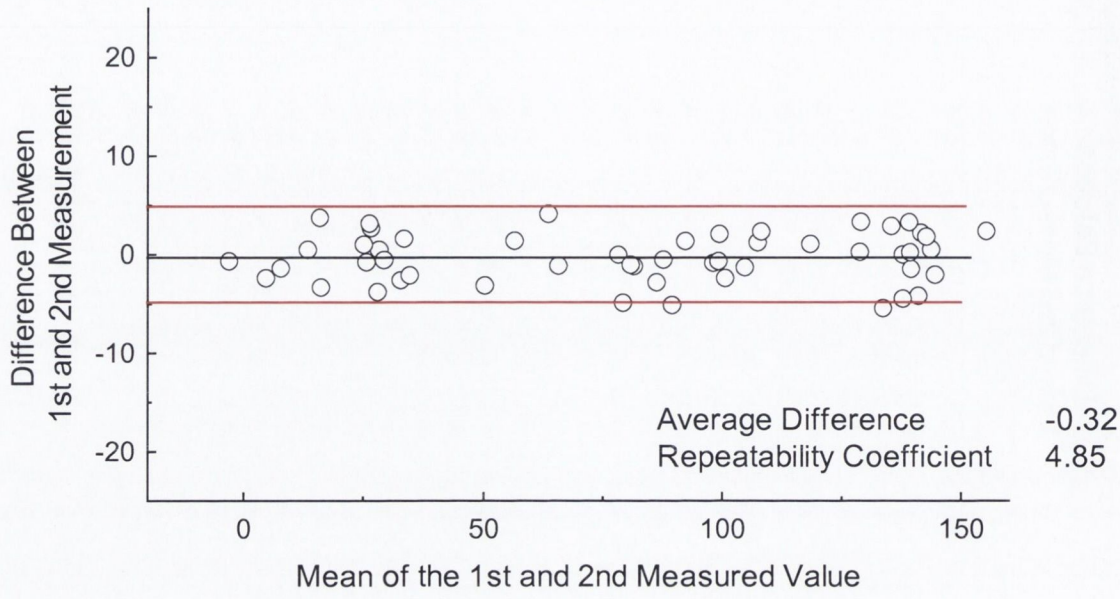
Intra-rater Reliability of the Knee Angle Measurments.
Uillinn Method (No Marker). Examiner 1 (n=51)



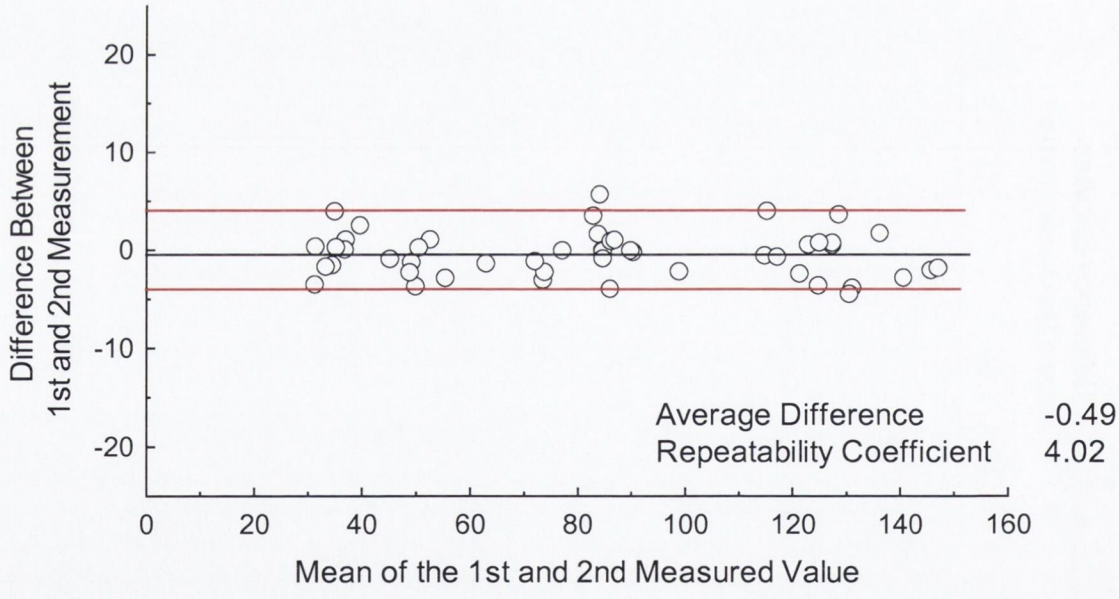
Intra-rater Reliability of the Knee Angle Measurments.
Uillinn Method (No Marker). Examiner 2 (n=51)



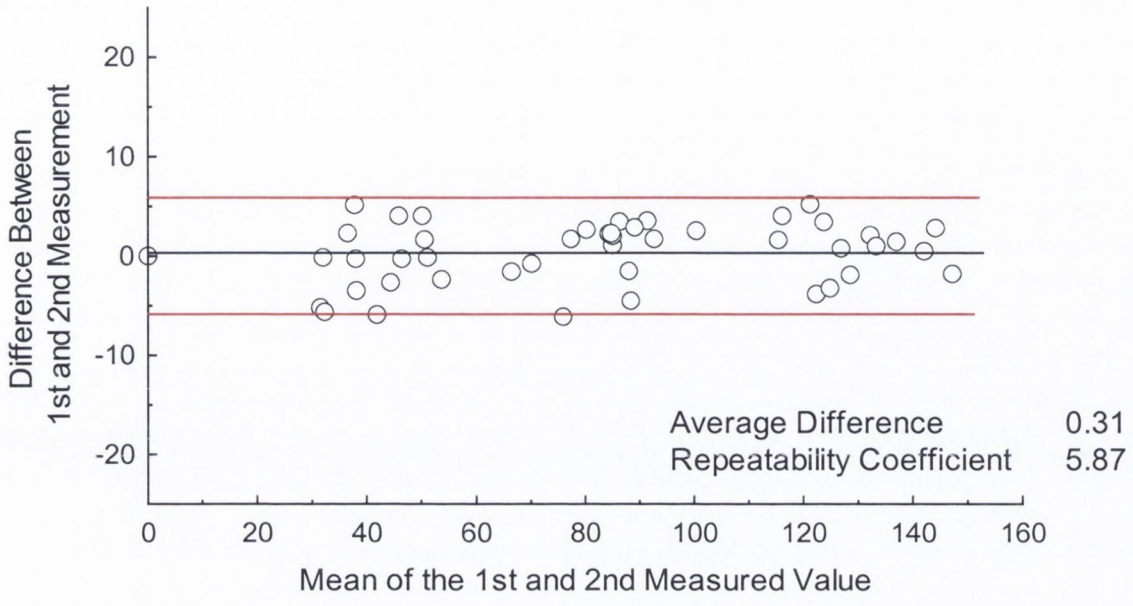
Intra-rater Reliability of the Knee Angle Measurments.
Uillinn Method (No Marker). Examiner 3 (n=51)



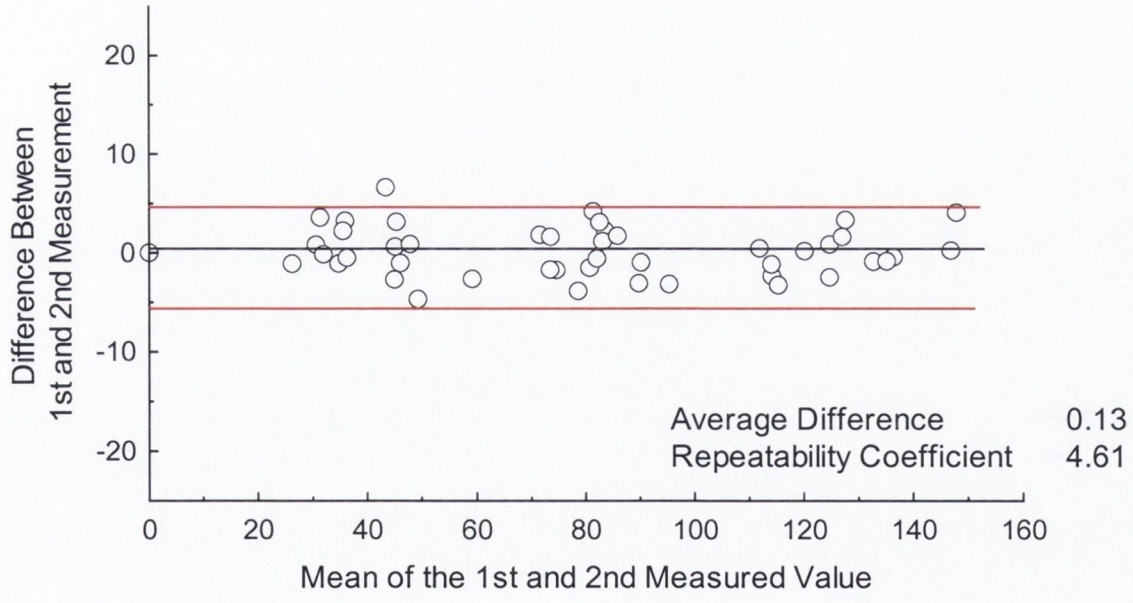
Intra-rater Reliability of the Elbow Angle Measurments.
Uillinn Method (No Marker). Examiner 1 (n=45)



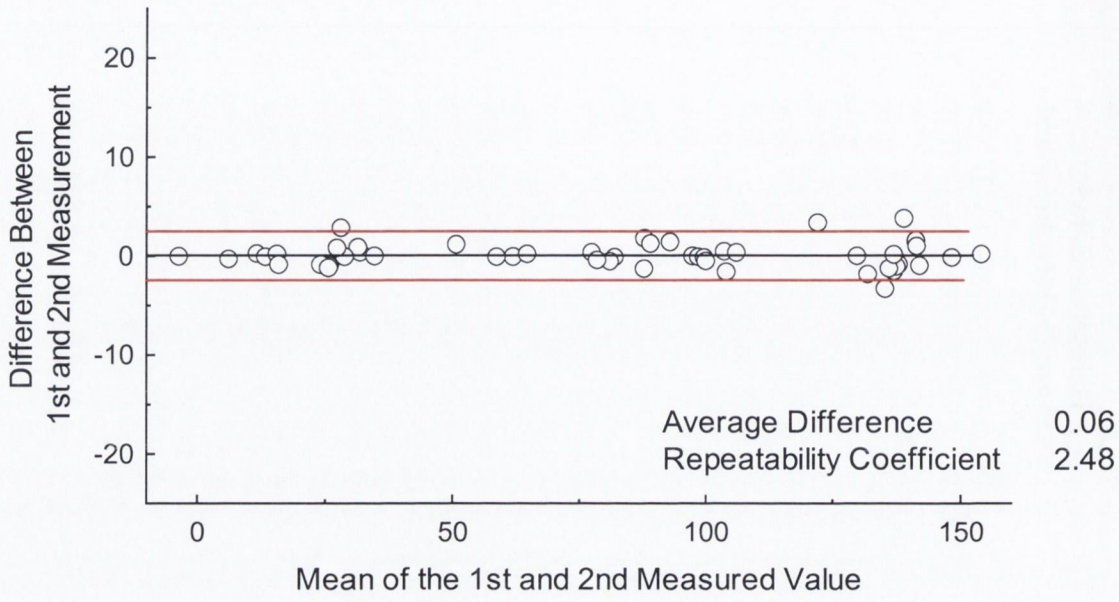
Intra-rater Reliability of the Elbow Angle Measurments.
Uillinn Method (No Marker). Examiner 2 (n=45)



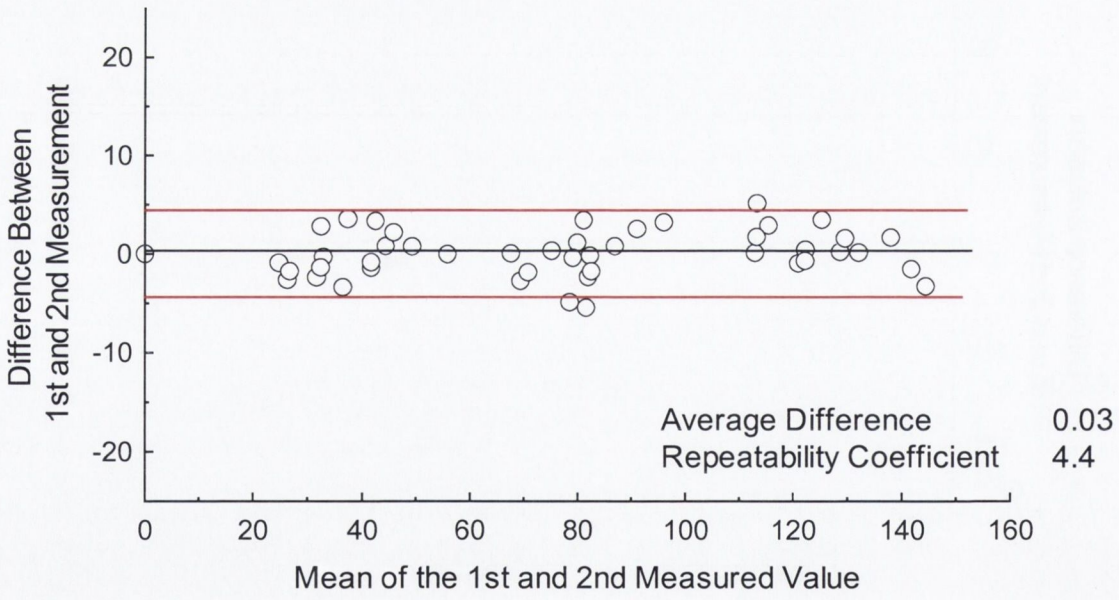
Intra-rater Reliability of the Elbow Angle Measurments.
Uillinn Method (No Marker). Examiner 3 (n=45)



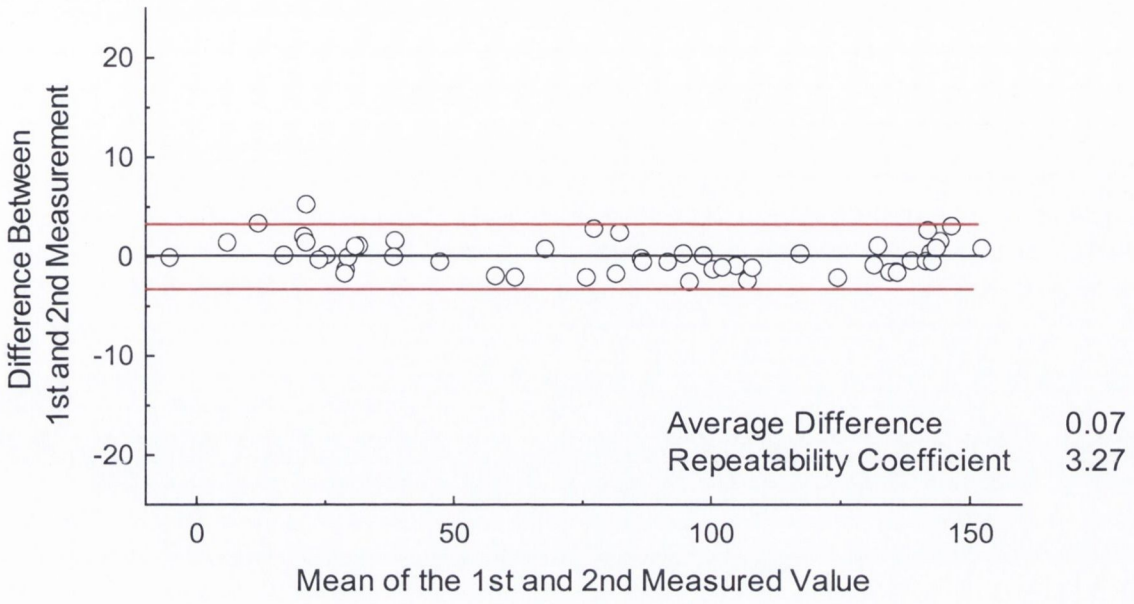
Intra-rater Reliability of the Knee Angle Measurments.
Uillinn Method (Parallel). Examiner 1 (n=51)



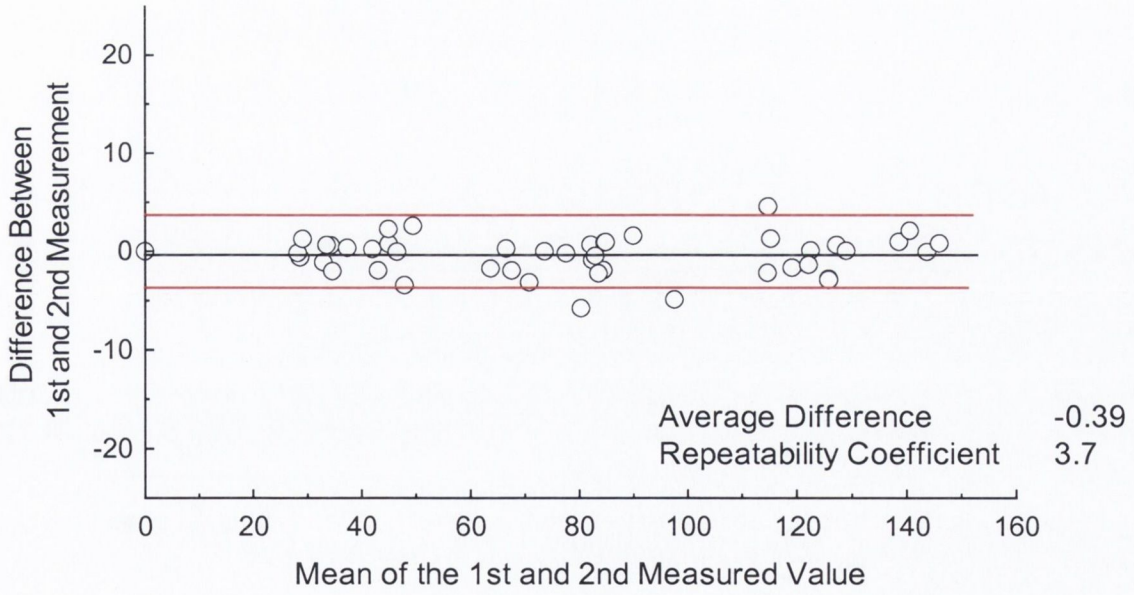
Intra-rater Reliability of the Elbow Angle Measurments.
Uillinn Method (Parallel). Examiner 2 (n=45)



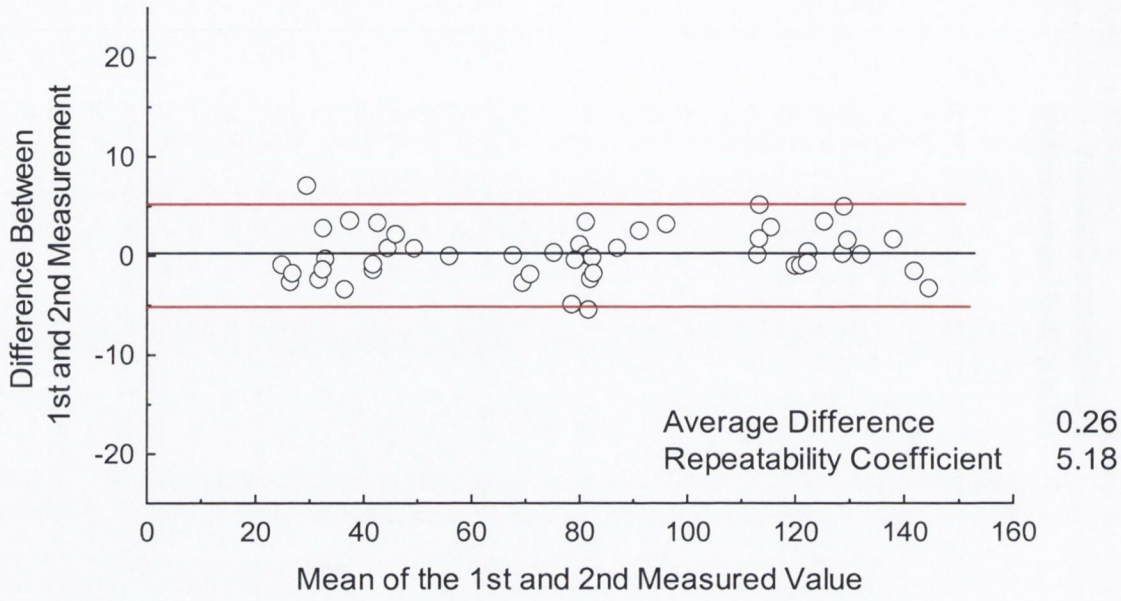
Intra-rater Reliability of the Knee Angle Measurments.
Uillinn Method (Parallel). Examiner 3 (n=51)



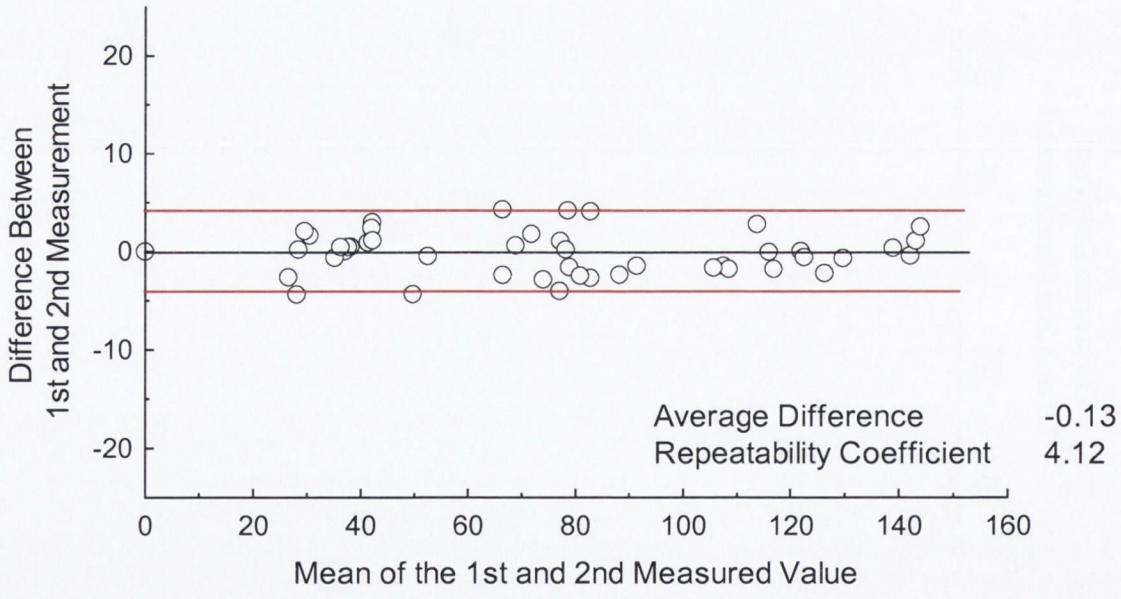
Intra-rater Reliability of the Elbow Angle Measurments.
Uillinn Method (Parallel). Examiner 1 (n=45)



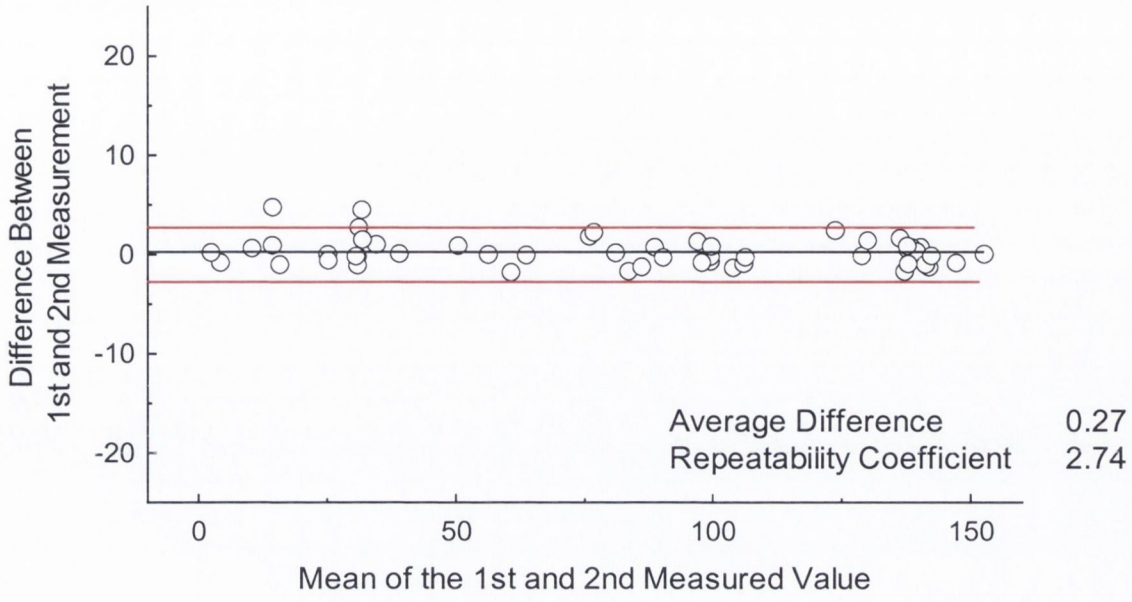
Intra-rater Reliability of the Elbow Angle Measurments.
Uillinn Method (Parallel). Examiner 2 (n=49)



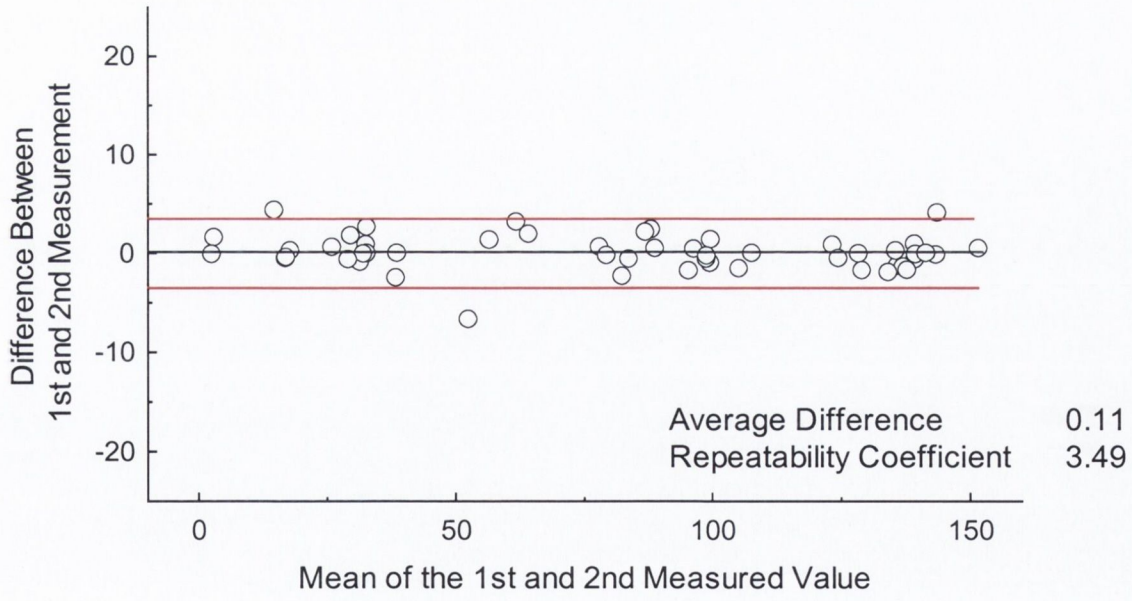
Intra-rater Reliability of the Elbow Angle Measurments.
Uillinn Method (Parallel). Examiner 3 (n=45)



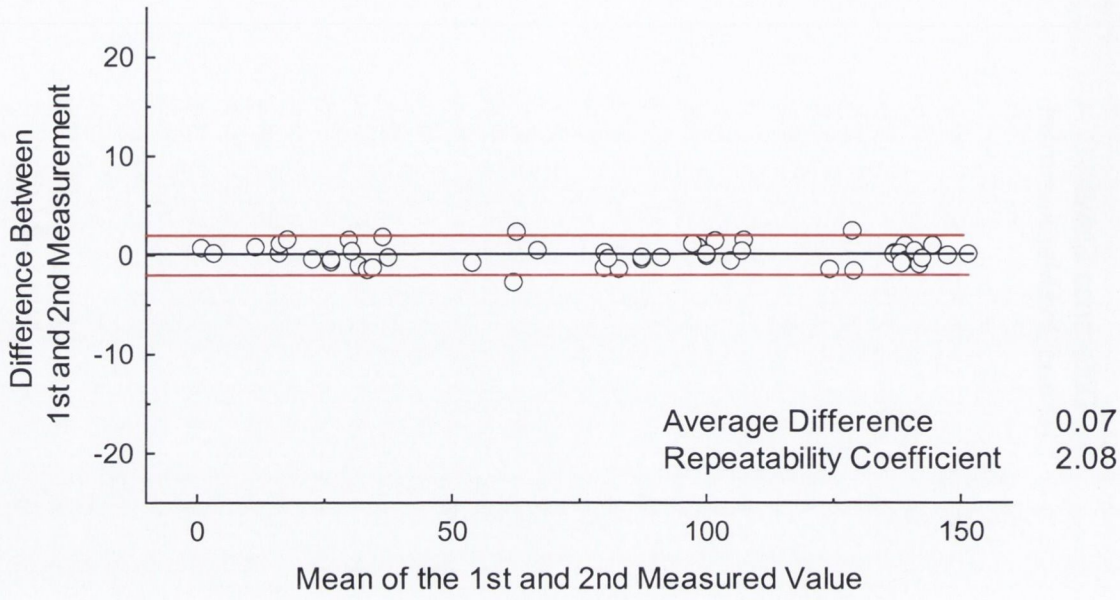
Intra-rater Reliability of the Knee Angle Measurments.
Uillinn Method (Mid-Point). Examiner 1 (n=51)



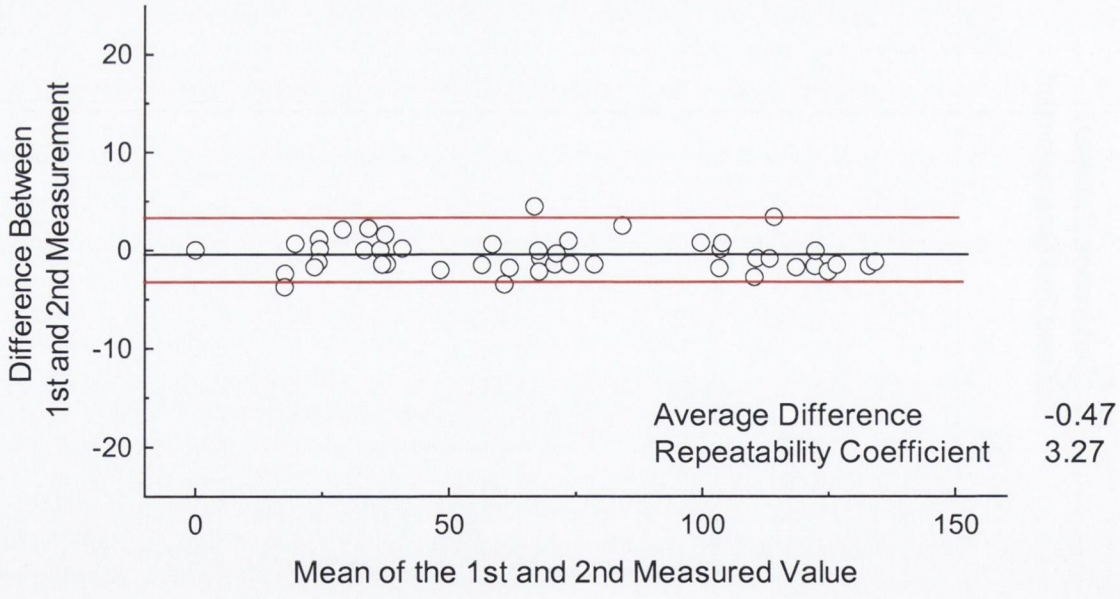
Intra-rater Reliability of the Knee Angle Measurments.
Uillinn Method (Mid-Point). Examiner 2 (n=51)



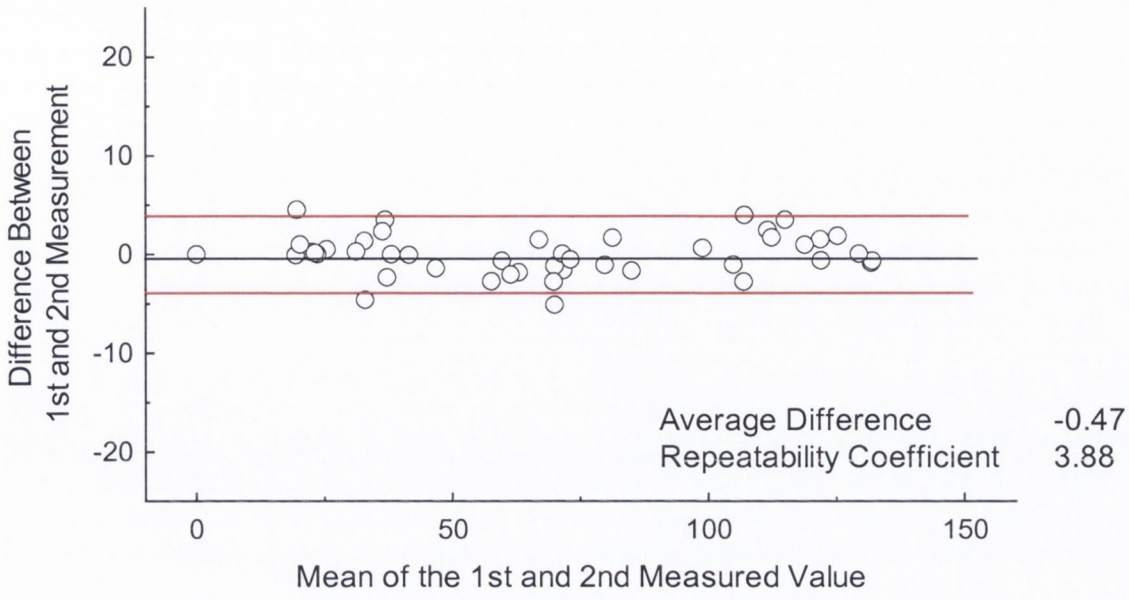
Intra-rater Reliability of the Knee Angle Measurments.
Uillinn Method (Mid-Point). Examiner 3 (n=51)



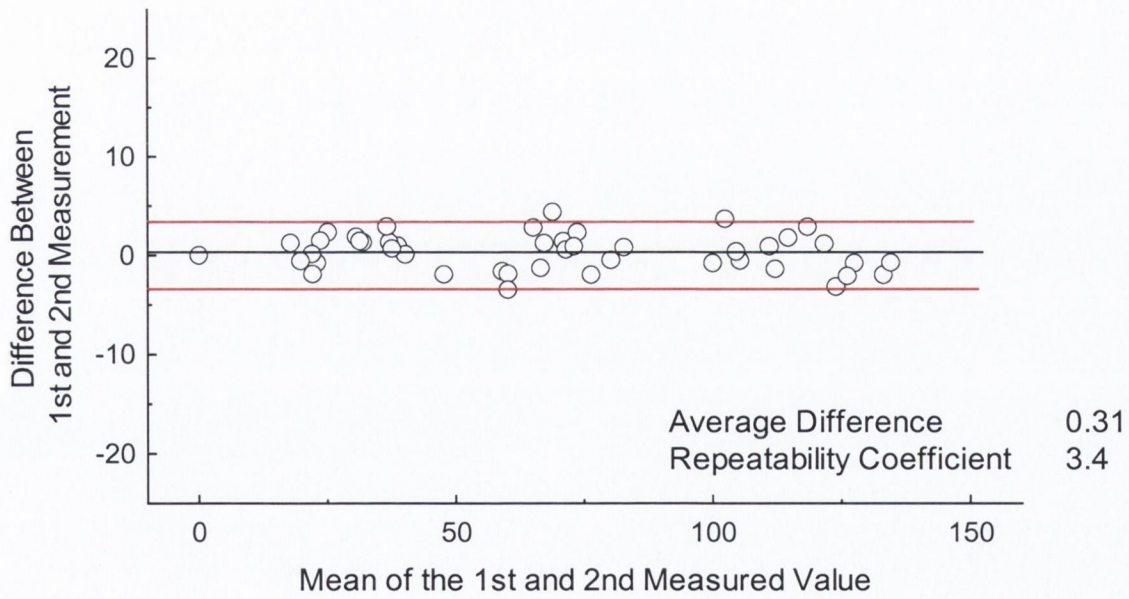
Intra-rater Reliability of the Elbow Angle Measurments.
Uillinn Method (Mid-Point). Examiner 1 (n=45)



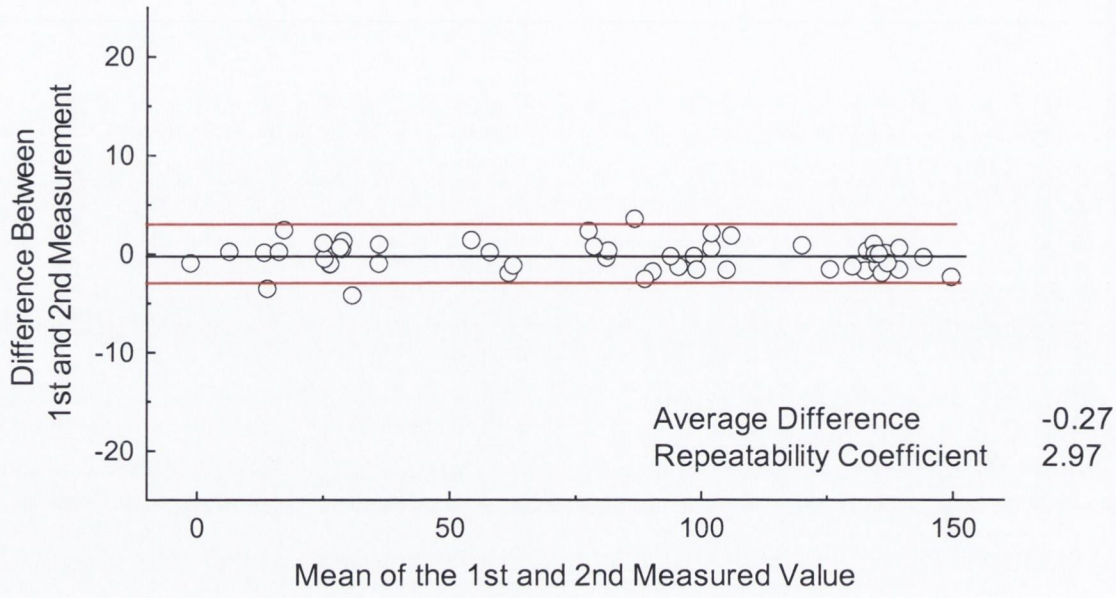
Intra-rater Reliability of the Elbow Angle Measurments.
Uillinn Method (Mid-Point). Examiner 2 (n=45)



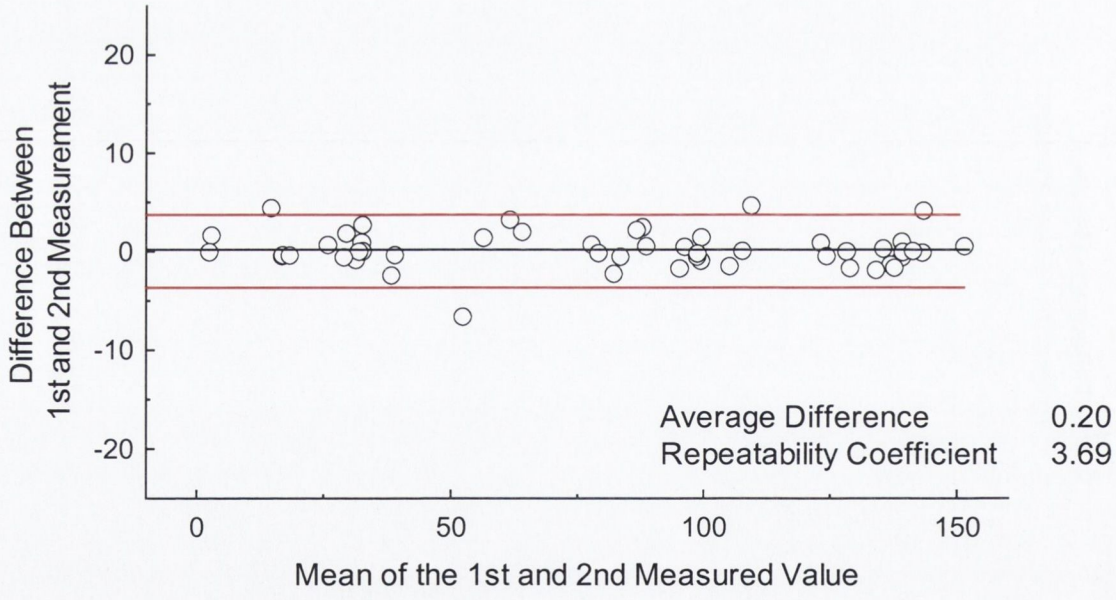
Intra-rater Reliability of the Elbow Angle Measurments.
Uillinn Method (Mid-Point). Examiner 3 (n=45)



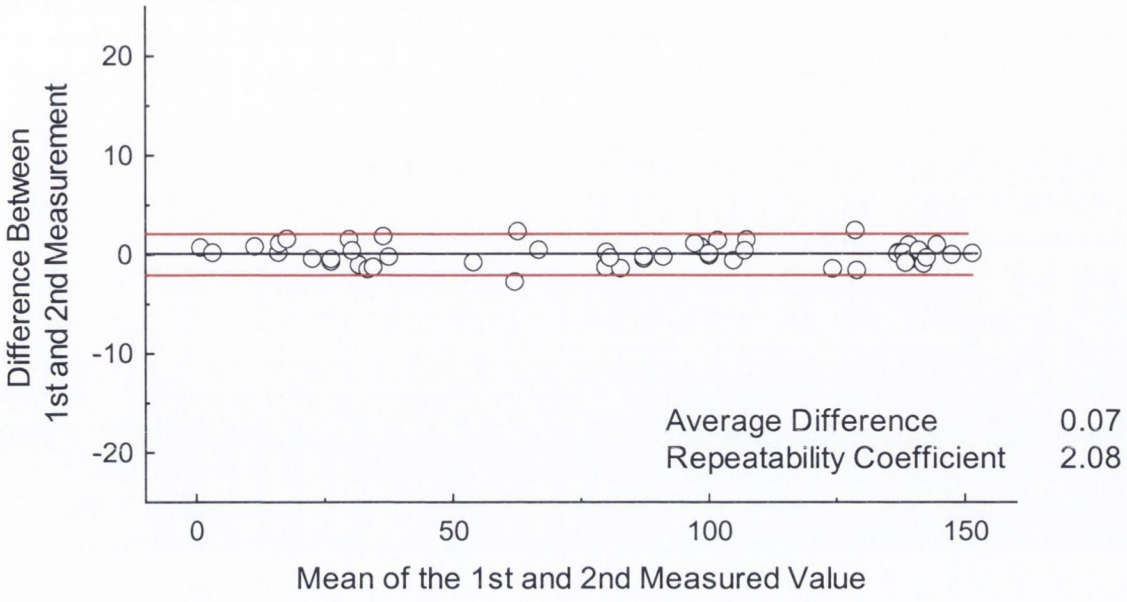
Intra-rater Reliability of the Knee Angle Measurments.
Uillinn Method (3 Markers). Examiner 1 (n=51)



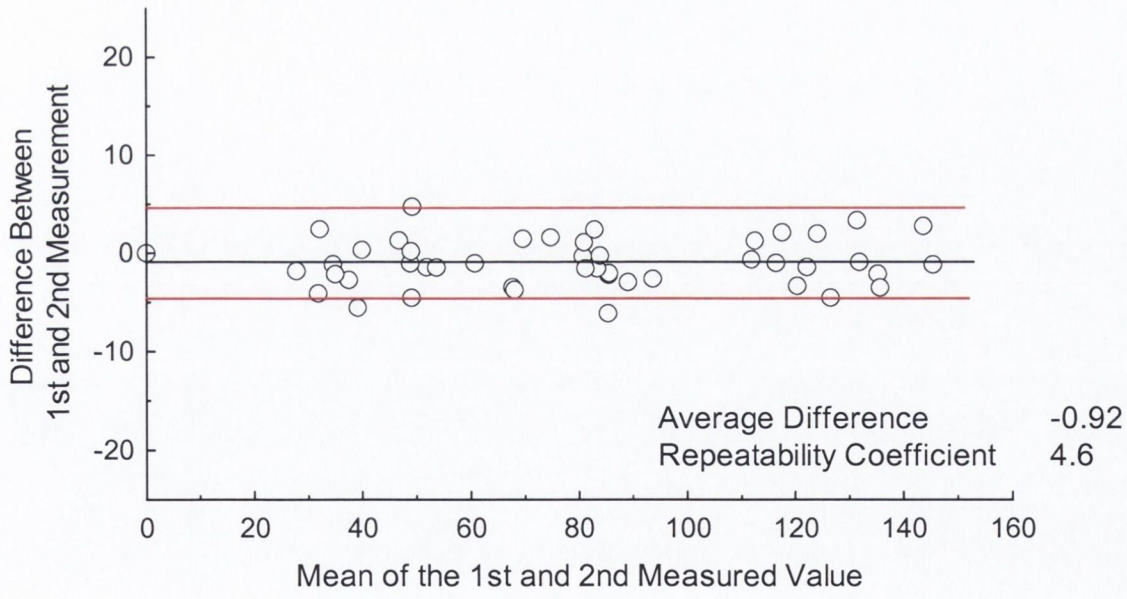
Intra-rater Reliability of the Knee Angle Measurments.
Uillinn Method (3 Markers). Examiner 2 (n=51)



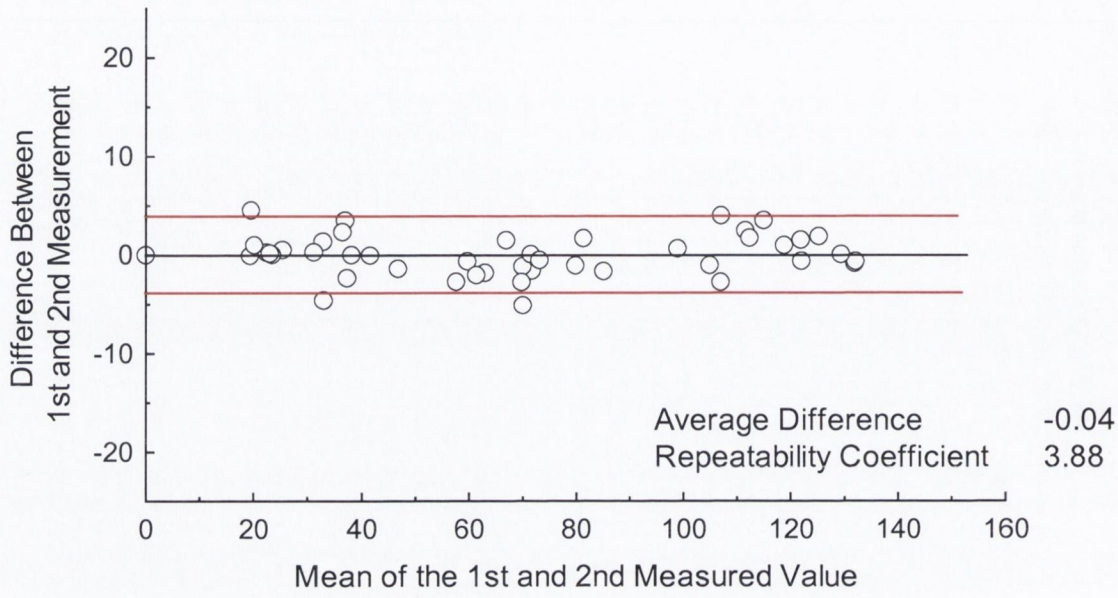
Intra-rater Reliability of the Knee Angle Measurments.
Uillinn Method (3 Markers). Examiner 3 (n=51)



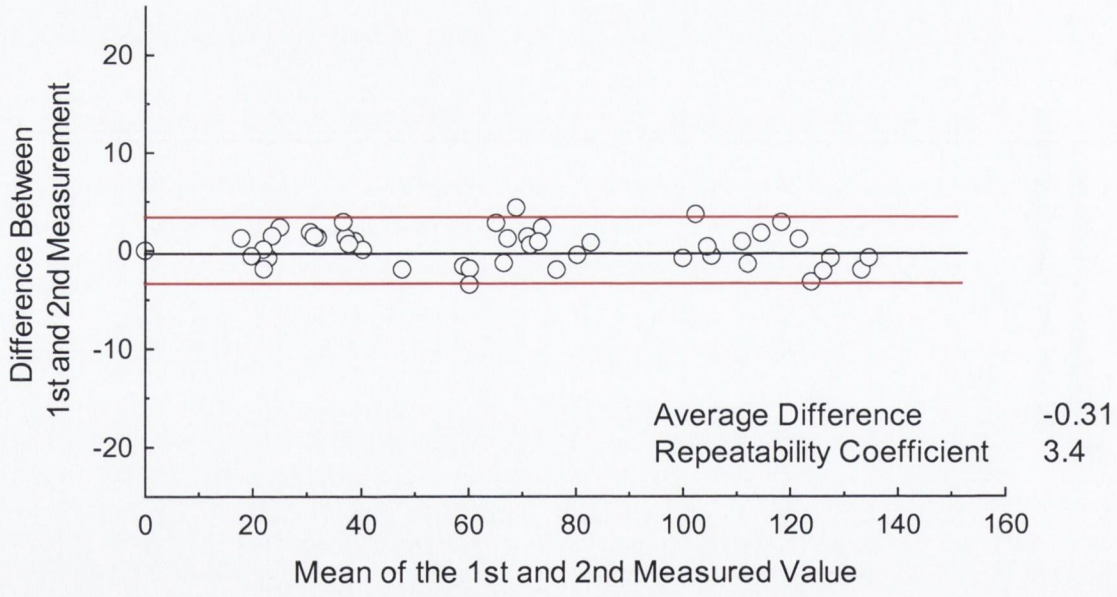
Intra-rater Reliability of the Elbow Angle Measurments.
Uillinn Method (3 Markers). Examiner 1 (n=45)



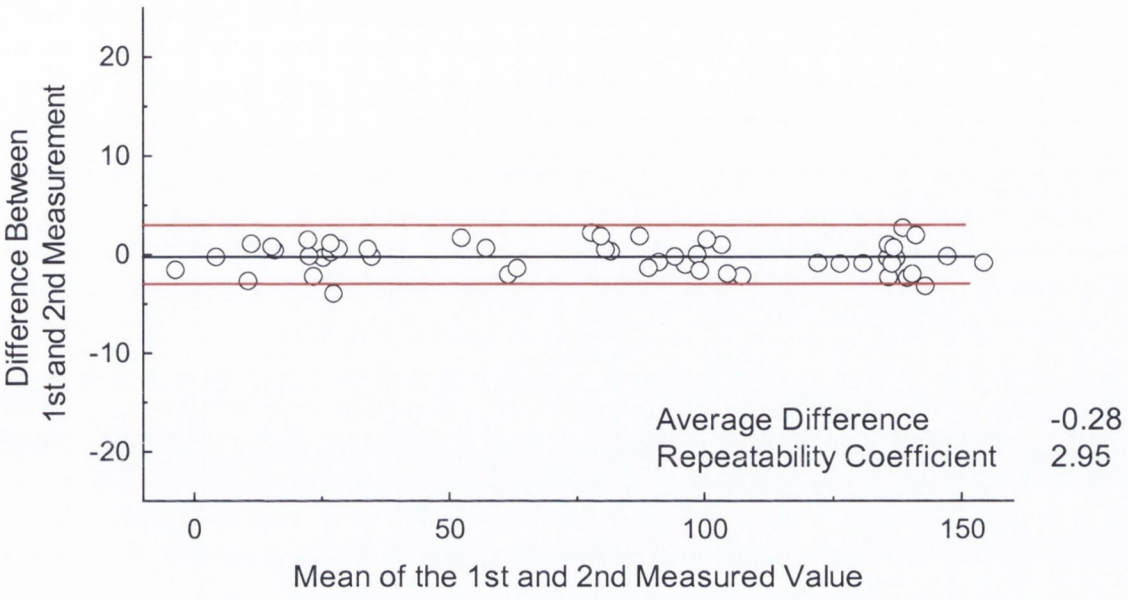
Intra-rater Reliability of the Elbow Angle Measurments.
Uillinn Method (3 Markers). Examiner 2 (n=45)



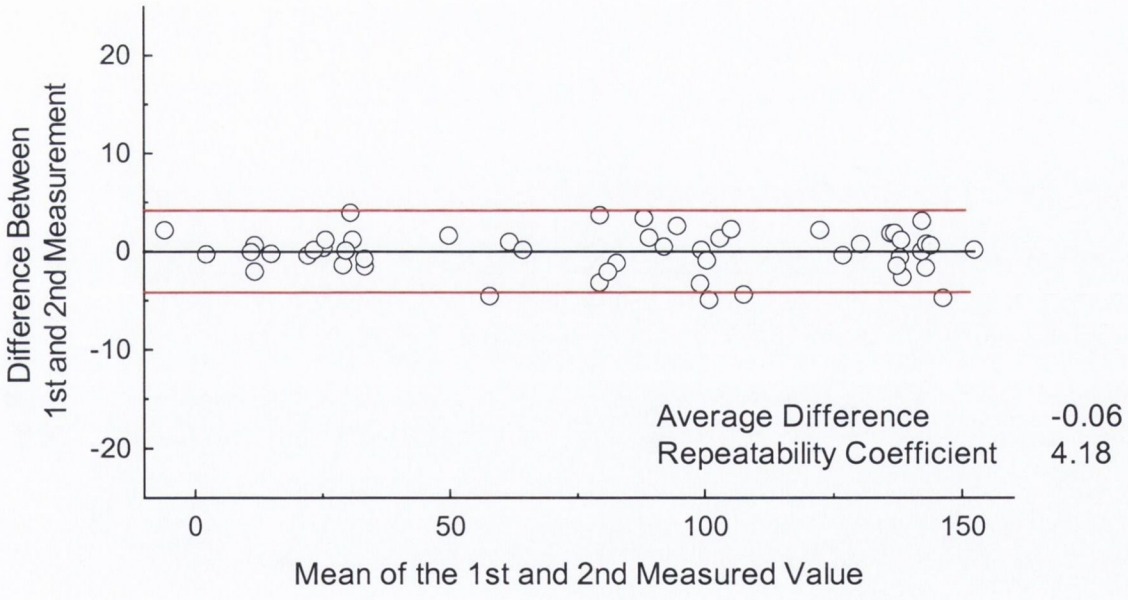
Intra-rater Reliability of the Elbow Angle Measurments.
Uillinn Method (3 Markers). Examiner 3 (n=45)



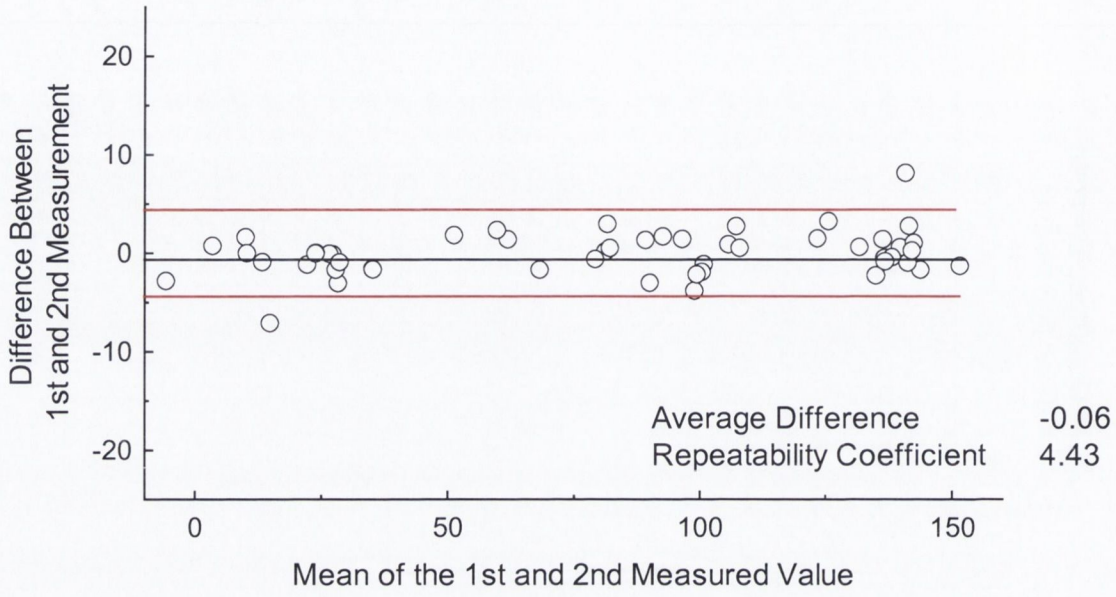
Intra-rater Reliability of the Knee Angle Measurments.
Uillinn Method (4 Markers). Examiner 1 (n=51)



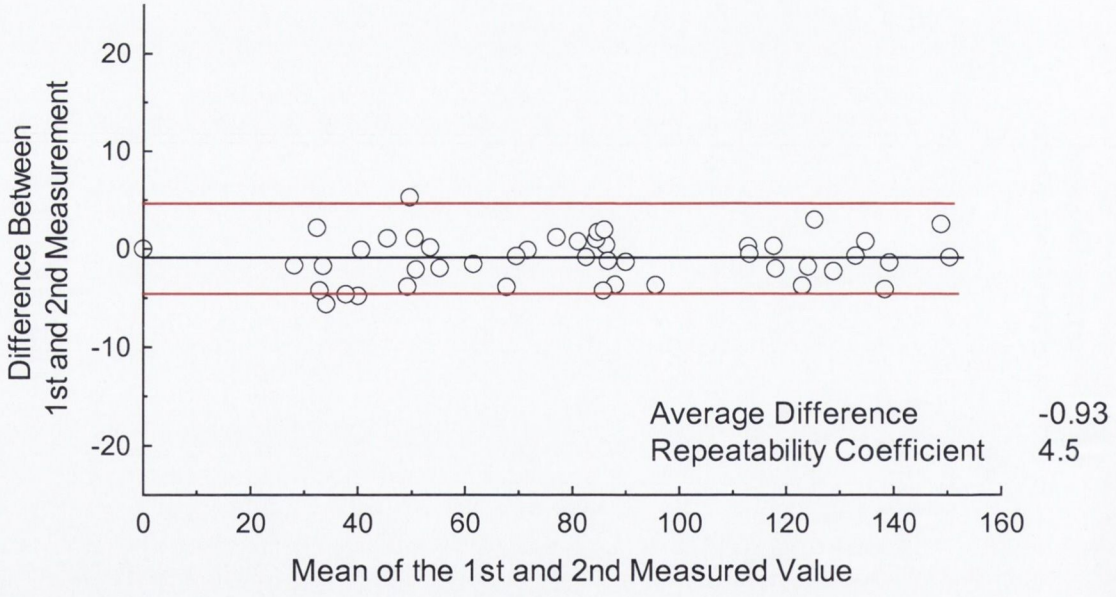
Intra-rater Reliability of the Knee Angle Measurments.
Uillinn Method (4 Markers). Examiner 2 (n=51)



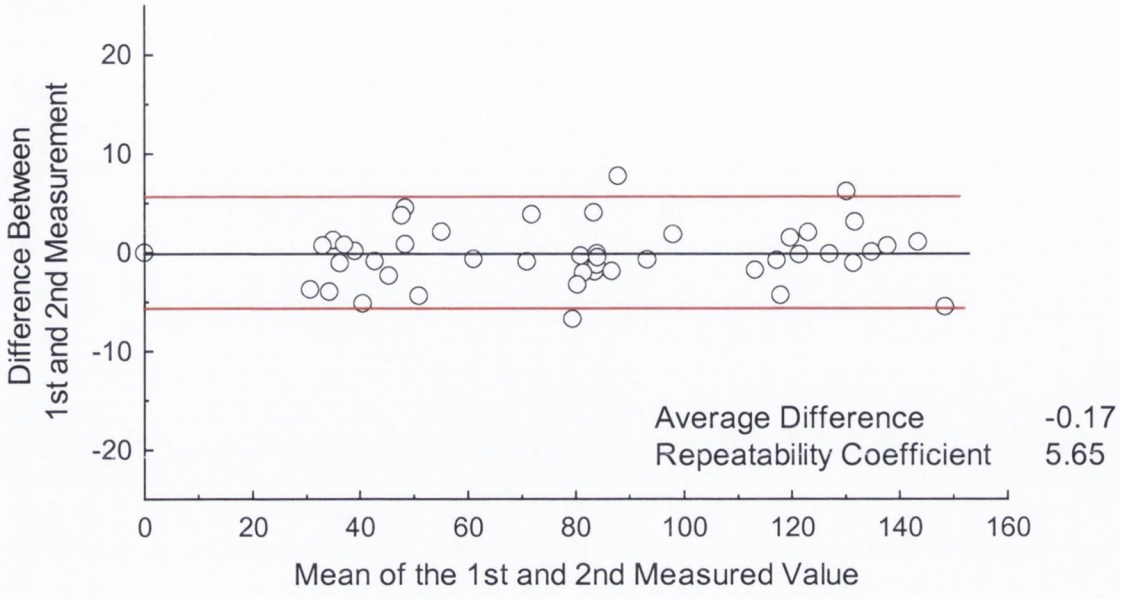
Intra-rater Reliability of the Knee Angle Measurments.
Uillinn Method (4 Markers). Examiner 3 (n=51)



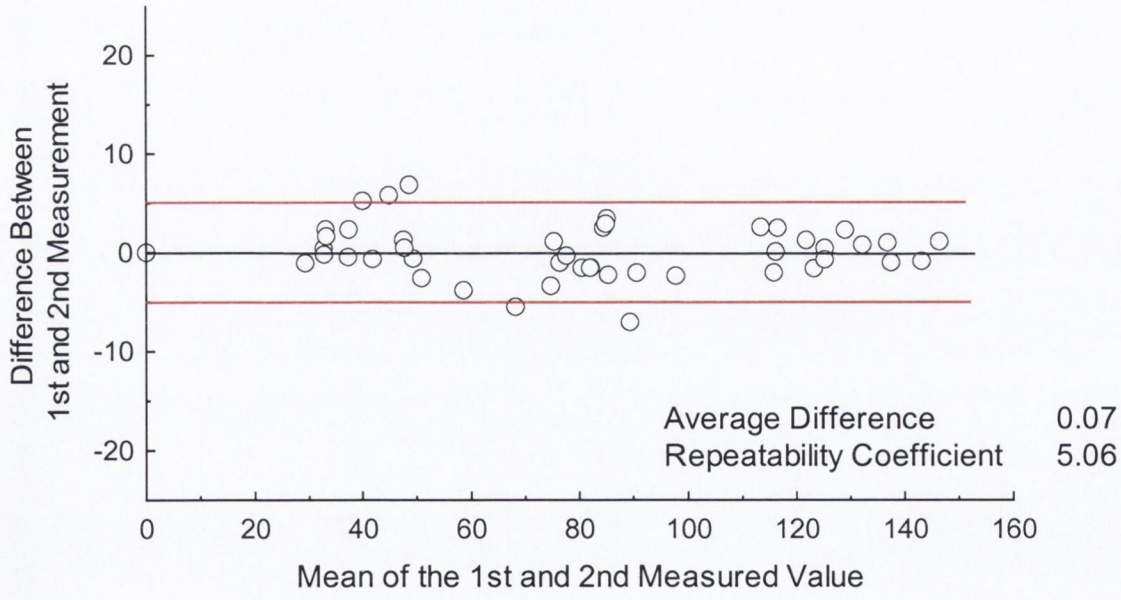
Intra-rater Reliability of the Elbow Angle Measurments.
Uillinn Method (4 Markers). Examiner 1 (n=45)



Intra-rater Reliability of the Elbow Angle Measurments.
Uillinn Method (4 Markers). Examiner 2 (n=45)



Intra-rater Reliability of the Elbow Angle Measurments.
Uillinn Method (4 Markers). Examiner 3 (n=45)



Appendix G

Significance Testing for Inter-rater Reliability

t-Test: Paired Two Sample for Means

UG KNEES	*Ex1*	*Ex2*	*Ex1*	*Ex3*	*Ex2*	*Ex3*
Mean	81.57	83.18	81.57	81.27	83.18	81.27
Variance	2159.61	2100.51	2159.61	2093.00	2100.51	2093.00
Observations	51.00	51.00	51.00	51.00	51.00	51.00
Pearson Correlation	0.99		1.00		0.99	
Hypothesized Mean Difference	0.00		0.00		0.00	
df	50.00		50.00		50.00	
t Stat	-1.76		0.57		2.31	
P(T<=t) one-tail	0.04		0.29		0.01	
t Critical one-tail	1.68		1.68		1.68	
P(T<=t) two-tail	0.09	NS	0.57	NS	0.03	NS
t Critical two-tail	2.01		2.01		2.01	

t-Test: Paired Two Sample for Means

UG Elbows	*Ex1*	*Ex2*	*Ex1*	*Ex3*	*Ex2*	*Ex3*
Mean	76.65	78.96	76.65	74.47	76.65	78.96
Variance	1457.83	1456.84	1457.83	1509.97	1457.83	1456.84
Observations	51.00	51.00	51.00	51.00	51.00	51.00
Pearson Correlation	0.96		1.00		0.96	
Hypothesized Mean Difference	0.00		0.00		0.00	
df	50.00		50.00		50.00	
t Stat	-1.48		4.06		-1.48	
P(T<=t) one-tail	0.07		0.00		0.07	
t Critical one-tail	1.68		1.68		1.68	
P(T<=t) two-tail	0.15	NS	0.00	Sig	0.15	NS
t Critical two-tail	2.01		2.01		2.01	

t-Test: Paired Two Sample for Means

Knee – No Markers	*Ex1*	*Ex2*	*Ex1*	*Ex3*	*Ex2*	*Ex3*
Mean	81.70	84.18	81.70	81.86	84.18	81.86
Variance	2283.82	2362.61	2283.82	2350.05	2362.61	2350.05
Observations	51.00	51.00	51.00	51.00	51.00	51.00
Pearson Correlation	1.00		1.00		1.00	
Hypothesized Mean Difference	0.00		0.00		0.00	
df	50.00		50.00		50.00	
t Stat	-6.10		-0.45		5.29	
P(T<=t) one-tail	0.00		0.33		0.00	
t Critical one-tail	1.68		1.68		1.68	
P(T<=t) two-tail	0.0000002	Sig	0.65	NS	0.000003	Sig
t Critical two-tail	2.01		2.01		2.01	

t-Test: Paired Two Sample for Means

Elbow – No Markers	*Ex1*	*Ex2*	*Ex1*	*Ex3*	*Ex2*	*Ex3*
Mean	83.90	85.51	83.90	81.88	85.51	81.88
Variance	1331.27	1382.46	1331.27	1384.13	1382.46	1384.13
Observations	49.00	49.00	49.00	49.00	49.00	49.00
Pearson Correlation	1.00				0.99	
Hypothesized Mean Difference	0.00		0.00		0.00	
df	48.00		48.00		48.00	
t Stat	-3.44		4.43		6.27	
P(T<=t) one-tail	0.00		0.00		0.00	
t Critical one-tail	1.68		1.68		1.68	
P(T<=t) two-tail	0.0012	Sig	0.0001	Sig	0.0000001	Sig
t Critical two-tail	2.01		2.01		2.01	

t-Test: Paired Two Sample for Means

Knee - Parallel	*Ex*	*Ex2*	*Ex1*	*Ex3*	*Ex2*	*Ex3*
Mean	81.88	81.35	81.88	82.42	81.35	82.42
Variance	2337.05	2487.47	2337.05	2271.16	2487.47	2271.16
Observations	51.00	51.00	51.00	51.00	51.00	51.00
Pearson Correlation	1.00		1.00		1.00	
Hypothesized Mean Difference	0.00		0.00		0.00	
df	50.00		50.00		50.00	
t Stat	1.19		-0.99		-1.65	
P(T<=t) one-tail	0.12		0.16		0.05	
t Critical one-tail	1.68		1.68		1.68	
P(T<=t) two-tail	0.24	NS	0.33	NS	0.11	NS
t Critical two-tail	2.01		2.01		2.01	

t-Test: Paired Two Sample for Means

Elbow - Parallel	*Ex*	*Ex2*	*Ex1*	*Ex3*	*Ex2*	*Ex3*
Mean	81.23	80.25	81.23	78.04	80.25	78.04
Variance	1408.98	1426.55	1408.98	1365.96	1426.55	1365.96
Observations	49.00	49.00	49.00	49.00	49.00	49.00
Pearson Correlation	1.00		0.99		0.99	
Hypothesized Mean Difference	0.00		0.00		0.00	
df	48.00		48.00		48.00	
t Stat	2.01		4.92		3.24	
P(T<=t) one-tail	0.03		0.00		0.00	
t Critical one-tail	1.68		1.68		1.68	
P(T<=t) two-tail	0.05	Sig	0.00001	Sig	0.002	Sig
t Critical two-tail	2.01		2.01		2.01	

t-Test: Paired Two Sample for Means

Mid-Point - Knee	*Ex1*	*Ex2*	*Ex1*	*Ex3*	*Ex2*	*Ex3*
Mean	82.66	82.36	82.66	83.21	82.36	83.21
Variance	2271.58	2148.06	2271.58	2281.02	2148.06	2281.02
Observations	51.00	51.00	51.00	51.00	51.00	51.00
Pearson Correlation	1.00		1.00		1.00	
Hypothesized Mean Difference	0.00		0.00		0.00	
df	50.00		50.00		50.00	
t Stat	0.70		-1.97		-1.93	
P(T<=t) one-tail	0.24		0.03		0.03	
t Critical one-tail	1.68		1.68		1.68	
P(T<=t) two-tail	0.49	NS	0.05	NS +/-	0.06	NS
t Critical two-tail	2.01		2.01		2.01	

t-Test: Paired Two Sample for Means

Mid-Point - Elbow	*Ex1*	*Ex2*	*Ex1*	*Ex3*	*Ex2*	*Ex3*
Mean	70.4	71.4	70.4	70.7	71.4	70.7
Variance	1342.2	1367.7	1342.2	1355.5	1367.7	1355.5
Observations	49.0	49.0	49.0	49.0	49.0	49.0
Pearson Correlation	1.0		1.0		1.0	
Hypothesized Mean Difference	0.0		0.0		0.0	
df	48.0		48.0		48.0	
t Stat	-2.9		-0.5		1.4	
P(T<=t) one-tail	0.0		0.3		0.1	
t Critical one-tail	1.7		1.7		1.7	
P(T<=t) two-tail	0.01	Sig	0.6	NS	0.2	NS
t Critical two-tail	2.0		2.0		2.0	

t-Test: Paired Two Sample for Means

3 Marker Knee	*Ex1*	*Ex2*	*Ex1*	*Ex3*	*Ex2*	*Ex3*
Mean	81.07	82.29	81.07	82.58	82.29	82.58
Variance	2185.71	2278.60	2185.71	2272.89	2278.60	2272.89
Observations	51.00	51.00	51.00	51.00	51.00	51.00
Pearson Correlation	1.00		1.00		1.00	
Hypothesized Mean Difference	0.00		0.00		0.00	
df	50.00		50.00		50.00	
t Stat	-3.76		-4.54		-1.05	
P(T<=t) one-tail	0.00		0.00		0.15	
t Critical one-tail	1.68		1.68		1.68	
P(T<=t) two-tail	0.0004	Sig	0.00004	Sig	0.30	NS
t Critical two-tail	2.01		2.01		2.01	

t-Test: Paired Two Sample for Means

3 Marker Elbow	*Ex1*	*Ex2*	*Ex1*	*Ex3*	*Ex2*	*Ex3*
Mean	81.91	81.50	81.91	81.70	81.50	81.70
Variance	1306.25	1273.32	1306.25	1286.56	1273.32	1286.56
Observations	49.00	49.00	49.00	49.00	49.00	49.00
Pearson Correlation	1.00		1.00		1.00	
Hypothesized Mean Difference	0.00		0.00		0.00	
df	48.00		48.00		48.00	
t Stat	0.85		0.50		-0.46	
P(T<=t) one-tail	0.20		0.31		0.32	
t Critical one-tail	1.68		1.68		1.68	
P(T<=t) two-tail	0.40	NS	0.62	NS	0.65	NS
t Critical two-tail	2.01		2.01		2.01	

t-Test: Paired Two Sample for Means

4 Marker Knee	*Ex1*	*Ex2*	*Ex1*	*Ex3*	*Ex2*	*Ex3*
Mean	81.19	81.93	81.19	82.01	81.93	82.01
Variance	2355.61	2396.53	2355.61	2418.67	2396.53	2418.67
Observations	51.00	51.00	51.00	51.00	51.00	51.00
Pearson Correlation	1.00		1.00		1.00	
Hypothesized Mean Difference	0.00		0.00		0.00	
df	50.00		50.00		50.00	
t Stat	-2.08		-2.62		-0.24	
P(T<=t) one-tail	0.02		0.01		0.40	
t Critical one-tail	1.68		1.68		1.68	
P(T<=t) two-tail	0.04	Sig	0.01	Sig	0.81	NS
t Critical two-tail	2.01		2.01		2.01	

t-Test: Paired Two Sample for Means

4 Marker Elbow	*Ex1*	*Ex2*	*Ex1*	*Ex3*	*Ex2*	*Ex3*
Mean	83.29	83.66	83.29	82.60	83.66	82.60
Variance	1371.04	1323.83	1371.04	1312.84	1323.83	1312.84
Observations	49.00	49.00	49.00	49.00	49.00	49.00
Pearson Correlation	0.99		1.00		1.00	
Hypothesized Mean Difference	0.00		0.00		0.00	
df	48.00		48.00		48.00	
t Stat	-0.68		1.49		2.29	
P(T<=t) one-tail	0.25		0.07		0.01	
t Critical one-tail	1.68		1.68		1.68	
P(T<=t) two-tail	0.50	NS	0.14	NS	0.03	NS
t Critical two-tail	2.01		2.01		2.01	

Appendix H

Validity Measurement Data (Knee).

MRN	PACS	UG	No Markers	Parallel	Mid Point	3 Markers	4 Markers
13081	72.38	71.5	70.1	66.34	66.84	73.7	75.2
22429	87.95	91.5	87.89	89.64	86.77	91.21	90.55
42004	63.66	63	62.7	57.7	57.81	67.1	65.2
53808	55.98	52	50	48.75	49.47	50.9	45
63067	63.45	72	61.6	57.78	63.14	66.2	60.9
74318	54.3	57.5	60.5	48.89	54.38	63.2	60.7
84051	40.8	42.5	37.3	41.77	42.31	54.8	37.3
120923	66.35	68	66	64.55	65.81	68.1	66.7
146405	25.9	29	31.3	25.88	25.42	21.1	20.3
166412	53.88	50.5	54.9	45.98	46.57	50.5	47.8
178617	51.68	54.8	52	52.42	50.81	52.2	50.7
207141	88.56	89	90.96	88.07	84.76	89.29	85.32
223978	64.29	64.5	65.9	65.43	63.91	64.5	64.1
253120	79.64	76	78.5	79.06	79.94	79.6	80.81
260111	83.47	76	75.2	76.45	71.56	80	77.9
269629	89.95	88.11	87.69	85.87	86.57	87.01	89.62
285392	78.81	73.5	78.7	74.66	78.77	78.8	77
289927	63.99	65	60.3	57.25	54.1	68.3	63.5
302740	82.41	86	75.6	72.37	76.68	78.7	77.3
308840	79.02	70.5	75.5	68.95	73.64	77.2	74.5
311999	73.66	67.5	67.7	66.49	66.93	69.9	68.8
314036	89.77	91.5	91.29	90.81	92.57	94.58	94.14
316121	66.95	66	65.8	61.45	62.51	61.1	63.8
317648	68.05	71.5	67.5	62.81	66.49	67.2	66.9
193834(Lt)	76.2	73	77.4	70.29	73.57	75.4	75.2
193834(Rt)	62.66	67.5	69.4	62.73	59.38	72.5	73.9
21690*	57.7	55.32	57.8	60.67	56.6	61.8	56
26129(Lt)	65.16	64.5	64.9	59	57.07	59.5	56.7
26129(Rt)	80.19	77	77.3	76.01	76.21	77.9	76.2
31128(Lt)	58.97	60	60.1	50.98	56.2	60.8	56.1
31128(Rt)	58.68	62.5	64.1	58.43	60.38	66.9	65.1
317953*	27.5	31.5	22.4	26.99	27.96	31.2	28.6
48629(Lt)	75.76	74.5	78.1	71.3	68.54	80.08	76.1
48629(Rt)	54.59	50.5	52.8	49.54	48.21	63.3	54
85805(I)	45.19	50	53.3	48.6	52.11	54.8	47.7
85805(II)	60.5	62	64.5	60.81	63.35	69.2	63.1
90704(Lt)	94.43	90.5	93.12	90.87	92.51	95.94	96.1
90704(Rt)	52.46	55	51.9	50.39	49.66	55.7	52.4
93253(Lt)	49.27	58	57.7	42.69	47.94	54	50.2
93253(Rt)	59.98	60.5	62.4	51.95	55.88	65.1	59.3
99124(Lt)	64.79	66.5	63.3	58.62	61.28	63.1	59.6
99124(Rt)	65.7	68.5	62	64.87	64.19	66	66.7

Validity Measurement Data (Elbow).

MRN	Ave PACS	UG	No Markers	Parallel	Mid-Point	3 Markers	4 Markers
11721	67.225	57	68.95	63.12	57.37	70.7	70.58
75885	84.21	83.5	93.05	88.15	75.75	94.89	93.64
85274	103.51	98	93.42	89.27	84.83	101.29	98.6
93422	106.51	100	99.8	98.32	90.39	100.39	97.38
97679	88.9	98	92.04	84.66	74.53	86.23	82.19
108578	108.36	100.5	104.29	98.79	93.54	105.08	105.27
109479	86.425	85	89.96	85.29	76.98	95.24	92.58
112634	81.5	86.5	89.87	84.12	80.12	96.07	94.82
128001	83.88	75	82.45	77.91	70.02	80.2	80.89
128661	89.81	90	86.84	79.74	71.53	85.36	85.85
133213	88.035	84.5	92.59	83.35	73.17	96.41	96.69
133674	82.855	74.5	74.96	74.32	67.83	77.77	80.1
146004	112.665	102.5	111.49	97.77	91.86	114.58	114.2
154005	84.12	82	89.72	80.12	72.12	94.73	90.21
164905	101.655	94.5	99.51	94.79	87.17	103.64	103.79
212342	78.54	90	98.76	89.26	82.89	91.78	90.6
234289	97.655	95	95.95	90.52	86.63	99.33	97.09
253629	88.965	85	95.45	84.84	75.66	95.36	90.46
260835	88.945	82.5	86.36	80.89	74.59	91.56	89.53
267850	89.34	81	84.54	81.35	75.19	83.57	85.12
276750	80.325	77	82.61	75.01	65.55	86.18	82.64
283913	76.02	74	71.07	70.79	64.97	89.93	88.73
306088	87.73	83	87.85	82.28	74.23	88.28	91.39
306088	87.77	87	88.19	78.49	68.62	84.47	83.55
306088	82.475	86	93.77	84.83	76.63	93.04	91.96
309688	100.62	94.5	89.49	95.6	82.12	94.18	96.37
310362	84.27	75.5	83.02	82.86	71.46	83.15	83.97
329467	86.115	87	91.65	87.22	81.43	93.26	91.93
330180	78.94	86	95.71	87.66	83.69	100.34	99.66
331072	63.245	51	65.22	58.52	48.61	69.02	64.55
343488	95.7	96	101.77	93.94	88.77	105.42	106.44
348191	86.55	92.5	92.44	87.64	76.57	95.61	91.55
348191	88.875	100	86.4	79.9	71.17	86.66	83.23
348191	88.235	101	85.28	86.03	77.1	80.43	81.29
348395	89.64	79	86.39	77.43	75.05	88.75	88.53
350047	97.89	92	97.54	94.13	86.59	103.42	102.89
353331	87.095	76	75.78	75.53	66.81	79.81	80.83
353614	98.835	93	100.99	92.58	84.54	104.8	101.45
354824	88.375	89.5	89.6	88.1	81.36	102.27	100
354824	79.62	108.5	81.65	78.37	71.28	86.22	84.09
354824	78.68	91.5	93.22	87.34	81.51	95.45	94.34
356215	83.68	85	93.12	85.12	75.78	95.46	92.71
356218	85.67	86.5	96.09	90.09	83.97	89.98	88.89
358299	112.21	100	103.73	98.49	93.47	104.71	104.71
360746	100.3	88.5	95.96	91.55	83.26	97.33	98.52
362092	75.78	69	85.35	72.8	64.99	84.18	82.63
362803	100.18	92.5	99.36	92.66	82.03	98.5	97.7
366296	88.55	78	84.03	78.9	68.73	83.64	83.63
366635	107.565	94	99.14	92.56	84.02	95.69	94.51
369377	80.665	86	96.4	87.74	81.11	95.67	89.3

Appendix I

Information sheet.

This information sheet is to be read by potential subjects for the study into measuring elbow flexion using the Uillinn Method©.

Title of Study.

An Investigation into the Clinical Application of the Uillinn Method© of Assessing Joint Angles.

Introduction.

Thank you for taking the time to read this information sheet. You are being asked to participate in a study concerned with the measurement of joint range movement. The aim of the study is to investigate a new method of measuring the angle of elbow flexion; i.e. the amount the arm is flexed at one particular moment. This new method is called the Uillinn Method© and it uses digital photography and a computer to establish the amount of degrees to which the joint (in this case the elbow) is bent.

Your participation in the study will involve having 2 photographs taken of your elbow, 1 of these photographs will require you to have 4 small adhesive markers placed on your arm. Both of the photographs will be taken immediately after your X-ray. The procedure will not add more than 1 or 2 minutes onto the length of time you will be asked to stay in the X-ray department. This is a once off examination and you will not be asked to participate further into the study.

Procedure.

The criteria for participation in this study are;

- That you are going for a routine X-ray of your elbow.
- You are of consenting age.
- You do not have hypersensitive skin or any lesions on your arm.
- That your elbow does not cause you so much pain that you can't hold it in one position for more than 2 minutes.

If you decide to participate in this study you will be asked to have the top of your shoulder expose during the examination. Appropriate attire will be provided for you. Then 4 small adhesive markers will be placed on your arm, 1 at your wrist, 2 at your elbow and 1 at your shoulder.

You will then enter the X-ray room. The radiographer will ask you to sit beside the X-ray table with your arm on the table. The radiographer will position you correctly and you will then have your X-ray taken. Immediately afterward the first digital photograph will then be taken. While holding your elbow in the same *exact* position, the markers on your arm will be removed and the final photograph will be taken. A physiotherapist will then quickly measure the angle that your elbow is bent with a universal goniometer (a small plastic tool). That will concluded your involvement in the study. The radiologist will then continue their assessment, if anything further is required.

Benefits.

Information regarding joint range of movement is useful to every health professional involved in assessing the musculoskeletal system. This type of information is objective data from which much useful information can be gained. There is some room for improvement regarding current joint measurement techniques and it is possible that digital methods (the Uillinn Method©) of measurement may improve objectivity in this area. It is hoped that this study will help to determine whether or not employing the Uillinn Method© improves competence of measuring joint angles.

Risks.

There are minimal risks posed from participation in this study. It is possible that some individuals may get a minor allergic reaction to the adhesive markers that will be placed on your arm.

Exclusion Criteria.

You can not be involved in this study if you;

are known to have hypersensitive skin,
have any skin lesions on your arm,
are under the age of consent,
are too painful to hold your elbow in one position for two minutes.

Confidentiality.

Your identity will remain confidential. Your name will not be published and will not be disclosed to anyone outside the hospital.

Compensation.

The Physiotherapist involved in this study, Colin Dunlevy, is qualified and is fully insured.

Voluntary Participation.

Participation in this study is completely voluntary. You may quit at any time. If you decide not to participate, or if you quit during the assessment, you will not be penalised and will not give up any benefits which you had before entering the study.

Stopping the Study.

You understand that the professionals involved can stop your participation in the study at any time if they feel you are not a suitable candidate.

Permission.

The trial has been approved by the hospital research ethics committee, the radiological staff and the orthopaedic surgical team.

Further Information.

You can get more information or answers to your questions about the study, your participation, and your rights from Colin Dunlevy (physiotherapist). He can be approached anytime during or after the study on (01) 414 3055.

Appendix J

Consent Form.

This consent form is to be completed by potential subjects for the study into measuring elbow flexion using the Uillinn Method©.

Title of Study.

An Investigation into the Clinical Application of the Uillinn Method© of Assessing Joint Angles.

The details of the above study and this consent form have been explained to me. The physiotherapist involved in the study has answered all my questions to my satisfaction. I believe I understand what will happen if I agree to be part of this study.

I have read or had explained to me the meaning of this consent form. I have had the opportunity to ask questions and all my questions have been answered to my satisfaction. I freely and voluntarily agree to be part of this research study, though without prejudice to my legal and ethical rights. I have received a copy of this agreement.

Participants Name :.....

Participants Signature:.....

Date :

Statement of investigator's responsibility: I have explained the nature, purpose, procedures, benefits and risk of this study. I have offered to answer any questions and fully answered such questions. I believe that the participant understands my explanation and has freely given informed consent.

Researchers

Signature:.....

Appendix K

Calculation of Control Data (X-rays) at the Knee			
MRN	**1st PACS**	**2nd PACS**	**Average**
13081	72.055	72.705	72.38
22429	86.305	89.595	87.95
42002	63.83	63.49	63.66
53808	55.005	56.955	55.98
63067	63.45	N/T	63.45
74318	54.3	N/T	54.3
84051	40.8	N/T	40.8
120923	66.35	N/T	66.35
146405	25.9	N/T	25.9
166412	53.205	54.555	53.88
178617	51.19	52.17	51.68
207141	88.56	N/T	88.56
223987	64.065	64.515	64.29
253120	79.63	79.65	79.64
260111	81.955	84.985	83.47
269629	90.195	89.705	89.95
285392	79.15	78.47	78.81
289927	63.99	N/T	63.99
302740	82.21	82.61	82.41
308840	78.76	79.28	79.02
311999	74.08	73.24	73.66
314036	89.77	N/T	89.77
316121	67.025	66.875	66.95
317648	69.52	66.58	68.05
193834(Lt)	76.2	N/T	76.2
193834(Rt)	62.66	N/T	62.66
21690	57.7	N/T	57.7
26129(Lt)	65.235	65.095	65.16
26129(Rt)	79.56	80.82	80.19
31128(Lt)	57.505	60.435	58.97
31128(Rt)	58.715	58.645	58.68
317953	27.5	N/T	27.5
48629(Lt)	75.76	N/T	75.76
48629(Rt)	54.59	N/T	54.59
85805(I)	45.19	N/T	45.19
85805(II)	60.5	N/T	60.5
90704(Lt)	94.28	94.58	94.43
90704(Rt)	52.235	52.685	52.46
93253(Lt)	49.27	N/T	49.27
93253(Rt)	59.98	N/T	59.98
99124(Lt)	64.79	N/T	64.79
99124(Rt)	65.7	N/T	65.7
N/T = Not Tested			

Calculation of Control Data (X-rays) at the Elbow			
MRN	**1st PACS**	**2nd PACS**	**Average**
11721	66.31	68.14	67.225
75885	84.9	83.52	84.21
85274	103.7	103.32	103.51
93422	107.48	105.54	106.51
97679	89.17	88.63	88.9
108578	108.17	108.55	108.36
109479	87.7	85.15	86.425
112634	81.355	81.645	81.5
128001	84.81	82.95	83.88
128661	89.98	89.64	89.81
133213	89.01	87.06	88.035
133674	82.93	82.78	82.855
146004	112.67	112.66	112.665
154005	84.87	83.37	84.12
164905	100.97	102.34	101.655
212342	77.52	79.56	78.54
234289	98.05	97.26	97.655
253629	89.28	88.65	88.965
260835	89.89	88	88.945
267850	89.65	89.03	89.34
276750	80.22	80.43	80.325
283913	76.07	75.97	76.02
306088	88.49	86.97	87.73
306088	87.9	87.64	87.77
306088	82.95	82	82.475
309688	101.09	100.15	100.62
310362	85.08	83.46	84.27
329467	85.75	86.48	86.115
330180	79.6	78.28	78.94
331072	62.87	63.62	63.245
343488	95.20	96.21	95.7
348191	85.485	87.615	86.55
348191	87.495	90.255	88.875
348191	87.71	88.76	88.235
348395	89.55	89.73	89.64
350047	97.96	97.82	97.89
353331	86.48	87.71	87.095
353614	99.75	97.92	98.835
354824	87.56	89.19	88.375
354824	79.29	79.95	79.62
354824	78.64	78.72	78.68
356215	83.8	83.56	83.68
356218	85.89	85.45	85.67
358299	112.38	112.04	112.21
360746	100.23	100.37	100.3
362092	75.45	76.11	75.78
362803	100.89	99.47	100.18
366296	88.67	88.43	88.55
366635	107.16	107.97	107.565
369377	81.36	79.97	80.665

Inter-rater Reliability of the PACS system.

Knees (n=7)		
Subject	IK	CGD
1	61.29	63.615
2	63.41	65.825
3	84.88	84.5
4	87.91	89.655
5	79.78	84.975
6	68.83	73.045
7	55	54.845

Elbows (n=7)		
MRN	IK	CGD
1	79.15	79.58
2	88.95	88.445
3	83.59	81.015
4	74.55	73.895
5	86.92	88.9
6	86.76	86.55
7	71.16	70.59

Appendix L

Significance Testing Between PACS (X-ray) data and each of the Experimental Measurements Systems – The Knee

Knee Measurements

t-Test: Paired Two Sample for Means (p=0.05)

	PACS	*UG*	
Mean	65.58619	65.82667	
Variance	247.3667	209.6965	
Observations	42	42	
Pearson Correlation	0.970074		
df	41		
t Stat	-0.39992		
P(T<=t) one-tail	0.345647		
t Critical one-tail	1.682879		
P(T<=t) two-tail	0.691295	**NS**	
t Critical two-tail	2.019542		

Knee Measurements

t-Test: Paired Two Sample for Means (p=0.05)

	PACS	*No Markers*	
Mean	65.58619	65.41548	
Variance	247.3667	226.8563	
Observations	42	42	
Pearson Correlation	0.969141		
df	41		
t Stat	0.285049		
P(T<=t) one-tail	0.388521		
t Critical one-tail	1.682879		
P(T<=t) two-tail	0.777042	**NS**	
t Critical two-tail	2.019542		

Knee Measurements

t-Test: Paired Two Sample for Means (p=0.05)

	PACS	*Parallel*	
Mean	65.58619	62.00262	
Variance	247.3667	236.7543	
Observations	42	42	
Pearson Correlation	0.972867		
df	41		
t Stat	6.380488		
P(T<=t) one-tail	6.22E-08		
t Critical one-tail	1.682879		
P(T<=t) two-tail	1.24E-07	**Significant**	
t Critical two-tail	2.019542		

Knee Measurements

t-Test: Paired Two Sample for Means (p=0.05)

	PACS	*Mid-Point*	
Mean	65.58619	62.82857	
Variance	247.3667	229.4003	
Observations	42	42	
Pearson Correlation	0.972357		
df	41		
t Stat	4.862451		
P(T<=t) one-tail	8.72E-06		
t Critical one-tail	1.682879		
P(T<=t) two-tail	1.74E-05	**Significant**	
t Critical two-tail	2.019542		

Knee Measurements

t-Test: Paired Two Sample for Means (p=0.05)

	PACS	*3 Markers*	
Mean	65.58619	67.3431	
Variance	247.3667	223.9278	
Observations	42	42	
Pearson Correlation	0.952879		
df	41		
t Stat	-2.38645		
P(T<=t) one-tail	0.010855		
t Critical one-tail	1.682879		
P(T<=t) two-tail	0.02171	**Significant**	
t Critical two-tail	2.019542		

Knee Measurements

t-Test: Paired Two Sample for Means (p=0.05)

	PACS	*4 Markers*	
Mean	65.58619	64.69119	
Variance	247.3667	267.5228	
Observations	42	42	
Pearson Correlation	0.967855		
df	41		
t Stat	1.409546		
P(T<=t) one-tail	0.083107		
t Critical one-tail	1.682879		
P(T<=t) two-tail	0.166214	**NS**	
t Critical two-tail	2.019542		

NS = No Significant Difference Significant = Significant Difference

Significance Testing Between PACS (X-ray) data and each of the Experimental Measurements Systems – The Elbow

Elbow Measurements

t-Test: Paired Two Sample for Means (p=0.05)

	PACS	*UG*	
Mean	89.0942	86.89	
Variance	112.4739	120.8703	
Observations	50	50	
Pearson Correlation	0.702764		
df	49		
t Stat	1.870056		
P(T<=t) one-tail	0.033728		
t Critical one-tail	1.676551		
P(T<=t) two-tail	0.067456	**NS**	
t Critical two-tail	2.009574		

Elbow Measurements

t-Test: Paired Two Sample for Means (p=0.05)

	PACS	*No Markers*	
Mean	89.0942	90.4558	
Variance	112.4739	84.1319	
Observations	50	50	
Pearson Correlation	0.735296		
df	49		
t Stat	-1.31567		
P(T<=t) one-tail	0.097205		
t Critical one-tail	1.676551		
P(T<=t) two-tail	0.194409	**NS**	
t Critical two-tail	2.009574		

Elbow Measurements

t-Test: Paired Two Sample for Means (p=0.05)

	PACS	*Parallel*	
Mean	89.0942	84.8146	
Variance	112.4739	74.13672	
Observations	50	50	
Pearson Correlation	0.814065		
df	49		
t Stat	4.913058		
P(T<=t) one-tail	5.22E-06		
t Critical one-tail	1.676551		
P(T<=t) two-tail	1.04E-05	**Significant**	
t Critical two-tail	2.009574		

Elbow Measurements

t-Test: Paired Two Sample for Means (p=0.05)

	PACS	*Mid-Point*	
Mean	89.0942	77.1516	
Variance	112.4739	82.17489	
Observations	50	50	
Pearson Correlation	0.789014		
df	49		
t Stat	12.887		
P(T<=t) one-tail	1.17E-17		
t Critical one-tail	1.676551		
P(T<=t) two-tail	2.34E-17	**Significant**	
t Critical two-tail	2.009574		

Elbow Measurements

t-Test: Paired Two Sample for Means (p=0.05)

	PACS	*3 Markers*	
Mean	89.0942	92.4004	
Variance	112.4739	85.60532	
Observations	50	50	
Pearson Correlation	0.703558		
df	49		
t Stat	-3.01796		
P(T<=t) one-tail	0.002015		
t Critical one-tail	1.676551		
P(T<=t) two-tail	0.00403	**Significant**	
t Critical two-tail	2.009574		

Elbow Measurements

t-Test: Paired Two Sample for Means (p=0.05)

	PACS	*4 Markers*	
Mean	89.0942	91.2314	
Variance	112.4739	85.49989	
Observations	50	50	
Pearson Correlation	0.748831		
df	49		
t Stat	-2.11392		
P(T<=t) one-tail	0.019818		
t Critical one-tail	1.676551		
P(T<=t) two-tail	0.039636	**Significant**	
t Critical two-tail	2.009574		

NS = No Significant Difference Significant = Significant Difference

Significance Testing of the Differences between PACS (X-ray) data and each of the Experimental Measurements Systems, in different ranges – The Knee

Knee Measurements

t-Test: Two-Sample Assuming Equal Variances

PACS - UG	Outer Range (25° to 59°)	Mid Range (60° to 95°)
Mean	-1.516	0.468148
Variance	13.25187	15.36215
Observations	15	27
Pooled Variance	14.61705	
df	40	
t Stat	-1.61156	
P(T<=t) one-tail	0.057461	
t Critical one-tail	1.683852	
P(T<=t) two-tail	0.114921	**NS**
t Critical two-tail	2.021075	(p=0.05)

Knee Measurements

t-Test: Two-Sample Assuming Equal Variances

PACS - No Markers	Outer Range (25° to 59°)	Mid Range (60° to 95°)
Mean	-1.442	1.066667
Variance	20.61496	10.32083
Observations	15	27
Pooled Variance	13.92378	
df	40	
t Stat	-2.0877	
P(T<=t) one-tail	0.021619	
t Critical one-tail	1.683852	
P(T<=t) two-tail	0.043239	**Significant**
t Critical two-tail	2.021075	(p=0.05)

Knee Measurements

t-Test: Two-Sample Assuming Equal Variances

PACS - Parallel	Outer Range (25° to 59°)	Mid Range (60° to 95°)
Mean	2.862667	3.984074
Variance	17.48298	11.01189
Observations	15	27
Pooled Variance	13.27677	
df	40	
t Stat	-0.9557	
P(T<=t) one-tail	0.172483	
t Critical one-tail	1.683852	
P(T<=t) two-tail	0.344966	**NS**
t Critical two-tail	2.021075	(p=0.05)

Knee Measurements

t-Test: Two-Sample Assuming Equal Variances

PACS - Mid-Point	Outer Range (25° to 59°)	Mid Range (60° to 95°)
Mean	1.531333	3.438889
Variance	13.64628	12.60433
Observations	15	27
Pooled Variance	12.96902	
df	40	
t Stat	-1.64485	
P(T<=t) one-tail	0.053919	
t Critical one-tail	1.683852	
P(T<=t) two-tail	0.107838	**NS**
t Critical two-tail	2.021075	(p=0.05)

Knee Measurements

t-Test: Two-Sample Assuming Equal Variances

PACS - 3 Markers	Outer Range (25° to 59°)	Mid Range (60° to 95°)
Mean	-3.962	-0.53185
Variance	30.65716	15.02498
Observations	15	27
Pooled Variance	20.49625	
df	40	
t Stat	-2.35277	
P(T<=t) one-tail	0.01182	
t Critical one-tail	1.683852	
P(T<=t) two-tail	0.02364	**Significant**
t Critical two-tail	2.021075	(p=0.05)

Knee Measurements

t-Test: Two-Sample Assuming Equal Variances

PACS - 4 Markers	Outer Range (25° to 59°)	Mid Range (60° to 95°)
Mean	1.044667	0.811852
Variance	20.59071	15.59479
Observations	15	27
Pooled Variance	17.34336	
df	40	
t Stat	0.173599	
P(T<=t) one-tail	0.431528	
t Critical one-tail	1.683852	
P(T<=t) two-tail	0.863056	**NS**
t Critical two-tail	2.021075	(p=0.05)

NS = No Significant Difference Significant = Significant Difference

Significance Testing of the Differences between PACS (X-ray) data and each of the Experimental Measurements Systems, in different ranges – The Elbow

Elbow Measurements

t-Test: Two-Sample Assuming Equal Variances

PACS - UG	Mid Range (60° to 89°)	Inner Range (90° to 115°)
Mean	0.209861	7.3325
Variance	77.46884	13.9209
Observations	36	14
Pooled Variance	60.25794	
df	48	
t Stat	-2.91316	
P(T<=t) one-tail	0.002709	
t Critical one-tail	1.677224	
P(T<=t) two-tail	0.005417	**Significant**
t Critical two-tail	2.010634	(p=0.05)

Elbow Measurements

t-Test: Two-Sample Assuming Equal Variances

PACS - No Markers	Mid Range (60° to 89°)	Inner Range (90° to 115°)
Mean	-3.31375	3.658214
Variance	52.03638	24.0629
Observations	36	14
Pooled Variance	44.46023	
df	48	
t Stat	-3.31971	
P(T<=t) one-tail	0.000863	
t Critical one-tail	1.677224	
P(T<=t) two-tail	0.001726	**Significant**
t Critical two-tail	2.010634	(p=0.05)

Elbow Measurements

t-Test: Two-Sample Assuming Equal Variances

PACS - Parallel	Mid Range (60° to 89°)	Inner Range (90° to 115°)
Mean	2.535972	8.763214
Variance	35.26094	18.02178
Observations	36	14
Pooled Variance	30.58471	
df	48	
t Stat	-3.57498	
P(T<=t) one-tail	0.000405	
t Critical one-tail	1.677224	
P(T<=t) two-tail	0.000811	**Significant**
t Critical two-tail	2.010634	(p=0.05)

Elbow Measurements

t-Test: Two-Sample Assuming Equal Variances

PACS - Mid-Point	Mid Range (60° to 89°)	Inner Range (90° to 115°)
Mean	10.35264	16.03107
Variance	43.90526	18.64309
Observations	36	14
Pooled Variance	37.06342	
df	48	
t Stat	-2.96132	
P(T<=t) one-tail	0.002375	
t Critical one-tail	1.677224	
P(T<=t) two-tail	0.004751	**Significant**
t Critical two-tail	2.010634	(p=0.05)

Elbow Measurements

t-Test: Two-Sample Assuming Equal Variances

PACS - 3 Markers	Mid Range (60° to 89°)	Inner Range (90° to 115°)
Mean	-5.01681	1.0925
Variance	60.17868	35.23058
Observations	36	14
Pooled Variance	53.4192	
df	48	
t Stat	-2.65383	
P(T<=t) one-tail	0.00538	
t Critical one-tail	1.677224	
P(T<=t) two-tail	0.010761	**Significant**
t Critical two-tail	2.010634	(p=0.05)

Elbow Measurements

t-Test: Two-Sample Assuming Equal Variances

PACS - 4 Markers	Mid Range (60° to 89°)	Inner Range (90° to 115°)
Mean	-3.65569	1.7675
Variance	49.31371	37.06323
Observations	36	14
Pooled Variance	45.99588	
df	48	
t Stat	-2.53879	
P(T<=t) one-tail	0.00721	
t Critical one-tail	1.677224	
P(T<=t) two-tail	0.014421	**Significant**
t Critical two-tail	2.010634	(p=0.05)

NS = No Significant Difference Significant = Significant Difference

Appendix M

Introduction to the Uillinn Method©.

In all areas of health care management a scientific approach by clinicians is essential in order to achieve optimal health outcomes. A scientific approach demands valid and reliable measurement tools that can deliver objective data. This implies that clinicians require objective measurement tools to assess human health, function and performance. Physiotherapy is no exception in this need for dependable measurement tools.

One very important type of measurement for physiotherapy is the calculation of joint range of motion (ROM), i.e. osteokinematic measurement. The *American Academy of Orthopaedic Surgeons* (AAOS) (1994)[1] state that joint ROM is directly related to human function and therefore it is an essential component of a patient assessment.

Universal Goniometry (UG) is probably the most commonly used method of objectively assessing available joint ROM. However, research has shown that the validity and reliability of this device is limited. The Uillinn Method© (UM) was devised as an alternative approach to joint ROM measurement. It was hoped that the UM may improve upon the performance of the UG and possibly offer some useful advantages over it and other measurement methods.

The Uillinn Method©.

The UM uses personal computer (PC) software to create a 'virtual' on-screen goniometer. The system has been designed to report angular changes of as low as 0.01°. While, it has been shown that the UM calculations (non-clinical) will remain within 0.5° of perfect accuracy for 95% of measurements. However, it should be noted that certain photographic procedures must be observed if this level of accuracy is to be maintained.

Research into the potential accuracy, clinical validity and clinical reliability of the UM is on going. To date clinical validity and reliability has been established for measurements of sagittal motion at the knee[2]. The findings of that research has

---

[1] *American Academy of Orthopaedic Surgeons. (1994) The Clinical Measurement of Joint Motion. 1st Edition Edited by Greene, W. B., Heckman, J. D.*

[2] *Dunlevy et al (2002) Unpublished Thesis.*

established that the UM agrees with X-ray measurements by 3.6° for 95% of measurements (95% LOA), with a high Pearson product moment correlation coefficient (PPMCC) (r=0.99). The investigation also established that the universal goniometer (UG) only achieved a 95% LOA of 7.6° (r=0.97) with X-ray data. These findings indicate that the validity of UM improves upon conventional UG by as much as 4°. Research into the validity for the UM for elbow measurements is not yet complete but initial results indicate outcomes similar to those seen at the knee.

The intra-rater reliability for the UM for assessment of the knee was shown to have a 95% LOA of 4.5° (r=0.99), while inter-rater reliability achieved a 95% LOA of 6.5° (ICC=0.99). It is difficult to compare these results to the performance of the UG, as reliability tests for the UG using the 95% limits of agreement (95% LOA) analysis have not been carried out.

The Photographic component of the Uillinn Method©.

During the actual taking of photographs for the UM, it is vital that certain procedures are observed. Following these 3 simple guidelines will ensure optimal results;

1) The camera lens should be centred on the approximate centre of the angle that is to be assessed. Specifically, the viewfinder crosshairs should be placed just below the approximate apex of the joint angle. This will help to avoid parallax error, the pillow effect or other causes of photographic misrepresentation.

2) An effort should be made to keep the incident angle of the lens as low as possible, to achieve optimal results this angle should be no more than 5°. To do this the photographer should stand in front of, and perpendicular to, the angle that is to be photographed. The photographer should check the alignment of the target angle and ensure that the camera is held approximately parallel to it. If the photographer is not certain of the orientation of the target angle with regard to their point of view, subsequent assessment of the image with the UM can not be trusted to be less than 1°.

4) The limb must be exposed adequately during photography.

The Computer based component of the Uillinn Method©.

Once the photographic component of the UM is complete, the image can be transferred to the PC using the Smart card reader (see Fig 1). Once the photographic file is transferred to the PC and placed in the relevant therapists folder. The file should then be renamed to include the patient's initials and the date on which the photograph was taken. If more than one photograph of the same patient is taken on the same day, then additional roman numerals should be used to identify the file. The Uillinn software can be used to display the photographic image.

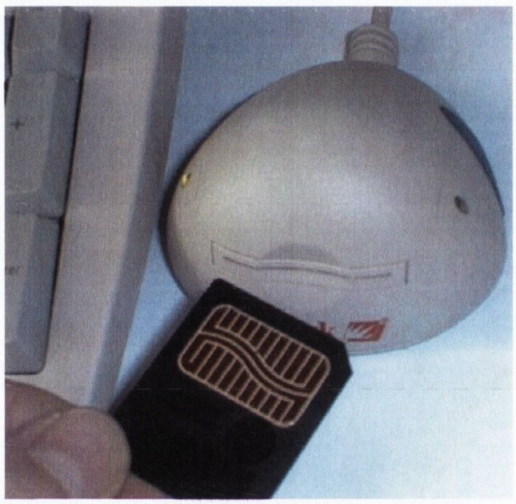


Fig 1. The Smart Media Disk and Smart Media Disk Drive.

The image can be zoomed in upon and positioned so the relevant limb segment fills the screen. Once the picture is properly positioned in the screen, the measurement arms can be manipulated to allow calculation of the joint angle. Each measurement arm should be placed so it is approximately parallel to the longitudinal axis of the limb segment above and below the joint. For optimal results the measurement arm should be made longer than the limb segment (see fig 2).

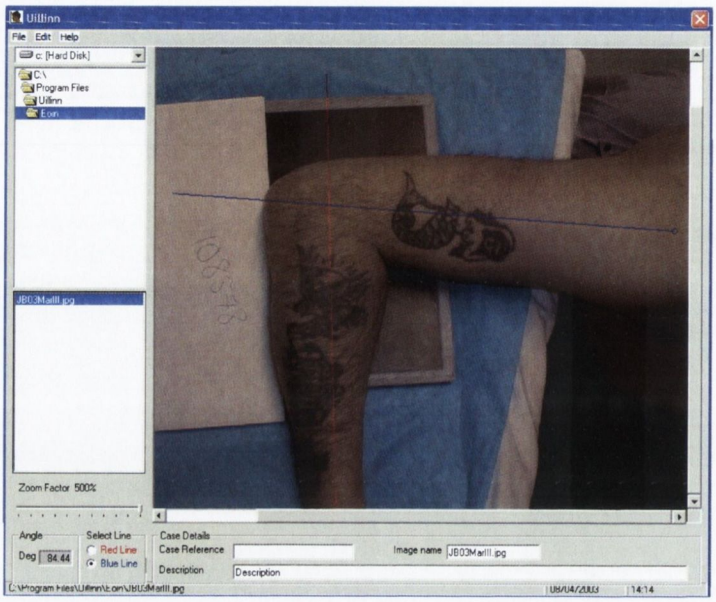


Fig 2. Screen Shot of the Uillinn Method©.

Appendix N

How to use the Uillinn Method©.

Introduction.

There are 3 basic steps that must be completed when using the Uillinn Method©, these are;

1. *The Photographic Procedure.*
2. *Transferring and Storing the photographic file.*
3. *Using the Uillinn Software*

.

The Photographic Procedure of the Uillinn Method©.

During the photographic component of the Uillinn Method© (UM), it is vital that certain procedures are observed. Following these 3 simple guidelines will minimise measurement error originating from the photograph;

1) The limb must be exposed adequately during photography.
2) The camera lens should be centred on the approximate centre of the angle that is to be assessed. Specifically, the viewfinder crosshairs should be placed just inside the apex of the joint angle.
3) The camera should be held perpendicular to the target joint, 5° of rotation between the camera and the joint will not add any error to the measurement, however, if 10° of rotation is present during the photographic procedure an additional 1.5° of error will be present in the measurement.

To control this adequately, the photographer should check the camera's alignment with regard to the target angle and ensure that the camera is held approximately perpendicular to it.

Transferring and Storing the Photograph to the Laptop.

When the photograph is taken the image must be transferred to the laptop and stored appropriately in each individual therapists folder. To complete this task;

8) Remove the Smart Media card from the camera.
9) Place the Smart Media card face down into the 'SanDisk' drive.
10) Double click on the "Shortcut to SanDisk ImageMate (E)" (see Fig 1.).

11) Then double click on the "IMOLYM" folder.

12) Right click on your photograph and choose the RENAME option (2nd from bottom).

13) Type the initials of your patient followed by the date. If you have taken more than one photograph of the same patient on that day indicate this with roman numerals. When this is done hit return.

Example;

John Murphy photographed on the *14 April 2003* = JM14Apr03

2nd photograph of *John Murphy* same day = JM14Apr03ii

14) Drag this file up to YOUR folder on the computers Desktop.

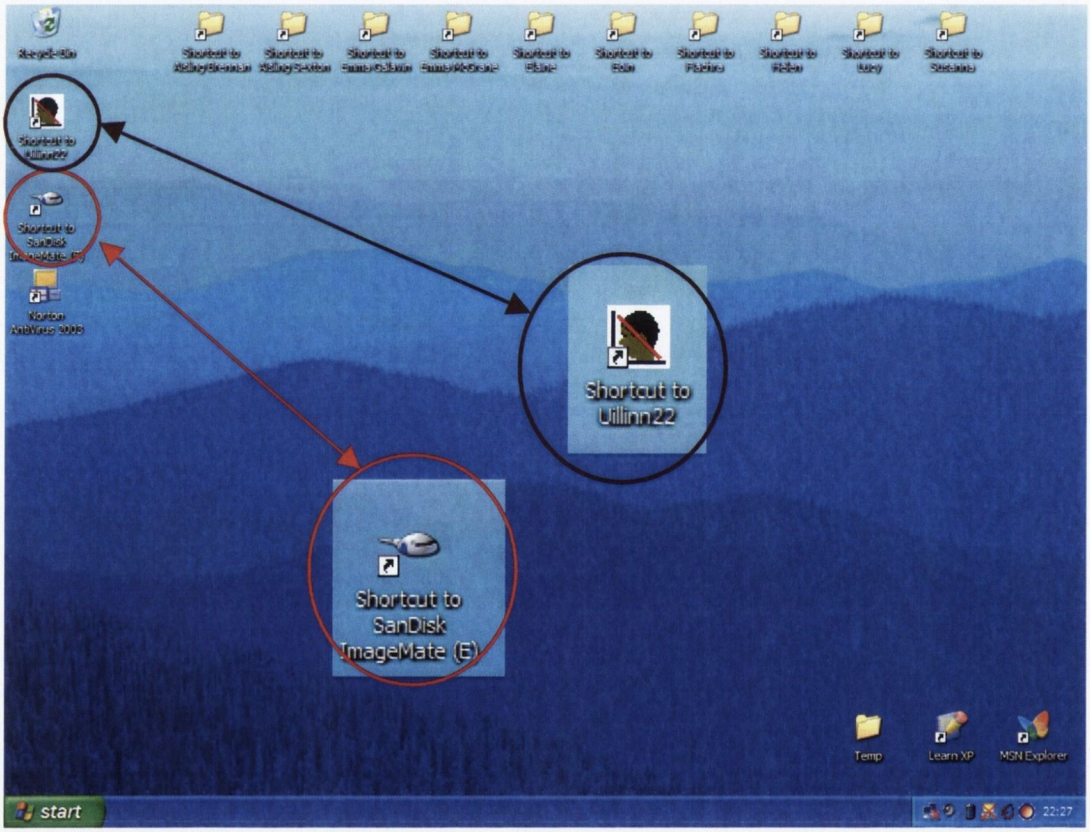
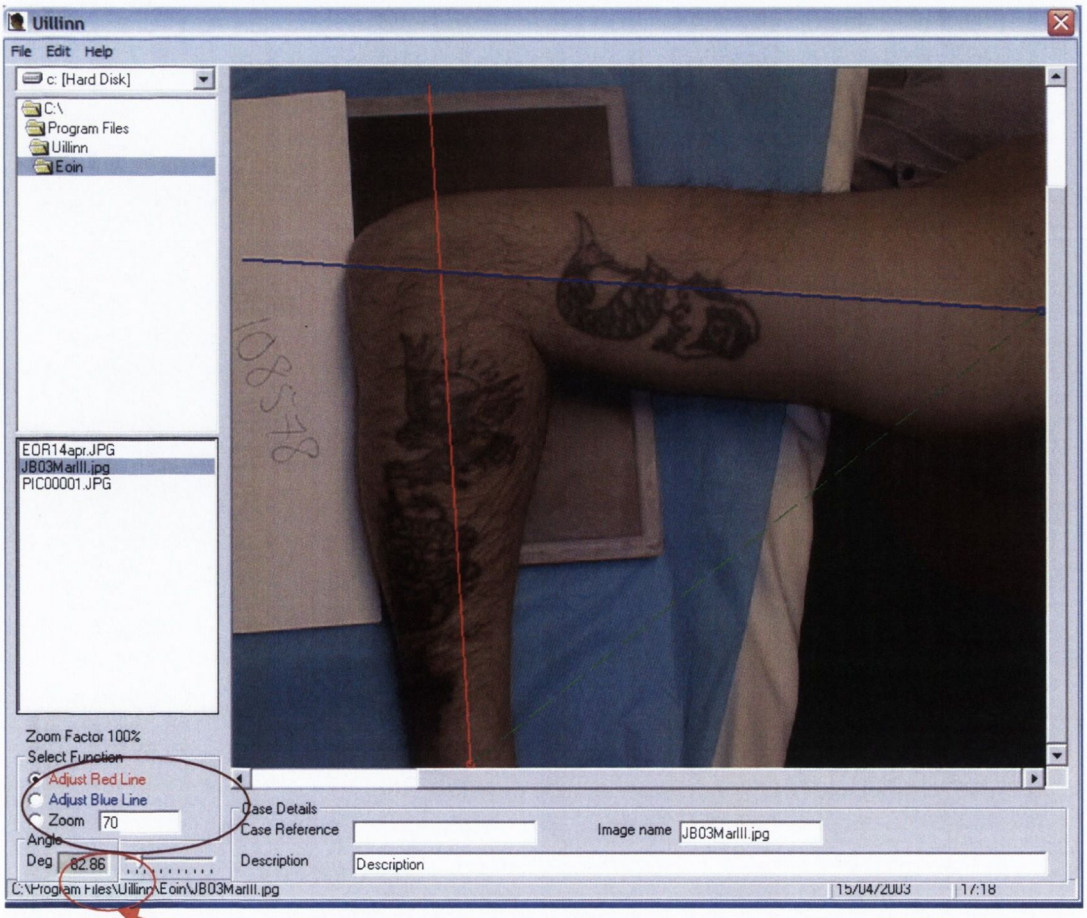


Fig. 1. Shortcuts on the Desktop.

Using the Uillinn Software.

When the photograph has been renamed and transferred to the laptop the Uillinn software can be used to calculate the joint angle.

9. Double click on the Shortcut to Uillinn 2.3 (fig. 1).
10. Double click on Your folder, within the Uillinn window.
11. Double click on the appropriate file (preview available by single click).
12. Zoom in and Adjust image as appropriate (see Fig 2).
13. The RED measurement line will automatically be active. Use the RIGHT and LEFT mouse buttons to manipulate this line. LEFT for the Apex of the angle, RIGHT for the base of the angle (Circular end). The line should be parallel to the longitudinal axis of the limb segment (see Fig. 3).
14. Click on the "Adjust Blue Line" button at the bottom left hand corner of the window.
15. Now manipulate this line so that it is parallel to the longitudinal axis of the other limb segment.
16. The dotted Green line will reflect the base of the triangle that you've just drawn, the apex is the joint angle. Read the angle from the "Angle Deg" box.



Results Box in "o"

Fig 2. Screenshot of the Uillinn Method©.



Fig 3. How to Manipulate the Measurement Line of the Uillinn Method© Software

Appendix O

Uillinn Method© Users Scoring Sheet Date / / 03

Therapist's Name..... Assessor's Name .....

Theory (Questions to the Therapists)

1. Do they understand the concept of the Uillinn Method© as a virtual goniometer? Y / N
2. Do they report they require further training in order to use the system? Y / N
3. Do they know how the camera and smart card system works? Y / N
4. Can they transfer and rename the photographic file (camera to PC) Y / N
5. Do they know how to display the photograph using the Uillinn software? Y / N
6. Do they understand the concept of the 'Parallel Approach' to the UM? Y / N
7. Do they know how to position the measurement arms of the Uillinn software? Y / N
8. Do you feel confident using the system? Y / N

Practical (Therapist to demonstrate)

9. Was the camera positioned appropriately during photography? Y / N
10. Was the photograph transferred, renamed and stored correctly? Y / N
11. Was the photograph aligned correctly (centred on the joint)? Y / N
12. Were the measurement arms of the Uillinn Method© used correctly? Y / N
13. Was the angular data read and recorded correctly? Y / N
14. Assessing the Elbow, what was their first angle reading (blind) .....

    What was their second angle reading (same photograph (blind)) .....

    What was the Assessors angle reading (same photograph (blind)) .....

    Was the intra-rater data within 4.5° of a blind repeated measurement? Y / N

    Was the inter-rater data within 9° of the Assessors results Y / N
15. Assessing the Knee, what was their first angle reading (blind) .....

    What was their second angle reading (same photograph (blind)) .....

    What was the Assessors angle reading (same photograph (blind)) .....

    Was the intra-rater data within 3.5° of a blind repeated measurement? Y / N

    Was the inter-rater data within 9° of the Assessors results Y / N

Appendix P

Questionnaire: Usage and Application of the Uillinn Method©.

THIS QUESTIONNAIRE IS TO BE COMPLETED BY PHYSIOTHERAPISTS WHO HAVE BEEN TRAINED IN THE USE OF THE UILLINN METHOD© AND HAVE HAD THE SYSTEM AVAILABLE TO THEM FOR AT LEAST 3 MONTHS

(IF YOU NEED MORE SPACE TO ANSWER YOUR QUESTIONS PLEASE USE THE REVERSE SIDE OF THIS SHEET OR USE ADDITIONAL PAPER.)

---

1. What area of the physiotherapy department do you work in?

   Out-patient department. ☐

   Elective Orthopaedic service. ☐

   Paediatric out-patients. ☐

   Hydrotherapy service. ☐

2. Was the training you received in the use of the Uillinn Method© adequate?

   Y / N

   If you answered NO, how would you recommend that the training be improved?.....

   .....

   .....

   .....

   .....

3. At the time of your training, approximately how many times did you need to use the Uillinn Method© before you were confident to use it by yourself?

   .....

4. Do you require further training <u>now</u>, in order to effective use of the Uillinn Method©? Y / N

5. The Uillinn Method© was available to you for a 3 month period. Did you use the system during this time? Y / N

   If you answered NO, please state the main reasons why you did not use the system.....
   .....
   .....
   .....
   .....(Please skip to question number 23).

Please Continue to respond to the questions below only if you answered YES to question 5.

---

6. Approximately how often did you use the Uillinn Method© per appropriate patient?

   Once Three Five
   Twice Four More than Five

7. How did you position your patients during the ELBOW measurement procedure?
   .....
   .....
   .....
   .....
   .....

8. How did you position your patients during the KNEE measurement procedure?
   .....
   .....
   .....

.....

.....

9. Which joint did you think was easier to assess using the UM?

Elbow ☐

Knee ☐

No Difference

☐

10. Approximately, how long did it take to perform each measurement (from photography to angle calculation)? (< = less than ; > = greater than)

<1min >1min >2mins >3mins >4mins

>5mins

11. In your opinion was this too much time? Y / N

12. How much inconvenience or "hassle factor" was involved in using the Uillinn Method©?

None Minimal Moderate A lot

13. Was using the Uillinn Method© worth the effort? Y / N

14. Of all the photographs that you took, approximately what percentage did you subsequently analyse using the Uillinn Method© software?

<25% ~50% ~75% 100%

If less than 100%, please state the main reason why;

.....

.....

.....

.....

.....

15. Are there any aspects of the Uillinn Method© or it's procedure that are prone to introducing error? Y / N

Please clarify (include how to minimise error) .....

.....

.....

.....

.....

16. Over the last 3 months, was having the Uillinn Method© available beneficial to you in your practice? Y / N

Please clarify

.....

.....

.....

.....

.....

17. Are there any area of physical medicine or otherwise, that would benefit from the use of the Uillinn Method©? Y / N

If YES, please clarify.....

.....

.....

.....

.....

18. Would you recommend the adoption of the Uillinn Method© by physiotherapy departments globally? Y / N

19. Different aspects of the Assessment and Treatment Cycle are laid out below. Please rate which of these the Uillinn Method© offers most to:

(1= most beneficial, 8 = least beneficial, n/a= not beneficial)

Initial Assessment ☐

Analysis ☐

Treatment Planning ☐

Treatment Implementation ☐

Evaluation ☐

Transfer of Care/ Discharge ☐

Documentation ☐

Communication ☐

Other (please specify) ..... ☐

20. Was there any advantage in storing pictures for analysis or comparison at a later date? Y / N

Please clarify.....

.....

.....

.....

21. Which patient group (diagnosis) did you feel were best suited to being assessed by the Uillinn Method©?

.....

.....

.....

.....

.....

22. Which other joints besides the knee or elbow, if any, do you feel could be assessed by the Uillinn Method©?

.....

.....

.....

.....

.....

23. After you had been trained in the use of the Uillinn Method©, how did you find the operational procedure?

    Extremely Difficult Difficulty Easy

    Extremely Easy

24. Would you change the Uillinn Method© to make it more user friendly? Y / N If YES please specify

    .....

    .....

    .....

    .....

    .....

25. When you were using the Uillinn Method©, including the training period, did you experience any procedural problems with the system? Y / N (e.g. Problems positioning the subject, problems with the camera, transferring the photographic file, renaming or storing the file, manipulating the softwares measurement arms, reading the data, etc.)

    If YES please specify

    .....

    .....

    .....

    .....

26. Can you please give an account of the barriers that exist to using the Uillinn Method© in a clinical setting?

    .....

    .....

    .....

    .....

    .....

27. Do you have any specific professional or operational concerns about the Uillinn Method© (e.g. data protection, equipment protection, ethical issues, etc.)?

.....

.....

.....

.....

.....

28. What are the major strengths and weaknesses of the Uillinn Method©?

STRENGHTS; .....

.....

.....

.....

..... WEAKNESSES;

.....

.....

.....

.....

.....

29. How confident in the use of the Uillinn Method© are you now?

No Confidence Moderate Confidence

Minimum Confidence Extreme Confidence

30. What, if any, are the main confidence limiting factors that you have towards using the Uillinn Method©?

.....

.....

.....

.....

.....

31. Please use this space for any additional comments you would like to make regarding the Uillinn Method© or the trial. .....

.....

.....

.....

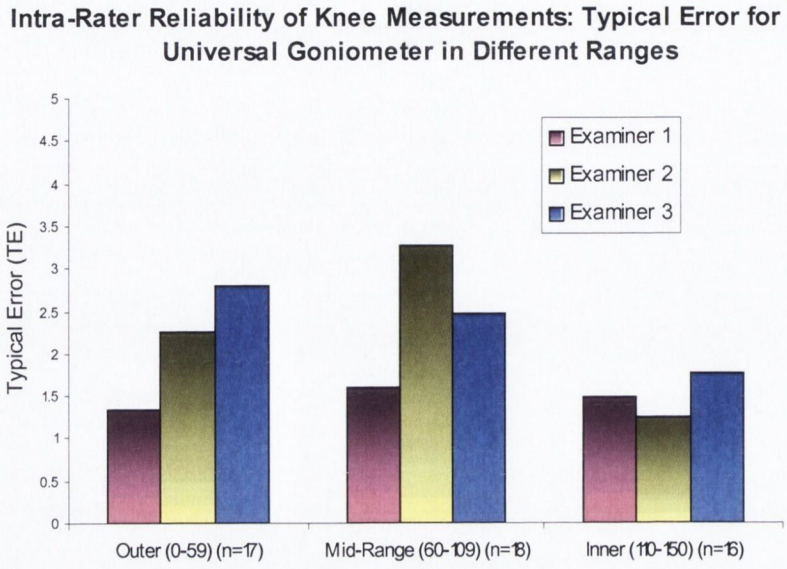
.....

.....

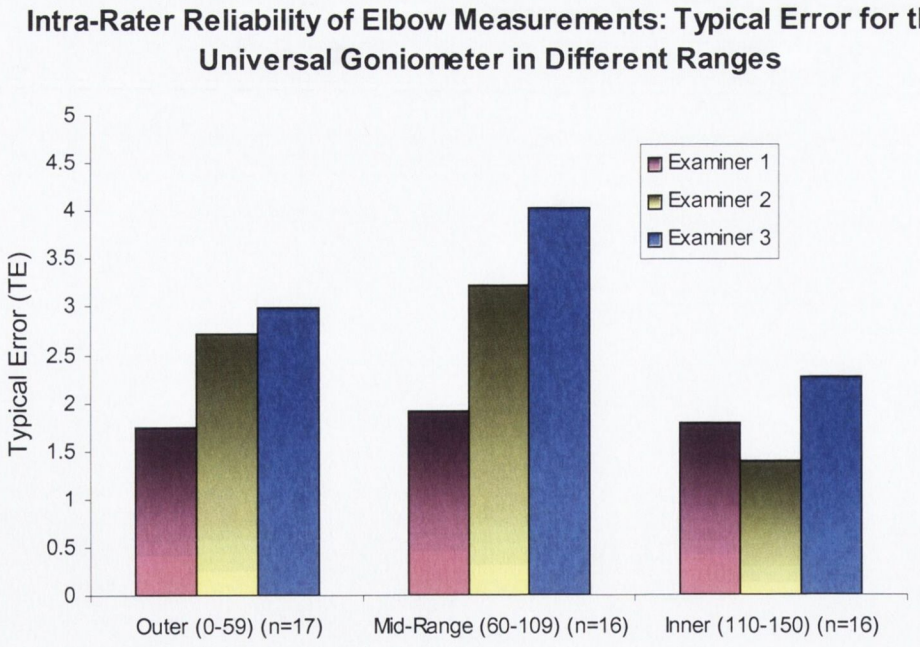
.....

.....

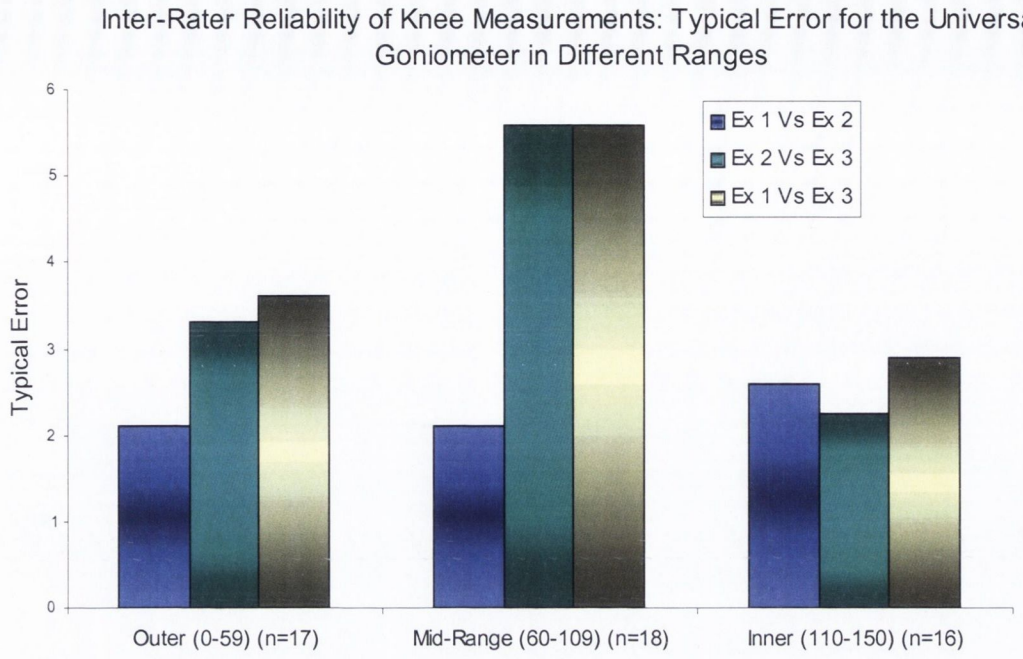
Appendix P



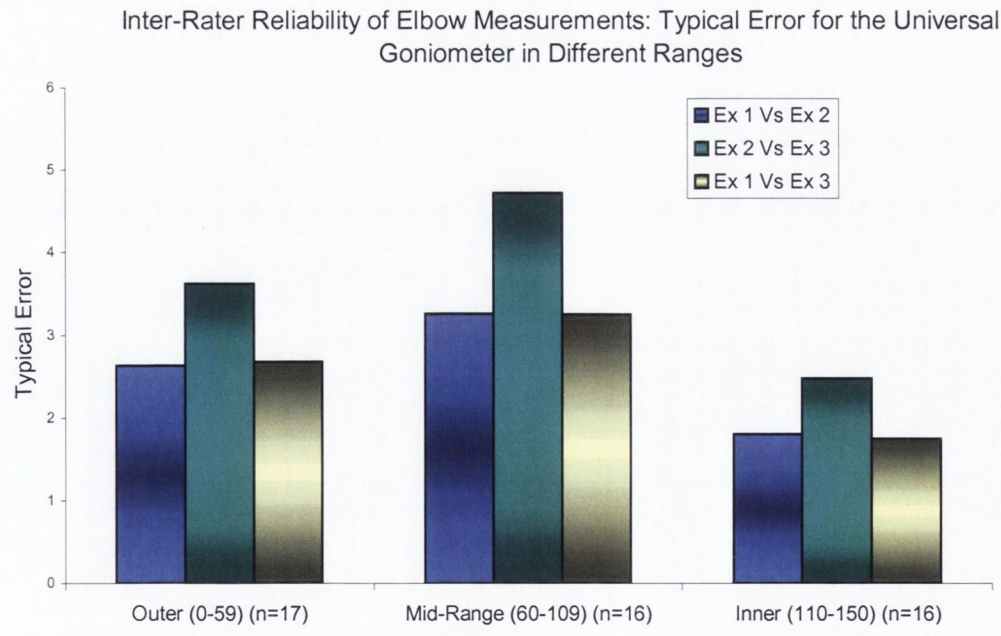
Graph 5.50. Intra-rater Reliability: Typical Error in the Different Joint Ranges (Universal Goniometer).



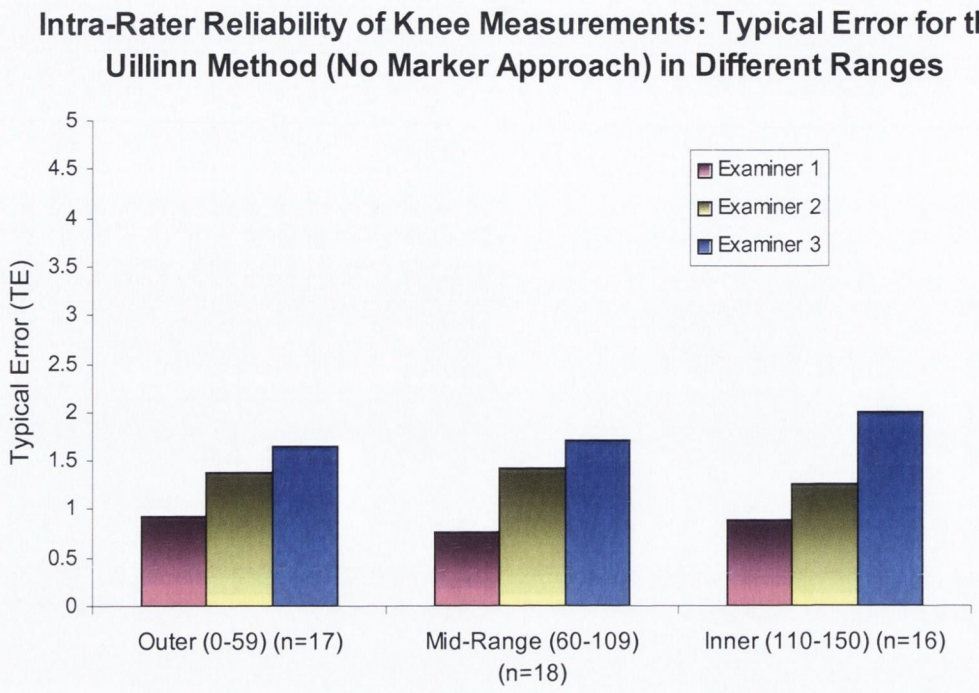
Graph 5.51. Intra-rater Reliability: Typical Error in the Different Joint Ranges (Universal Goniometer).



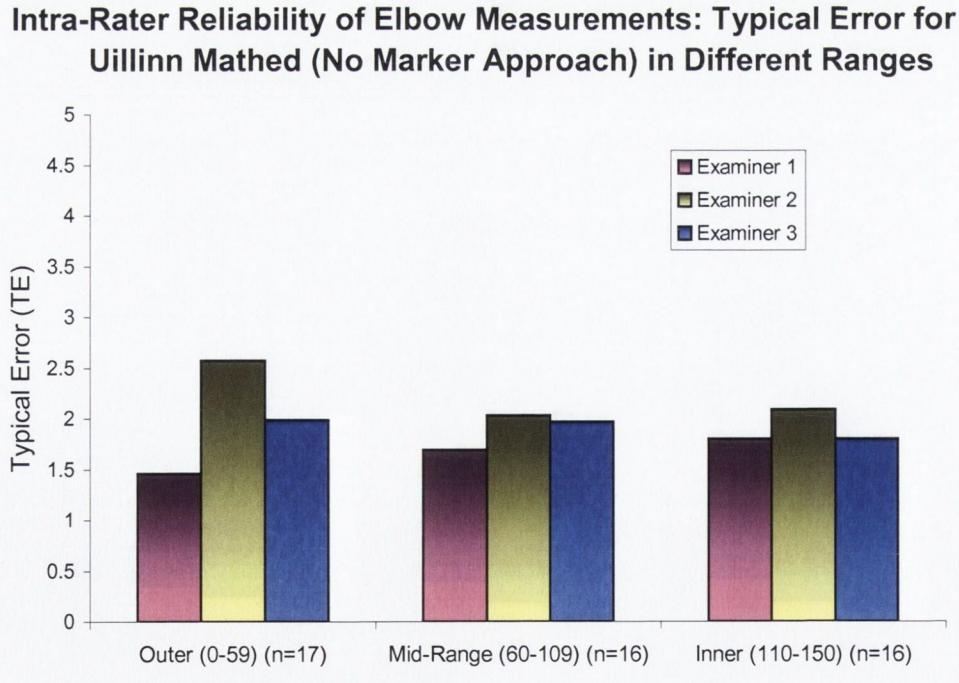
Graph 5.52. Inter-rater Reliability: Typical Error in the Different Joint Ranges (Universal Goniometer).



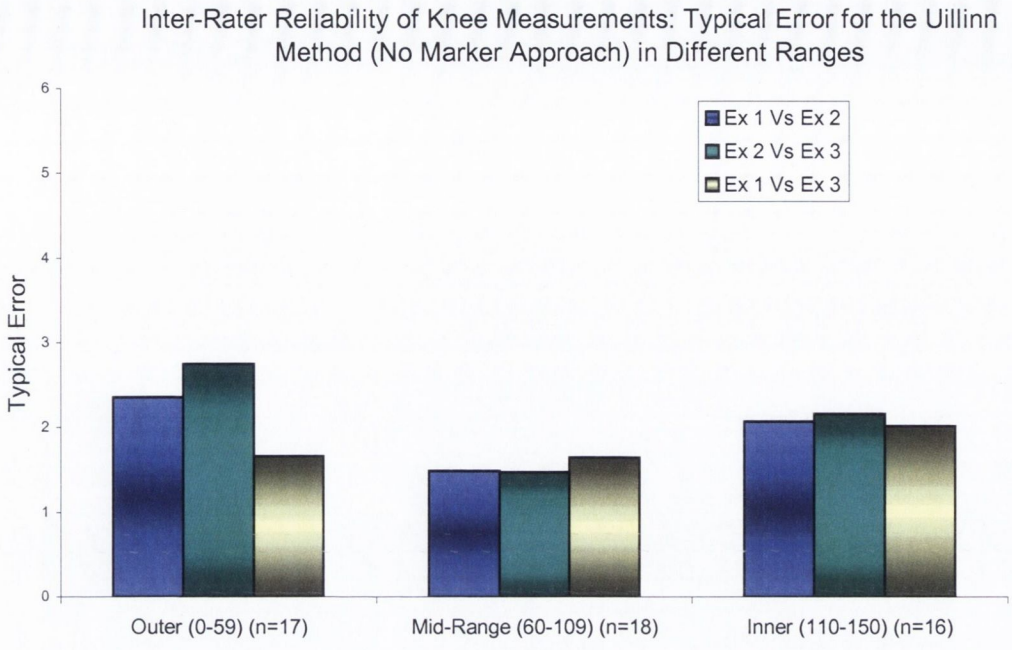
Graph 5.53. Inter-rater Reliability: Typical Error in the Different Joint Ranges (Universal Goniometer).



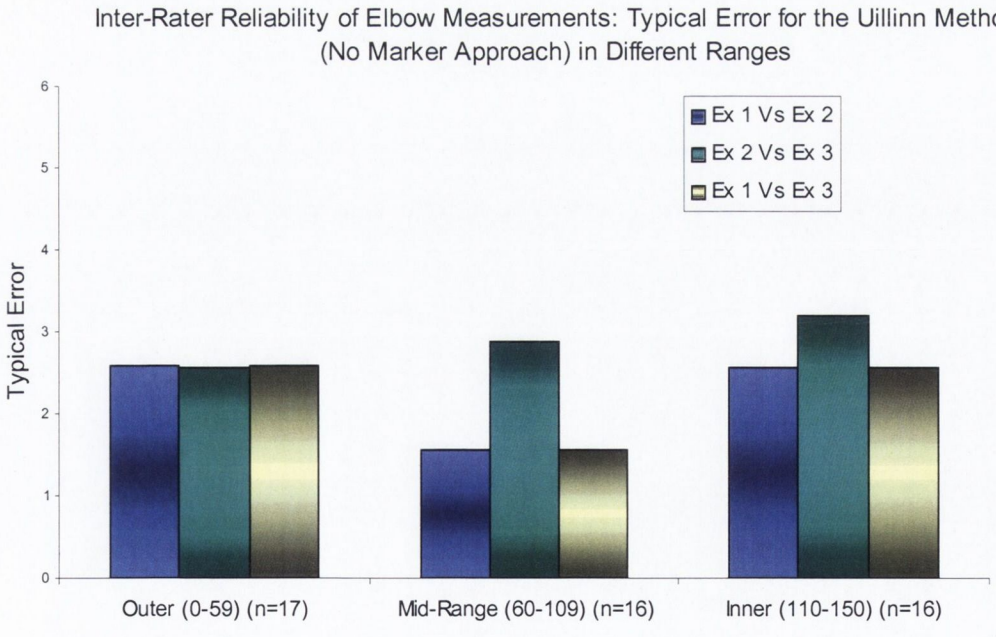
Graph 5.54. Intra-rater Reliability: Typical Error in the Different Joint Ranges (Uillinn Method - No Marker).



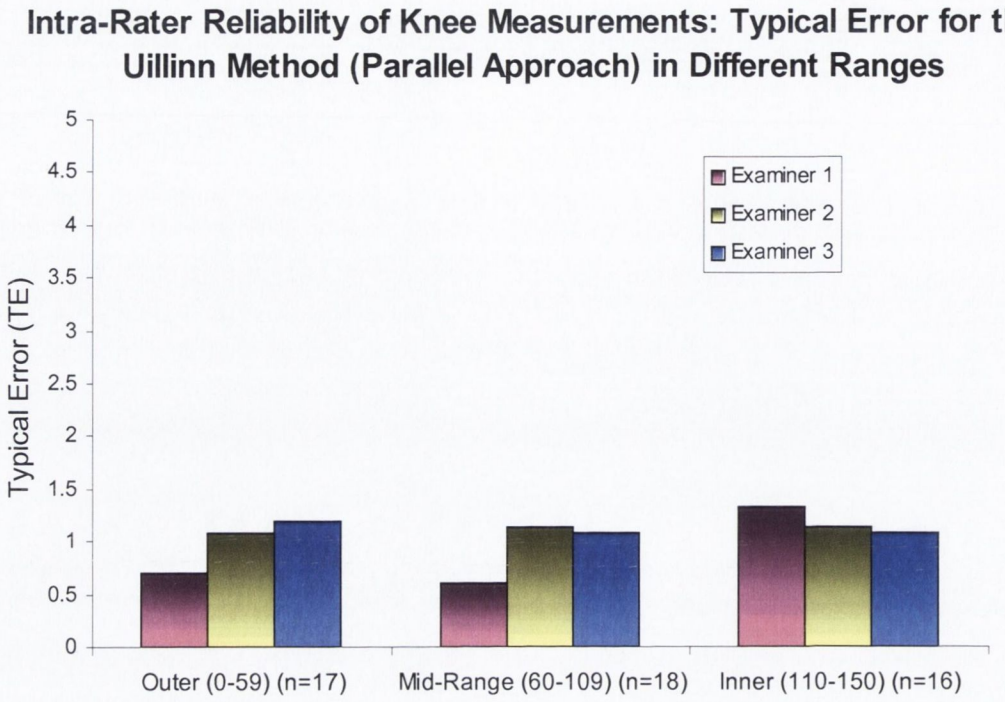
Graph 5.55. Intra-rater Reliability: Typical Error in the Different Joint Ranges (Uillinn Method - No Marker).



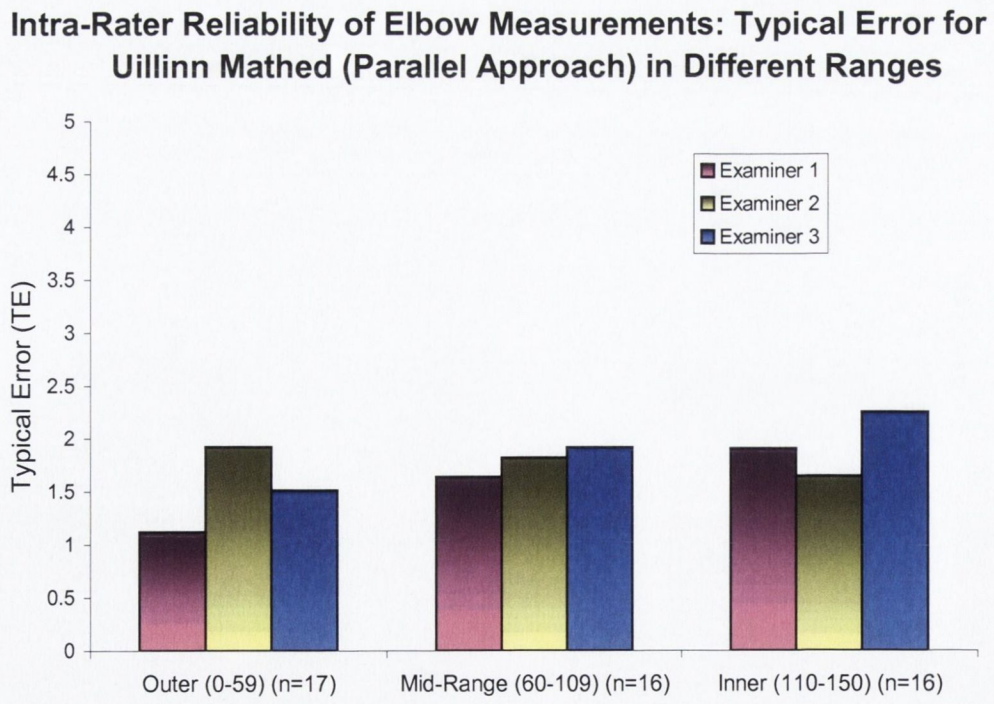
Graph 5.56. Inter-rater Reliability: Typical Error in the Different Joint Ranges (Uillinn Method - No Marker).



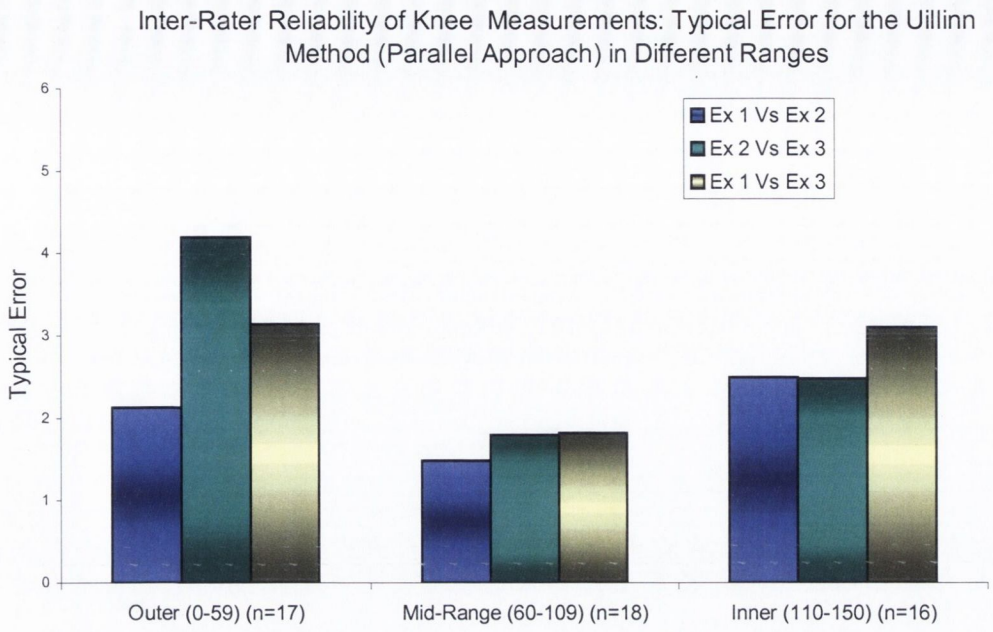
Graph 5.57. Inter-rater Reliability: Typical Error in the Different Joint Ranges (Uillinn Method - No Marker).



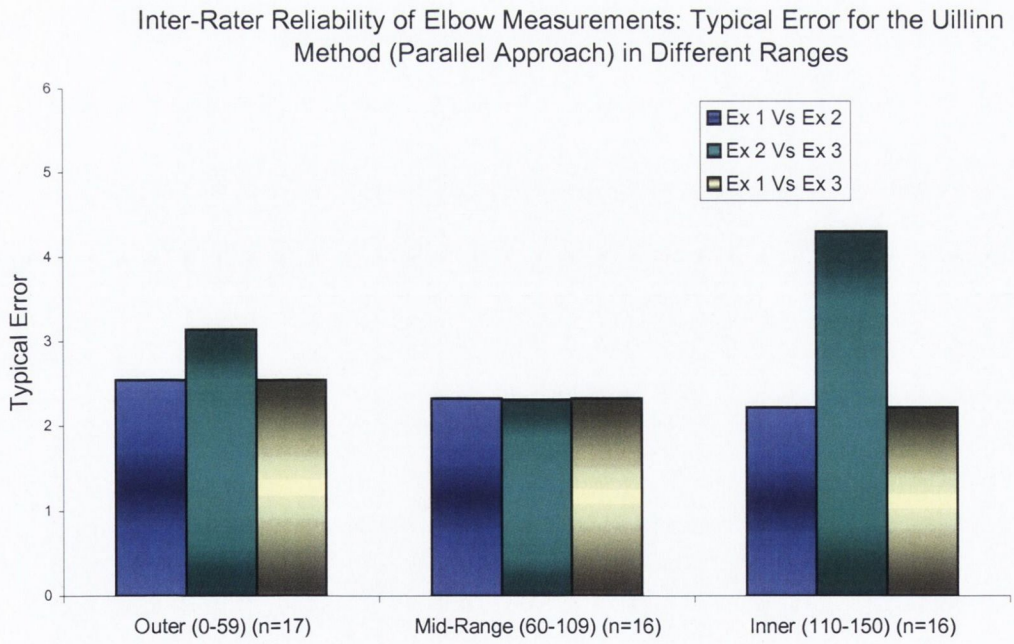
Graph 5.58. Intra-rater Reliability: Typical Error in the Different Joint Ranges (Uillinn Method - Parallel).



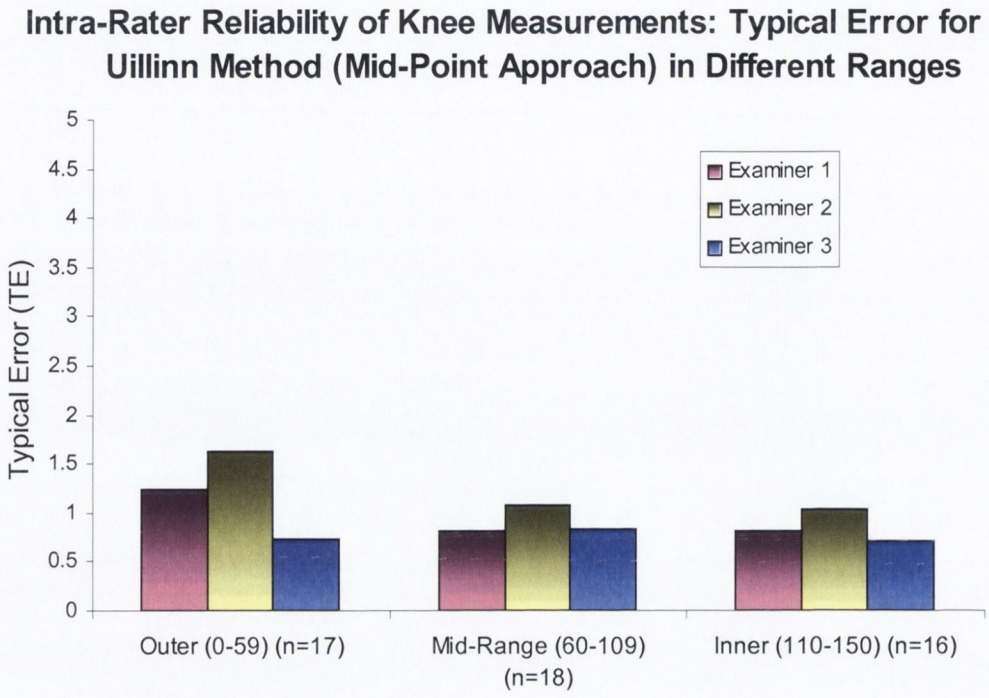
Graph 5.59. Intra-rater Reliability: Typical Error in the Different Joint Ranges (Uillinn Method - Parallel).



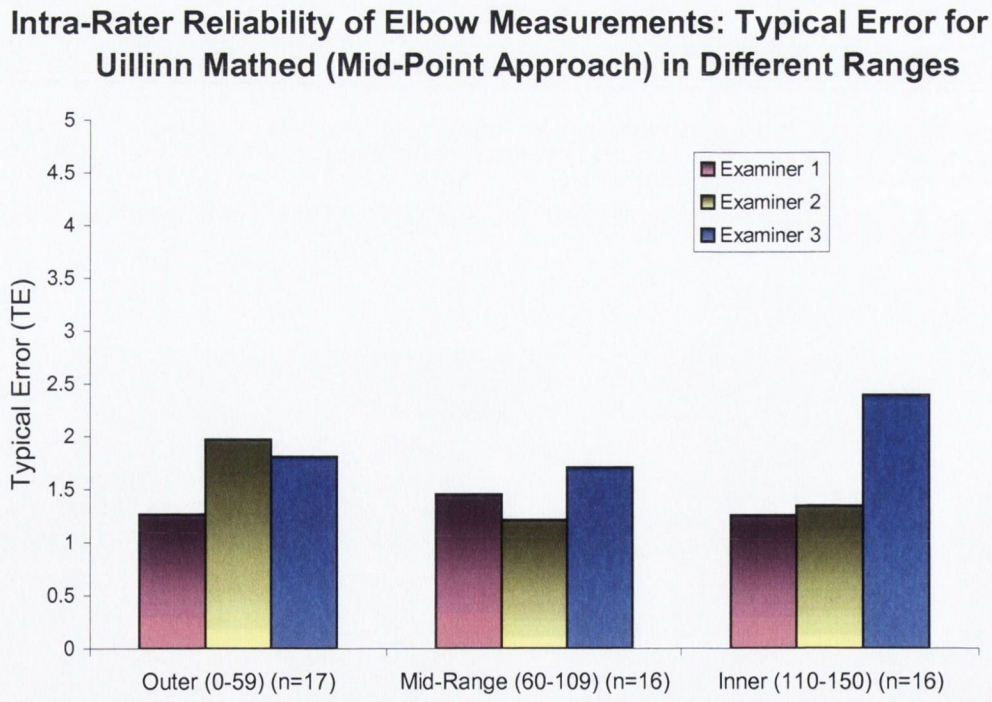
Graph 5.60. Inter-rater Reliability: Typical Error in the Different Joint Ranges (Uillinn Method - Parallel).



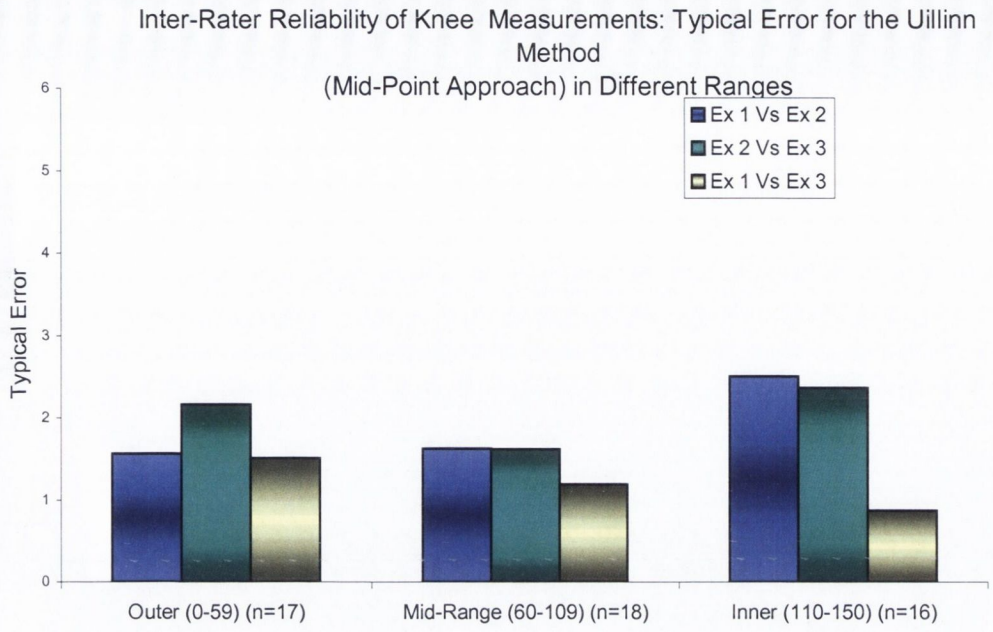
Graph 5.61. Inter-rater Reliability: Typical Error in the Different Joint Ranges (Uillinn Method - Parallel).



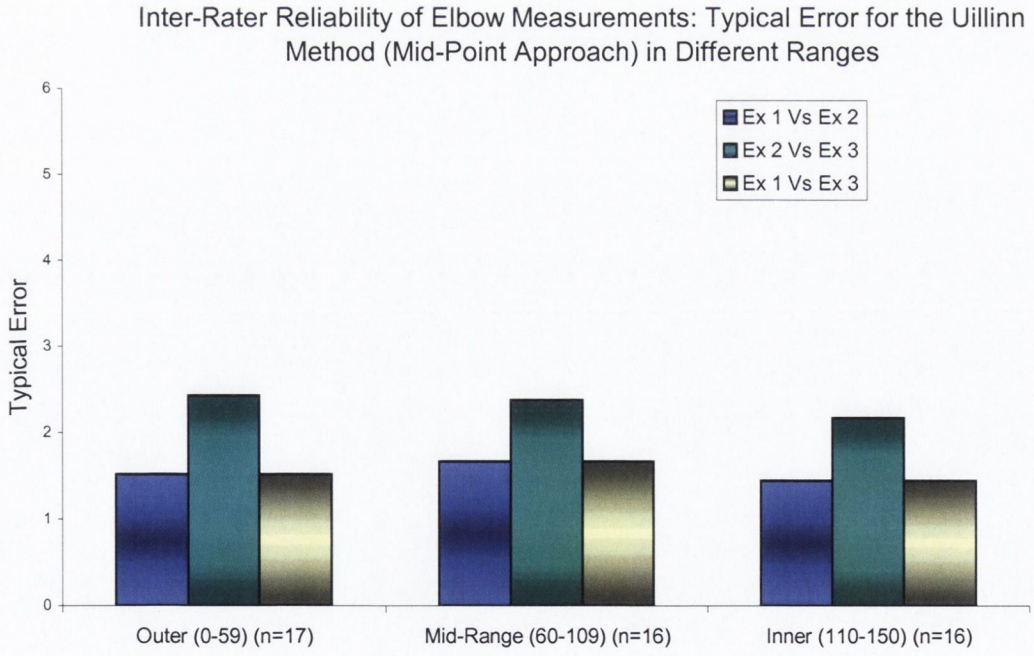
Graph 5.62. Intra-rater Reliability: Typical Error in the Different Joint Ranges (Uillinn Method – Mid-Point).



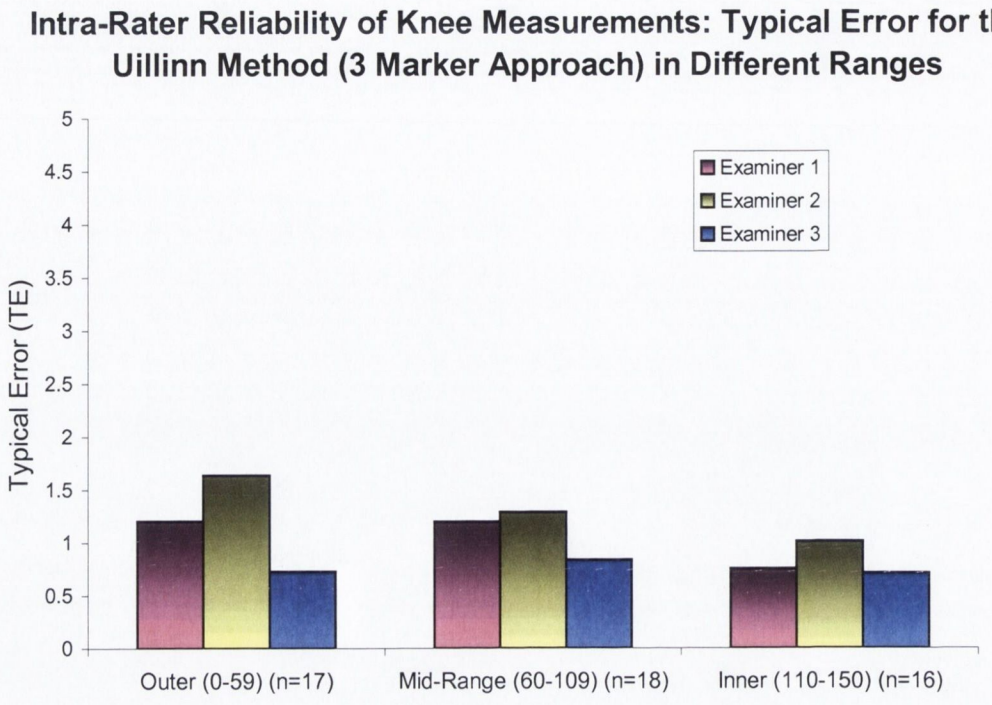
Graph 5.63. Intra-rater Reliability: Typical Error in the Different Joint Ranges (Uillinn Method – Mid-Point).



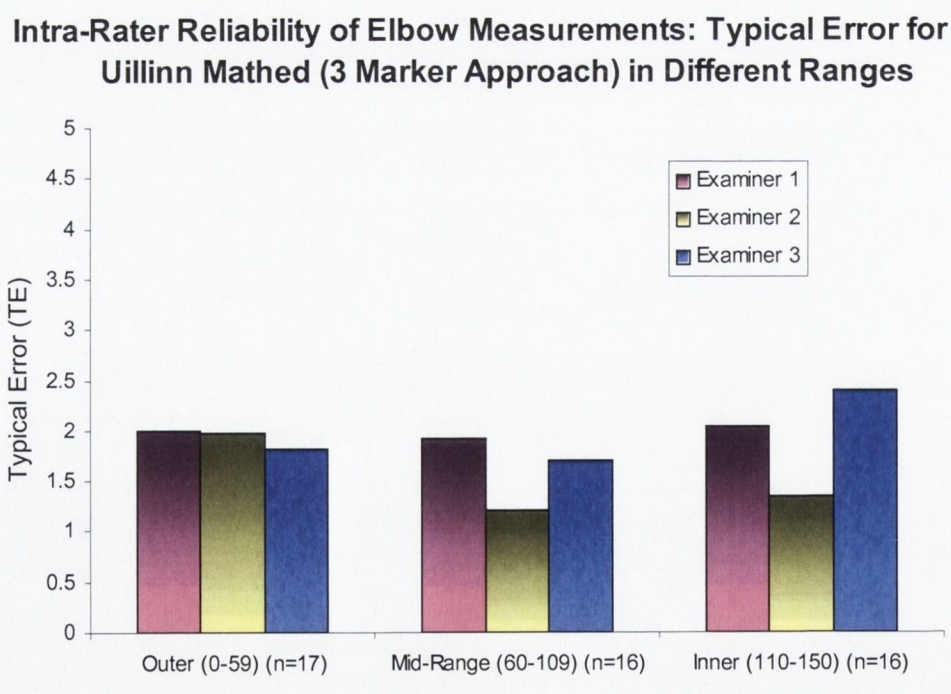
Graph 5.64. Inter-rater Reliability: Typical Error in the Different Joint Ranges (Uillinn Method – Mid-Point).



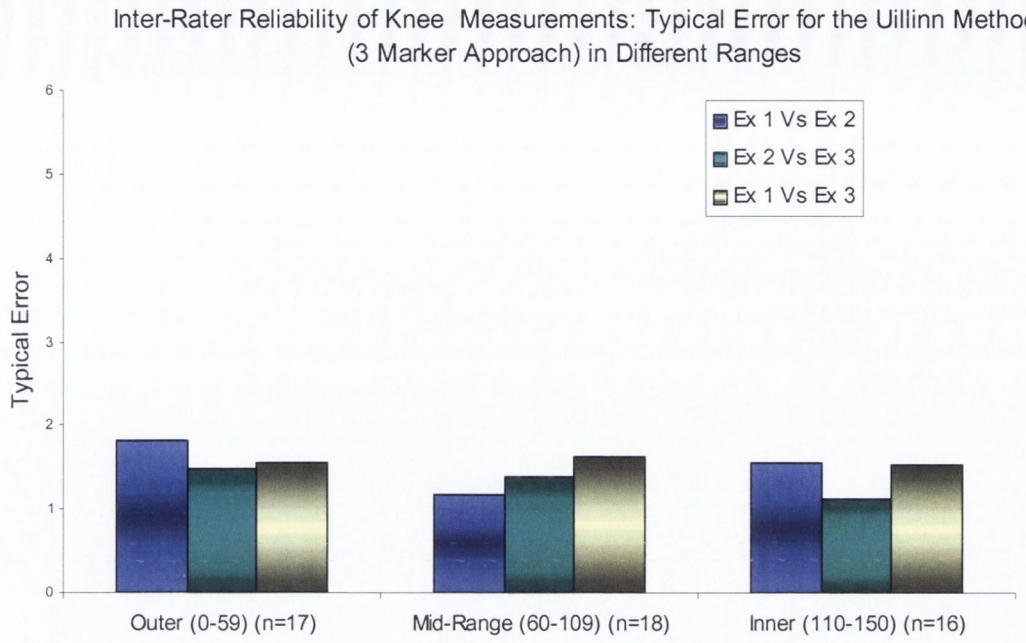
Graph 5.65. Inter-rater Reliability: Typical Error in the Different Joint Ranges (Uillinn Method – Mid-Point).



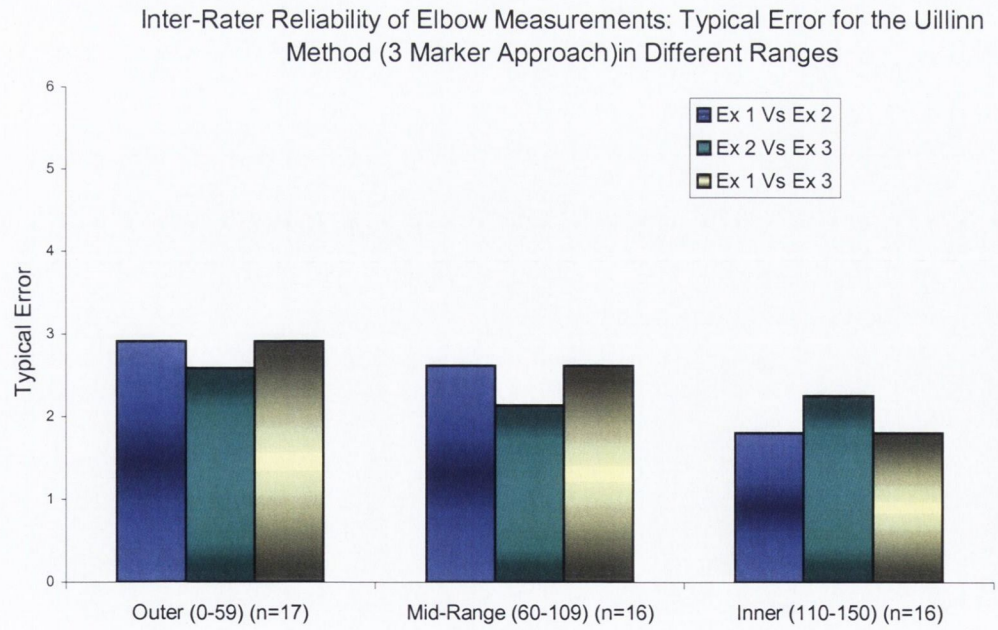
Graph 5.66. Intra-rater Reliability: Typical Error in the Different Joint Ranges (Uillinn Method – 3-Marker).



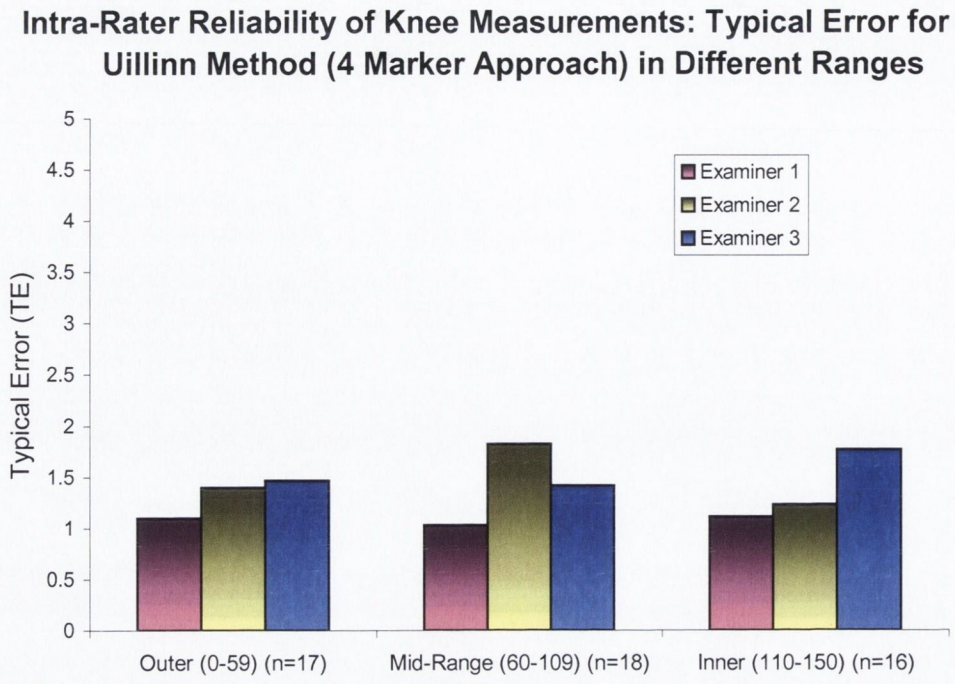
Graph 5.67. Intra-rater Reliability: Typical Error in the Different Joint Ranges (Uillinn Method – 3-Marker).



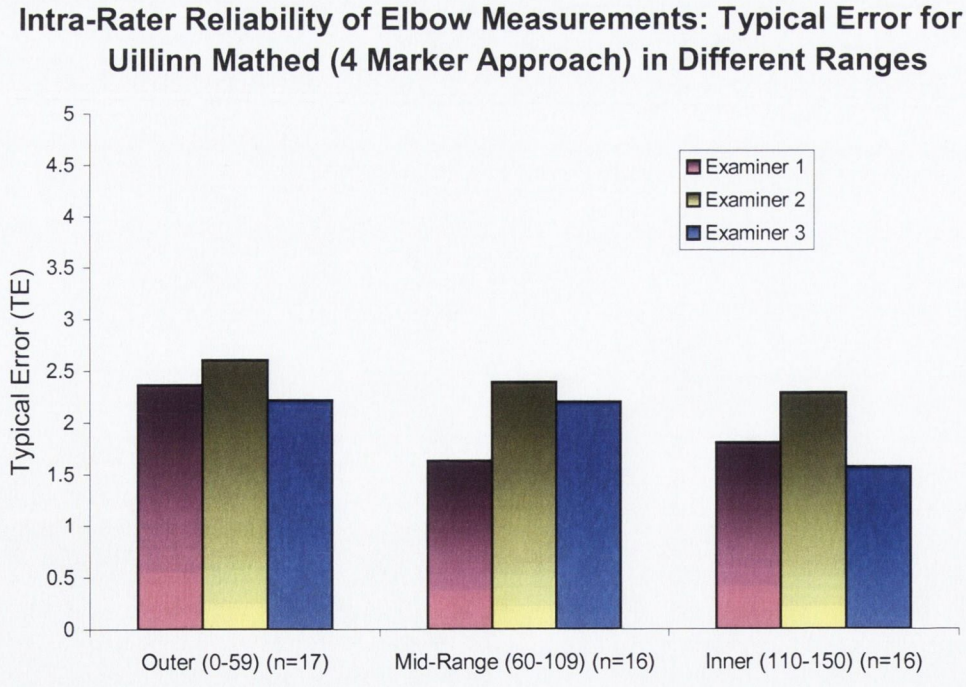
Graph 5.68. Inter-rater Reliability: Typical Error in the Different Joint Ranges (Uillinn Method – 3-Marker).



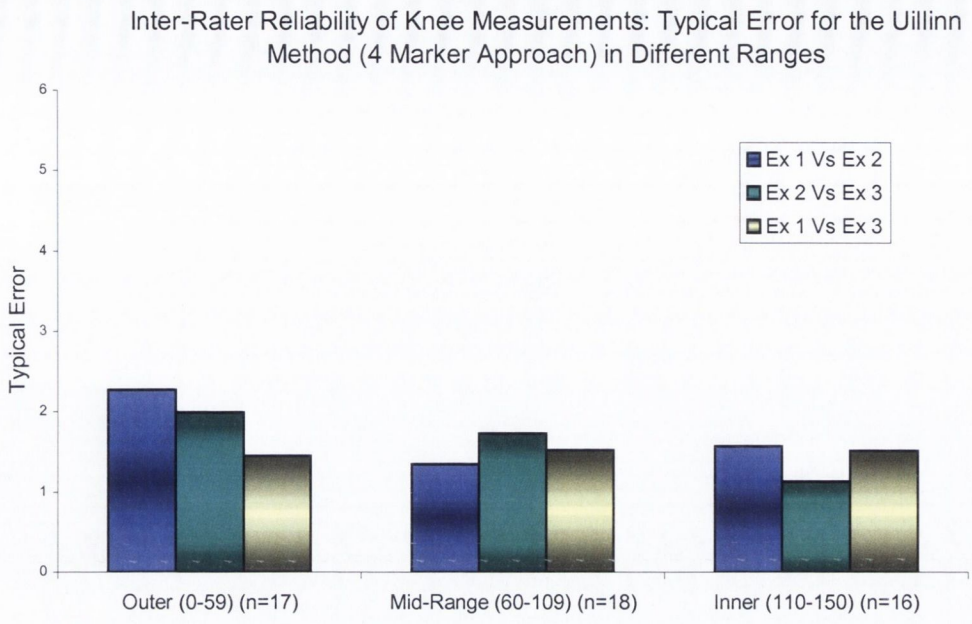
Graph 5.69. Inter-rater Reliability: Typical Error in the Different Joint Ranges (Uillinn Method – 3-Marker).



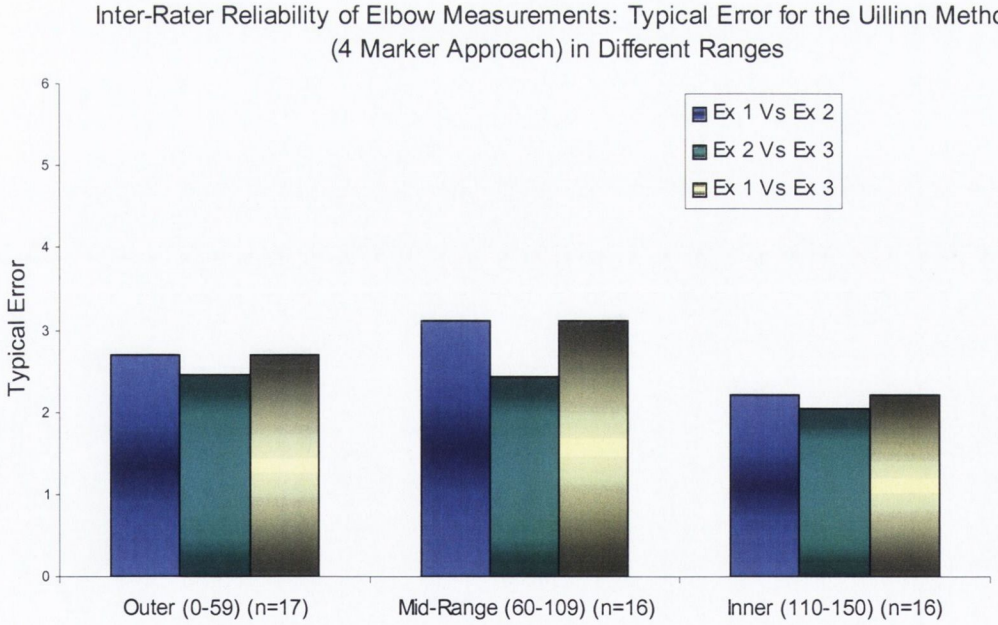
Graph 5.70. Intra-rater Reliability: Typical Error in the Different Joint Ranges (Uillinn Method – 4-Marker).



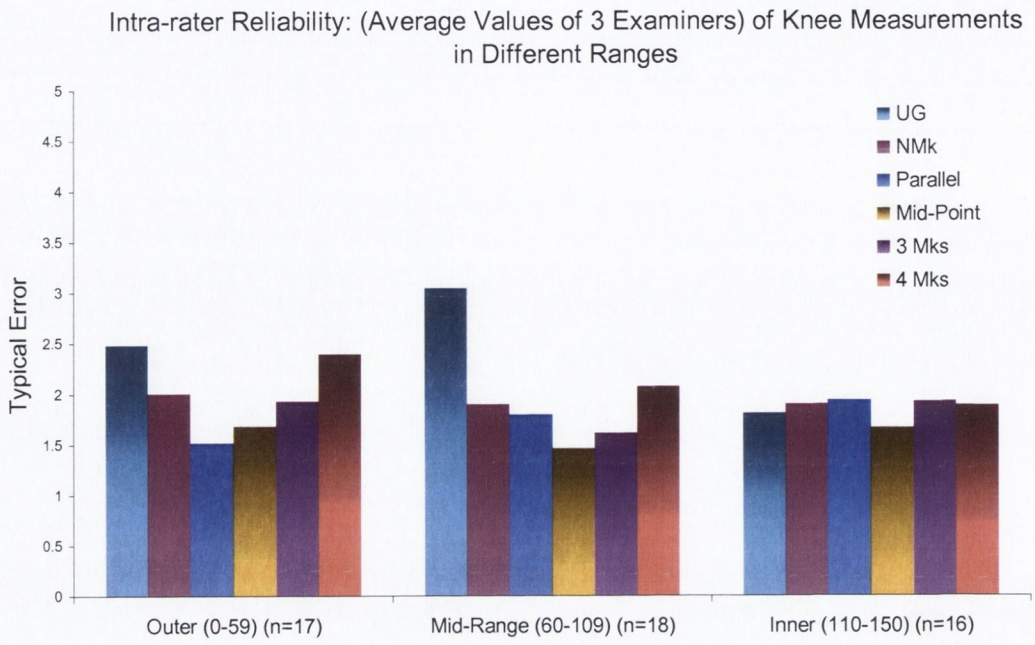
Graph 5.71. Intra-rater Reliability: Typical Error in the Different Joint Ranges (Uillinn Method – 4-Marker).



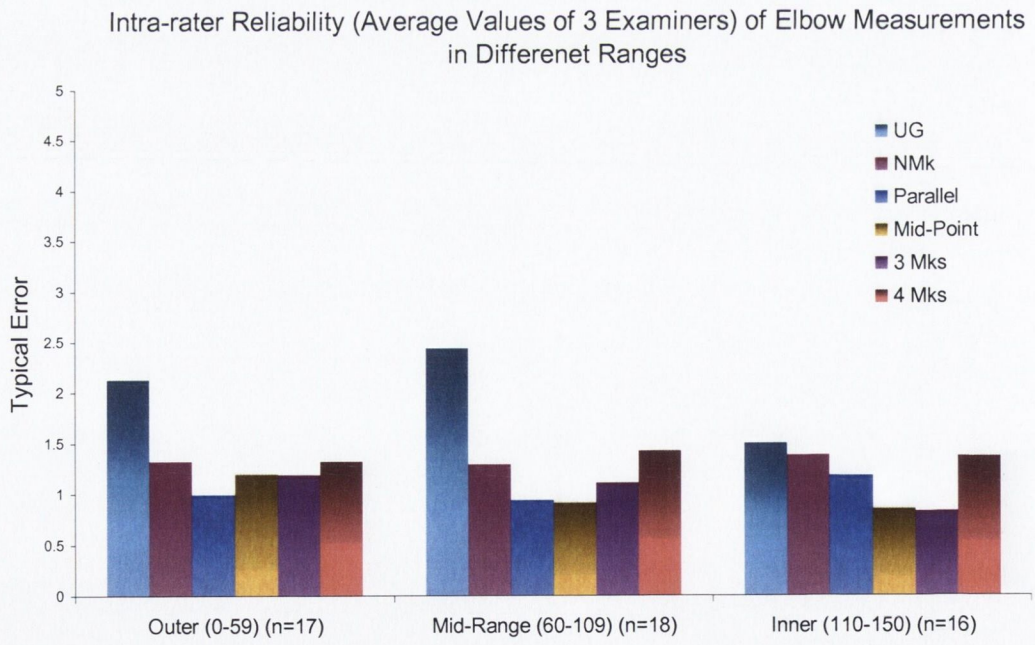
Graph 5.72. Inter-rater Reliability: Typical Error in the Different Joint Ranges (Uillinn Method – 4-Marker).



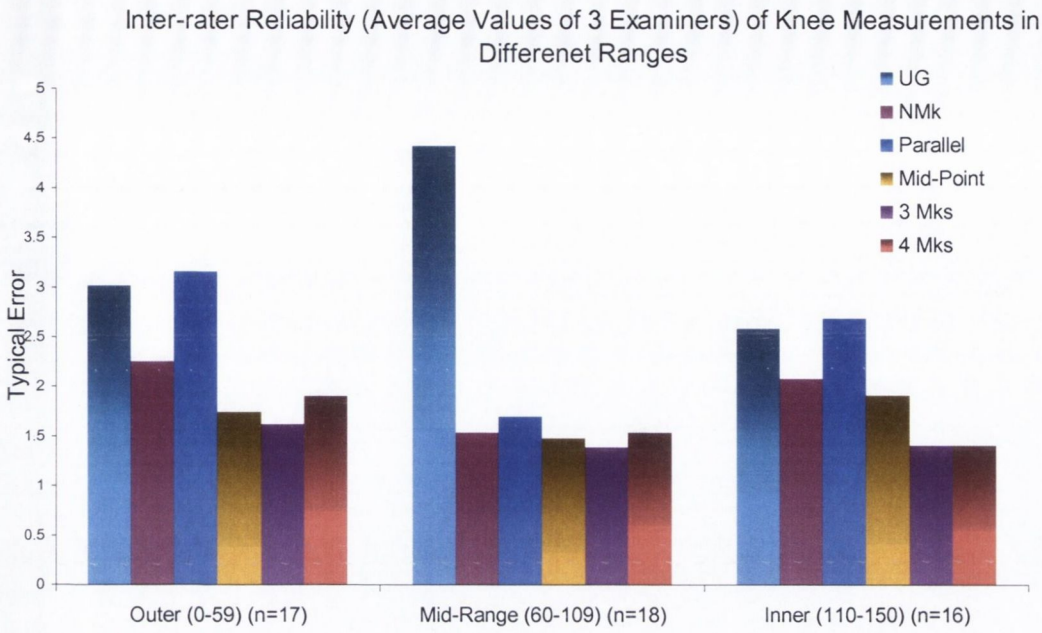
Graph 5.73. Inter-rater Reliability: Typical Error in the Different Joint Ranges (Uillinn Method – 4-Marker).



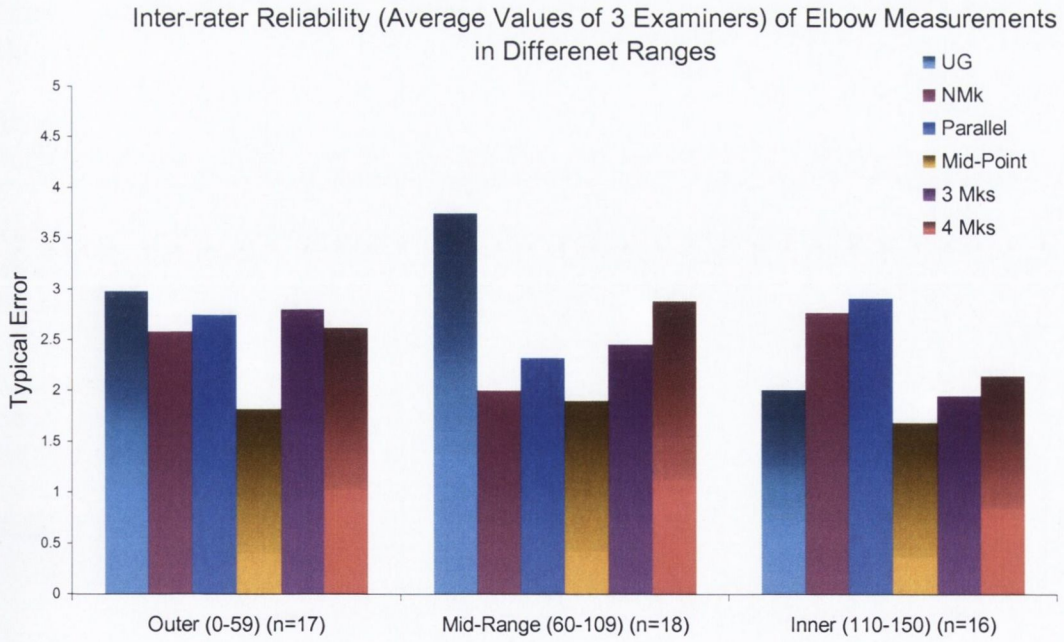
Graph 5.74. Intra-rater Reliability: Average Typical Error for the Knee in the Different Positions.



Graph 5.75. Intra-rater Reliability: Average Typical Error for the Elbow in the Different Positions.



Graph 5.76. Inter-rater Reliability: Average Typical Error in the Different Knee Positions.



Graph 5.77. Inter-rater Reliability: Average Typical Error in the Different Elbow Positions.

Appendix Q

Data Analysis Formulae;

$\overline{X}$	= Mean
SD	= Standard Deviation
SD (diff)	= Standard Deviation of the Differences (between two datasets)

Bland & Altmans 95% Limits of Agreement (LOA)

LOA = $\overline{X}$ - (SD (diff) * 1.96), $\overline{X}$ + (SD (diff) * 1.96)

Repeatability Coefficient (RC)

RC = SD (diff) * 1.96

Typical Error (TE)

TE (bi-variant) = SD (diff) /$\sqrt{2}$

TE (multi-variant) = $\sqrt{(\overline{X}\ (SD\ (for\ each\ subject))^2)}$ (Hopkins, 2000b).

Intraclass Correlation Coefficient (ICC (3,1))

ICC= (F - 1)/(F + k - 1)

(k = (number of observations - number of tests)/(number of subjects - 1).

References

Altman, DG. (1999) Practical Statistics for Medical Research. Chapman & Hall/ CRC. London.

American Academy of Orthopaedic Surgeons. (1994) The Clinical Measurement of Joint Motion. 1st Edition Edited by Greene, W. B., Heckman, J. D.

American Sports Medicine Institute (2003). http://www.asmi.org/SportsMed/anatomy/anatomy_ elbow.html. Cited Jan. 2004.

An, KN., Morrey, BF., Chao, EY. (1984) Carrying angle of the human elbow joint. [Journal Article] Journal of Orthopaedic Research. 1(4):369-78.

Andrews, JR., Timmerman, LA. (1995) Outcome of elbow surgery in professional baseball players... presented at the interim meeting of the AOSSM, New Orleans, Louisiana, February 1994. American Journal of Sports Medicine. 23(4):407-13.

Aoyagi, K., Ross, PD., Okano, K., Hayashi, T., Moji, K., Kusano, Y., Takemoto, T. (2002) Association of body mass index with joint pain among community-dwelling women in Japan. Aging-Clinical Experimental Research. 14(5): 378-81.

Armstrong, A., MacDermid, J., Chinchalkar, S., Stevens, R., King, G. (1998) Reliability of range-of-motion measurement in the elbow and forearm. Journal of Shoulder and Elbow Surgery. (7)573 – 580.

Arnheim, D.D., Prentice, W. E. (1993) Principles of Athletic Training. 8th Edition. Boston: Mosby Year Book.

Atkinson, G., Nevill, AM. (1998) Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. Sports Medicine. 26(4):217-38.

Atkinson, G., Nevill, A. M. (2000) Statistical Methods for Assessing Measurement Error (reliability) invariables relevant to sports medicine. Sports Medicine. 30(5): 375-81.

Balis, UJ. (1997) Imaging input technology. Clinics in Laboratory Medicine. 17(2): 151-74.

Barck, AL. (1997) Concordance between clinical and radiographic measurement of the mediolateral knee angle 5 years after unicompartmental knee replacement. The Knee (4)31-33.

Batterham, AM., George, KP. (2000) Reliability in evidence-based clinical practice: a primer for allied health professionals. Physical Therapy in Sport. 1(2):54-62.

Berryman Reese, N., Bandy, WD. (2002) Joint Range of Motion and Muscle Length Testing. W.B. Saunders Philadelphia.

Besier, TF., Sturnieks, DL., Alderson, JA., Lloyd, DG. (2003). Repeatability of gait data using a functional hip joint centre and a mean helical knee axis. Journal of Biomechanics. (36)1159-1168

Bland, JM., Altman, DG. (1986) Statistical Methods for Assessing Agreement Between Two Methods of Clinical Measurement. The Lancet. (8): 307-310.

Bland, JM., Altman, DG. (1996) Comparing several groups using analysis of variance. British Medical Journal. 312: 1472-3.

Bland, JM., Altman, DG. (1996) Education and Debate. Statistics Notes: Measurement error and correlation coefficients. British Medical Journal. 313: 744.

Bland, JM., Altman, DG. (1999) Variables and parameters. BMJ 1999;318:1667.

Boone, DC., Azen, SP. (1979) Normal Range of Movement of Joints in Male Subjects. Journal of Bone and Joint Surgery. 61A: 756-759.

Borchers, L., Kee, CC. (1999) An experience in telenursing. Clinical Nurse Specialist. 13(3): 115-8.

Bovens, A., van Baak, M., Vrencken, J., Wijnen, J., Verstappen, F. (1990) Variability and reliability of joint measurement. The American Journal of Sports Medicine.
18 (1): 58-63.

Boyd, RN., Graham, HK. (1999) Objective measurement of clinical findings in the use of botulinum toxin type A for the management of children with cerebral palsy. European Journal of Neurology 6 (suppl 4): 523-S35.

Braun, BL., Amundson, LR. (1989) Quantitative assessment of head and shoulder posture. Archives of Physical Medicine & Rehabilitation. 70 (4):322-9.

Braun, BL. (1991) Postural differences between asymptomatic men and women and craniofacial pain patients. Archives of Physical Medicine & Rehabilitation.
72 (9): 653-6.

Brosseau, L., Tousignant, M., Budd, J., Chartier, N., Duciaume, L., Plamondon, S. (1997) Intratester and intertester reliability and criterion validity of the parallelogram and universal goniometers for measuring maximum active knee flexion in healthy subjects. Physiotherapy Research International. 2: 150-166.

Brosseau, L., Balmer, S., Tousignant, M., O'Sullivan, J. P., Goudreault, C., Goudreault, M., Gringras, S. (2001) Intra- and intertester reliability and criterion validity of the parallelogram and universal goniometers for measuring maximum active knee flexion and extension of patients with knee restrictions. Archives of Physical Medicine & Rehabilitation. 82 (3): 396-402.

Bruton, A., Ellis, B., Goddard, J. (1999) Comparison of Visual Estimation and Goniometry for Assessment of the Metacarpophalangeal Joint Angle. Physiotherapy. 85(4). 201-208.

Bruton, A., Conway, JH., Holgate, ST. (2000). Reliability: What is it, and how is it measured?. Physiotherapy 86 (2): 94-9.

Bunker, TD., Apthorp, H. (1995) The knee. In: Foy, M. A , Fagg, P. S. Editor(s). Medicolegal Reporting in Orthopaedic Trauma Vol. 3.3. Edinburgh: Churchill Livingstone: 1-30.

Buntic, RF., Siko, PP., Buncke, GM., Ruebeck, D., Kinf, GM., Buncke, HJ. (1997) Using the Internet for Rapid Exchange of Photograph and X-ray Images to Evaluate Potential Extremity Replantation Candidates. Journal of Trauma. 43 (2): 342-4.

Cameron, DM., Bohannon, RW. (1993) Relationship between active knee extension and active straight leg raise test measurements. Journal of Orthopaedic & Sports Physical Therapy. 17 (5): 257-60.

Clapper, MP., Wolf, SL. (1988) Comparison of the reliability of the Orthoranger and the standard goniometer for assessing active lower extremity range of motion. Physical Therapy. 68 (2): 214-8.

Churchill, DL., Incavo, SJ., Johnson, CC., Beynnon, BD. (1998) The transepicondylar axis approximates the optimal flexion axis of the knee. Clinical Orthopaedics & Related Research. (356):111-8.

Coxford, P., Jones, K., Barker, K. (1998) Inter-tester comparison between visual estimation and goniometric measurement of ankle dorsiflexion. Physiotherapy Theory and Practice. 14, 107-113.

Croft, P., Pope, D., Boswell, R., Rigby, A., Silman, A. (1994) Observer variability in measuring elevation and external rotation of the shoulder. Primary Care Rheumatology Society Shoulder Study Group. British Journal of Rheumatology. 33 (10): 942-6.

Cunningham, LS., Kelsey, JL. (1984) Epidemiology of musculoskeletal impairments and associated disability. American Journal of Public Health. 74 (6): 574-9.

Curtis, E., White, P. (2002) Resistance to Change: Causes and Solutions. Nursing Management. 8(10) 15-20.

Daft, RL. (2001) The Leadership Experience. 2nd Ed. Harcourt College Publishers, London.

Daly, LE., Bourke, GJ. (2000) Interpretation and Uses of Medical Statistics. 5th ed. Blackwell Science Ltd, Oxford, UK.

Department of Health & Children Ireland. (2001) Quality and Fairness A Health System for You. Dublin: The Stationary Office.

Domholt, E. (1993) Physical therapy research, principles and application. Philadelphia: W.B. Saunders.

Duckworth, M. (1999) Outcome Measurement Selection and Typology. Physiotherapy. 85 (1): 21-27.

Dunlevy, C., Cooney, M., Gormley J. (2000) Objective measurements of head and neck alignment using information technology. Physiotherapy Ireland. 21 (2): 24-5.

Dunmore, T. (2003) Image is everything. Stuff; Gadgets, Gear, Technology. PO Box 280, Sittingbourne, Kent, ME9 8FB. April.

Dunnett, CW. (1964) New tables for multiple comparisons with a control. Biometrics. 482-491.

Dutoit, M. (1998) Percutaneous epiphysiodesis in the treatment of adolescent genuvalgum. Revue de Chirurgie Orthopedique et Reparatrice. 84 (7): 623-7.

Ekstrand, J., Wiktorsson, M., Oberg, B., Gillquist, J. (1982) Lower extremity goniometric measurements: a study to determine their reliability. Archives of Physical Medicine & Rehabilitation. 63(4): 171-5.

Eliasziw, M., Young, SL., Woodbury, MG,, Fryday-Field, K. (1994) Statistical methodology for the concurrent assessment of interrater and intrarater reliability: using goniometric measurements as an example. Physical Therapy. 74(8):777-88.

Ellis, H., Dussek, J. (1995) Surface Anatomy. Gray's Anatomy. 38th Edition. Edited: Bannister, LH., Berry, MM., Collins, P., Dyson, M., Dussek, J., Ferguson, MWJ. Churchill Livingstone, NY.

Ellis, H. (1997) Clinical Anatomy: A Revision and Applied Anatomy for Clinical Students. 9th Edition. Oxford: Blackwell Science: 247-250.

Enwemeka, CS. (1986) Radiographic verification of knee goniometry. Scandinavian Journal of Rehabilitation Medicine. 18 (2): 47-9.

Finch, E., Brooks, D., Stratford, P., Mayo, N. (2002) Physical Rehabilitation Outcome Measures. A Guide to Enhanced Clinical Decision Making. Canada, Lippincott, Williams & Wilkins.

Fish, DR., Wingate, L. (1985) Sources of Goniometric Error at the Elbow. Physical Therapy. 65(11) 1666-1670.

Fisher, LD., van Belle, G. (1993) Biostatistics: A Methodology for the Health Sciences. 1st Edition New York: Wiley-Interscience Publications, John Wiley & Sons: 32.

Fleiss, JL. (1986). The Design and Analysis of Clinical Experiments. New York: John Wiley and Sons.

Fosang, AL., Galea, MP.. McCoy, AT., Reddihough, DS. (2003) Story I. Measures of muscle and joint performance in the lower limb of children with cerebral palsy. Developmental Medicine & Child Neurology. 45(10):664-70.

Fuji, K., Suzuki, N., Ishijima, S., Kita, K., Sonoda, T., Dezawa, M., Sugita, K., Niimi, H. (1997) Abnormal DNA synthesis activity induced by X-rays in nevoid basal cell carcinoma syndrome cells. Biochemical & Biophysical Research Communications. 240 (2): 269-72.

Gajdosik, RL., Albert, CR., Mitman, JJ. (1994) Influence of hamstring length on the standing position and flexion range of motion of the pelvic *angle*, lumbar *angle*, and thoracic *angle*. Journal of Orthopaedic & Sports Physical Therapy. 20(4): 213-9.

Ganjikia, S., Duval, N., Yahia, L., de Guise, J. (2000) Three-dimensional Knee analyzer validation by simple fluoroscopic study. The Knee. (7): 221-231.

George, K., Batterham, A., Sullivan, I. (2000) Validity in clinical research: A review of basic concepts and definitions. Physical Therapy In Sport. (1): 19-27.

Gleeson F. Digital camera sales skyrocket in Europe. Electronic News Network http://www.enn.ie/frontpage/news-9371983.html August 18 2003.

Gogia, PP., Braatz, JH., Rose, SJ., Norton, BJ. (1987) Reliability and Validity of Goniometric Measurement at the Knee. Physical Therapy. 67 (2): 192-195.

Goodwin, J., Clark, C., Deakes, J., Burdon, D., Lawrence, C. (1992) Clinical methods of goniometry: a comparative study. Disability and Rehabilitation. 14(1). 10-15.

Granberry, WM., Noble, PC., Woods, GW. (1990) Evaluation of an electrogoniometric instrument for measurement of laxity of the knee. Journal of Bone & Joint Surgery - American Volume. 72 (9): 1316-22.

Gray, RJ., Quayle, AA., Horner, K., Al-Gorashi, AJ. (1991) The effects of positioning variations in transcranial radiographs of the temporomandibular joint: a laboratory study. The British Journal of Oral & Maxillofacial Surgery. 29 (4): 241-9.

Green, J., Britten, N. (1998) Qualitative research and evidence based medicine. British Medical Journal. 316: 1230-1232

Greene, BL., Wolf, SL. (1989) Upper extremity joint movement: comparison of two measurement devices. Archives of Physical Medicine & Rehabilitation. 70 (4): 288-90.

Greene WB, Heckman JD (1994) The Clinical Measurement of Joint Motion. Rosemont, IL, American Academy of Orthopedic Surgeons

Greene, CG., Edwards, TB., Wade, MR., Caeson, EW. (2001) Reliability of the Quadriceps Angle Measurement. American Journal of Knee Surgery.

(14): 97-103.

Grohmann, JEL. (1983) Comparison of Two Methods of Goniometry. Physical Therapy. 63(6)923-925.

Hageman, PA., Sorensen, TA. (1991) Eccentric Isokinetics. Eccentric Muscle Training in Sports and Orthopaedics. Edited: Albert, M., Tighe, A. Churchill Livingston: NY, Tokyo.

Hallen, LG., Lindahl, O. (1966) The screw-home movement in the knee joint. Acta Othopaedica Scandinavica. 37: 97-106.

Hayes, K, Walton, J.R., Szomor, Z.L., Murrell, G.A.C. (2001) Reliability of five methods for assessing shoulder range of motion. University of New South Wales. Australian Journal of Physiotherapy. 47. 289-294.

Heath, CH., Staheli, LT. (1993) Normal limits of knee angle in white children--genu varum and genu valgum. Journal of Pediatric Orthopedics. 13 (2): 259-62.

Heiderscheit, BC., Hamill, J., Caldwell, GE. (2000) Influence of Q-angle on lower-extremity running kinematics. Journal of Orthopaedic & Sports Physical Therapy. 30 (5): 271-8.

Heidt, RS., Sweeterman, LM., Richelle, LC., Traub, JA., Tekulve, ATC. (1997) Avoidance of Soccer Injuries with Preseason Conditioning. American Journal of Sports Medicine. 28 (5): 659-662.

Hellwig, E., Perrin, D. (1995) The mechanical and clinical reliability of the kinetic communicators gravity correction procedure. Isokinetic Exercise Science. 5: 85-91.

Hicks, CM. (1999) Research Methods for Clinical Therapists. Applied project design and analysis. London: Churchill Livingstone.

Hicks CM. (2004). Research Methods for Clinical Therapists. Applied Project Design and Analysis. 4rd ed. London: Churchill Livingstone.

Holm, I., Bolstad, B., Lutken, T., Ervik, A., Rokkum, M., Steen, H. (2000) Reliability of goniometric measurements and visual estimation of hip ROM in patients with osteoarthrosis. Physiotherapy Research International. 5(4) 241-248.

Hopkins, WG. (2000a) Measures of reliability in sports medicine and science. Sports Medicine. 30(1):1-15.

Hopkins WG (2000b). Reliability from consecutive pairs of trials (Excel spreadsheet). In: A new view of statistics. sportsci.org: Internet Society for Sport Science, sportsci.org/resource/stats/xrely.xls

Ishii, Y., Ohmori, G., Bechtold, JE., Sherman, RE., Gustilo, RB. (1995) Accuracy of the short radiograph in the measurement of the tibiofemoral angle. The Knee. (2): 81-84.

Ishii,Y., Ohmori, G., Bechtold, JE., Gustilo, RB. (1995) Extramedullary versus intramedullary alignment guides in total knee arthroplasty. Orthopaedics & Related Research. (318): 167-75.

Iwasawa, T., Inaba, Y., Nishimura, G., Aida, N., Kameshita, K., Matsubara, S. (1999) MR findings of bowlegs in toddlers. Pediatric Radiology. 29 (11): 826-34.

Jackson, JP., Waugh, W. (1984) Surgery of the knee joint. London: Chapman and Hall.

Jamshidi, M., Smith, AW. (1996) Clinical measurement of spasticity using the pendulum test: comparison of electrogoniometric and videotape analyses. Archives of Physical Medicine & Rehabilitation. 77 (11): 1129-32.

Kapandji, IA. (2001) The Physiology of the Joints; Volume One, the upper limb. 5th Edition . Churchill Livingstone. London.

Kandemir, U., Fu, FH., McMahon, P.J. (2002) Elbow injuries. [Review] Current Opinion in Rheumatology. 14(2):160-7.

Karkouti, E., Marks, R. (1997) Reliability of Photographic Range of Motion Measurements in a Healthy Sample: Knee and Ankle Joint Measurement. Physiotherapy Canada. (4): 24-31.

Kendall, FP., McCreary, EK., Provance, PG. (1993) Muscles testing and function, with posture and pain. 4th Edition. Baltimore: Williams and Wilkins. Pp: 420.

Klingele, KE. Kocher, MS. (2002) Little league elbow: valgus overload injury in the paediatric athlete. Sports Medicine. 32(15): 1005-15.

Koh, TJ., Grabiner, MD., De Swart, RJ. (1992) In vivo tracking of the human patella. Journal of Biomechanics. 25: 637-643.

Koper, L., Photomag 2001. [Cited 06 – 06 – 02]. http://www.photomag.co.uk/Data/Glossary/Pages/Parallax%20error.htm. 2001.

Lafortune, MA., Cavanagh, PR., Sommer, HJ., Kalenak, A. (1992) Three-dimensional kinematics of the human knee during walking. Journal of Biomechanics. (25): 247-57.

Lantz, CA., Chen, J., Buch, D. (1999) Clinical Validity and Stability of Active and Passive Cervical Range of Motion With Regard to Total and Unilateral Uniplanar Motion. SPINE. 24: 1082.

Lathinghouse, LH., Trimble, MH. (2000) Effects of isometric quadriceps activation on the Q-angle in women before and after quadriceps exercise. Journal of Orthopaedic & Sports Physical Therapy. 30 (4): 211-6.

Leen T'Jonck, S., Schacke, R., Lysens, E., Witvrouw, K., Delvaux, K. (1997) Peers. Intertester and intratester reliability of the standard goniometer and the cybex EDI 320 for active and passive shoulder range of motion in normals and patients. Proceedings of the First Conference of the ISG.

Li, X. M., Liu, B., Deng, B., Zhang, SM. (1996) Normal six-degrees-of-freedom motions of the knee joint during level walking. Journal of Biomechanical Engineering. 188: 258-61.

Livingston, LA., Mandigo, JL. (1997) Bilateral within-subject Q angle asymmetry in young adult females and males. Biomedical Sciences Instrumentation. 33: 112-7.

Livingston, LA. (1998) The quadriceps angle: a review of the literature. Journal of Orthopaedic & Sports Physical Therapy. 28 (2): 105-9.

LoBiondo-Wood,G., Haber, J. (2002) Reliability and Validity. Nursing Research, Methods, Critical Appraisal and Utilization. 5th Edition. St. Louis: Mosby. 311-330.

Lonner, JH., Laird, MT., Stuchin, SA. (1996) Effect of rotation and knee flexion on radiographic alignment in total knee arthroplasties. Clinical Orthopaedics & Related Research. (331): 102-6.

Lythgo, N.(2004) http://isb.ri.ccf.org/biomch-l/archives/biomch-l-1998-04/00083.html. Cited April 2004.

Mackey, A.H., Walt, S.E. Lobb, G., Stott, NS. (2004) Intraobserver reliability of the modified Tardieu scale in the upper limb of children with hemiplegia. Developmental Medicine & Child Neurology. 46(4):267-72.

Maitland, GD. (1991) Peripheral Manipulation. 3rd Edition. Butterworth-Heinemann Ltd. London. p 23.

Macnicol, MF. (1995) The Problem Knee. 2nd Edition. Butterworth-Heinemann Ltd. London. p24.

Manly, BFJ., Hopkins, WG. (1989) Errors in assigning grades based on tests of finite validity. Research Quarterly in Exercise in Sport. 60: 180-2.

Marks, R. (1996) Further evidence of impaired position sense in *knee* osteoarthritis. Physiotherapy Research International. 1 (2): 127-36.

Marks, R., Karkouti, E. (1996) Evaluation of the reliability of reflective marker placements. Physiotherapy Research International. 1 (1): 50-61.

Marks, R., Quinney, AH. (1997) From the field. Position sense perception at the *knee* in healthy persons and persons with chronic degenerative arthritis. Clinical Kinesiology: Journal of the American Kinesiotherapy Association. 50 (4): 76-81.

Mayerson, NH., Milano, RA. (1984) Goniometric Measurement Reliability in Physical Medicine. Archives of Physical Medicine & Rehabilitation. 65 (2): 92-94.

McMinn, RMH. (1994) Last Anatomy: Regional and Applied. 9th Edition. London: Churchill Livingston: 176-184.

Melby, A. 3rd., Askew, MJ. (1991) Evaluation of an electrogoniometric instrument for measurement of laxity of the knee. [letter; comment]. Journal of Bone & Joint Surgery – American. 73 (9): 1429.

Miyaska, KC., Daniel, DM., Stone, ML., Hirshman, P. (1991) The incidence of knee ligament injuries in the general population. American Journal of Knee Surgery. (4): 3-8.

Moore, KL. (1992) The Upper Limb, Clinically Oriented Anatomy. 3rd Edition. Edited by Satterfield, TS., Napora, L., Lumpkin, K. Williams & Wilkins.

Morris, S. (2002) Ashworth and Tardieu Scales: their clinical relevance for measuring spasticity in adult and paediatric neurological populations. Journal of Ethnopharmacology. 7(1):53-62.

National Highway Traffic Safety Administration. Vehicle Research and Test Centre. Pedestrian and Applied Biomechanics Division. http://www.nhtsa.dot.gov. Cited: September 16, 2003.

National Instruments Corporation. (2002) http://zone.ni.com/devzone/conceptd.nsf/webmain/ba741c90a118ea778625685e00805643?OpenDocument#7. Cited: 08/08/02.

Nicol, AC. (1989) Measurement of joint motion. Clinical Rehabilitation. (3): 1-9.

Nigel Keenan, W., Rodda, J., Wolfe, R., Roberts, S., Borton, DC., Kerr Graham, H. (2004) The static examination of children and young adults with cerebral palsy in the gait analysis laboratory: technique and observer agreement. Journal of Pediatric Orthopaedics B. (13)1.

Nordin, M., Frankel, VH. (1989) Basic Biomechanics of the Musculoskeletal System. 2nd Edition. Lea and Febiger: Philadelphia 115-134.

Norkin, C., Levangie PK. (1992) Joint Structure and Function. 2nd Edition. Philadelphia: F.A. Davis Company.

Norkin, C., White, DJ. (1995) Measurement of Joint Motion: A Guide to Goniometry. 2nd Edition. Philadelphia. F.A. Davis and Co.

Norkin, C., White, DJ. (2003) Measurement of Joint Motion: A Guide to Goniometry. 3rd Spiral edition. Philadelphia. F.A. Davis Company.

Oakley J. (2003) Digital Imaging. A Primer for Radiographers Radiologists and Health Professionals. Greenwich Medical Media Limit, London.

Oginni, LM., Olalekan, S., Badru, MB., Sharp, CA., Davie, MWJ., Worsfold, M. (2004) Knee Angles and Rickets in Nigerian Children. Journal of Pediatric Orthopaedics. 24(4).

Ojima, H., Miyake, S., Kumashiro, M., Togmai, H., Suzuki, K. (1991) Dynamic analysis of wrist circumduction : a new application of the biaxial flexible electrogoniometer. Clinical Biomechanics. 6: 221-229.

Orthoteers. (2002) http://www.orthoteers.co.uk/Nrujp~ij33lm/Orthelbowinjpaed1.htm. Cited Jan 2003

Palmer, JM. (1999) Radiometry and photometry FAQ. www.optics.arizona.edu/Palmer/. Cited: February, 2002. Version 1.01 (6).

Paquet, N., Malouin, F., Richards, CL., Dionne, JP., Comeau, F. (1991) Validity and reliability of a new electrogoniometer for the measurement of sagittal dorsolumbar movements. Spine. 16 (5): 516-9.

Peat, G., McCarney, R., Croft, P. (2001) Knee pain and osteoarthritis in older adults: a review of community burden and current use of primary health care. Annals of the Rheumatic Diseases. 60 (2): 91-7.

Petherick, M., Rheault, W., Kimble, S., Lechner, C., Senear, V. (1988) Concurrent Validity and Intertester Reliability of Universal and Fluid-based Goniometers for Active Range of Motion. Physical Therapy. 68:966-969.

Petty, NJ., Moore, AP. (1998) Neuromuscularskeletal Examination and Assessment. A Handbook for Therapists. Churchill Livingstone: London. pp 319.

Piazza, SJ., Cavanagh, PR. (2000) Measurement of the screw-home motion of the knee is sensitive to errors in axis alignment. Journal of Biomechanics. 33: 1029-34.

Pope, C., Mays (1998). Qualitative interviews in health research. 2nd Edition. British Medical Journal Publishing Group: London.

Popovic, N., Lemaire, R. (2002) Hyperextension trauma to the elbow: radiological and ultrasonographic evaluation in handball goalkeepers. British Journal of Sports Medicine. 36(6):452-6.

Portney, L. G., & Watkins, M. P. (1993). Foundations of clinical research: Applications to practice. Norwalk, Connecticut: Appleton & Lange.

Raine, S., Twomey, L. (1994) Posture of the head, shoulders and thoracic spine in comfortable erect standing. Australian Journal of Physiotherapy. 40 (1): 25-32.

Raine, S., Twomey, LT. (1997) Head and shoulder posture variations in 160 asymptomatic women and men. Archives of Physical Medicine & Rehabilitation. 78 (11): 1215-23.

Rankin, G., Stokes, M. (1998) Reliability of assessment tools in rehabilitation: an illustration of appropriate statistical analyses. Clinical Rehabilitation. 12: 187-199.

Rawes, ML., Richardson, JB., Dias, JJ. (1996) A new technique for the assessment of wrist movement using a biaxial flexible electrogoniometer. Journal of Hand Surgery. (B &E Volume) 21B: 5: 600-603.

Refshauge, K., Goodsell, M., Lee, M. (1994) Consistency of cervical and cervicothoracic posture in standing. Australian Journal of Physiotherapy. 40 (4): 235-40.

Remondino, F. (2001) History of Photography. http://www.isprs.org/society/history.html. Cited: 11/02/02.

Rheault, W., Miller, M., Nothnagel, P., Straessle, J., Urban, D.(1988) Intertester reliability and concurrent validity of fluid-based and universal goniometers for active knee flexion. Physical Therapy. 68 (11): 1676-8.

Riddle, DL., Rothstein, JM., Lamb, RL. (1987) Goniometric reliability in a clinical setting. Shoulder measurements. Physical Therapy. 67 (5): 668-73.

Roach, KE., Miles, TP. (1991) Normal Hip and Knee Active Range of Motion: The Relationship to Age. Physical Therapy. (71): 656-665.

Romanes, GJ. (1987) Cunningham's Manual of Practical Anatomy. 15th Ed. Upper and Lower Limb. Oxford: Oxford Medical Publications: 214-222.

Rome, K., Cowieson, F. (1996) A reliability study of the universal goniometer, fluid goniometer, and electrogoniometer for the measurement of ankle dorsiflexion. Foot & Ankle International. 17 (1): 28-32.

Rothstein, JM., Miller, PJ., Roettger, RF. (1983) Goniometric Reliability in a Clinical Setting. Elbow and Knee Measurements. Physical Therapy. 63: 1611-1615.

Rothstein, JM. (2001) Sick and Tired of Reilability. Physical Therapy. http://www.ptjournal.org/Feb01/FebruaryEdNote.htm. Cited 02 – 07 - 2001.

Rothstein J. (2003) Living with Error. Physical Therapy. May (5).

Rowe, PJ., Nicol, AC., Kelly, IG. (1989) Flexible goniometer computer system for the assessment of hip function. Clinical Biomechanics. 4 (2): 68-72.

Rowe, PJ., Myles, CM., Hillmann, SJ., Hazelwood, ME. (2001) Validation of Flexible Electrogoniometry as a Measure of Joint Kinematics. Physiotherapy. 87 (9): 479-488.

Russell, TG., Jull, GA., Wootton, R. (2003) Can the Internet be used as a medium to evaluate knee angle? Manual Therapy. 8(4) 242-246.

Sandelowski, M. (2000) Thermometers & Telephones. American Journal of Nursing. 100(10) 82-86.

Sanfridsson, J., Svahn, G., Jonsson, K., Ryd, L. (1997) Computed radiography for characterisation of the weight-bearing knee. Acta Radiologica. 38 (4 Pt 1): 514-9.

Sati, M., de Guise, JA., Larouche, S., Drouin, G. (1996) Quantitative assessment of skin-bone movement at the knee. The Knee. (3): 121-138.

Scholz, JP. (1989) Reliability and validity of the WATSMART three-dimensional optoelectric motion analysis system. Physical Therapy. 69 (8): 679-89.

Selfe, J. (1998) Validity and Reliability of the measurements taken by the Peak 5 motion analysis system. Journal of Medical Engineering & Technology. 22 (5): 200-5.

Shakhatreh, FM. (2001) Reproductive health of male radiographers. Saudi Medical Journal. 22 (2): 150-2.

Sherwood, L. (1997) Human Physiology: from calls to systems. 3rd Edition. Wadsworth publishing Co., California, USA.

Shircore, L., Love, S., Gibson, N., Price, C. (2004) Tardieu Scale to measure spasticity in children with cerebral palsy. Proceeding of the 5th National (Australian) Paediatric Physiotherapy Conference.

Shrout, PE., Fleiss, JL. (1979) "Intraclass correlations: uses in assessing rater reliability," Psychological Bulletin. 86, 420-428.

Sim, J., Arnell, P. (1993) Measurement validity in physical therapy research. Physical Therapy. 73 (2): 102-10.

Soames, RW. (ed.) (1995) Skeletal System. Gray's Anatomy. 38th Edition. New York: Churchill Livingstone.

Soames, R. (2003) Joint Motion, Clinical Measurement and Evaluation. 1st Edition. Churchill Livingstone. London.

Sprigle, S., Flinn, N., Wootten, M., McCorry, S. (2003) Development and testing of a pelvic goniometer designed to measure pelvic tilt and hip flexion. Clinical Biomechanics. (18) 462-465.

Startford, P., Agostino, V., Brazeau, C., Gowitzke, B. (1984) Reliability of joint measurement. A discussion of methodological issues. Physiotherapy Canada. 36 (1): 5-9.

Stergiou, N. (2004) Innovative Analyses of Human Movement. Human Kinetics. Sheridan Books, Champaign Il.

Stokdijk, M., Meskers, CG., Veeger, HE., de Boer, YA.(1999) Rozing PM. Determination of the optimal elbow axis for evaluation of placement of prostheses. Clinical Biomechanics. 14(3):177-84.

Stokdijk, M., Nagels, J., Garling, EH., Rozing, PM. (2003) The kinematic elbow axis as a parameter to evaluate total elbow replacement: A cadaver study of the iBP elbow system. Journal of Shoulder & Elbow Surgery. 12(1):63-8.

Stratford, P., Agostino, V., Brazeau, C., Gowitzke, BA. (1984) Reliability of joint angle measurement: a discussion of methodology issues. Physiotherapy Canada. 36(1): 5-9.

Struass, A., Corbin, J. (1990). Basics of qualitative research, techniques and procedures for developing grounded theory. Sage Publication: London.

Stude, DE., Goertz, C., Gallinger, M. (1994) Inter- and intraexaminer reliability of a single, digital inclinometric range of motion measurement technique in the assessment of lumbar range of motion. Journal of Manipulative & Physiological Therapeutics.
17 (2): 83-7.

Szulc, P., Lewandowski, J., Marecki, B. (2000) Verification of selected anatomical landmarks used as reference points for universal goniometer positioning during knee joint mobility range measurements. Diagnostic and Medical Technology. Medical Science Monit. 7 (2): 312-5.

Thomson, LC., Handoll, HHG., Cunningham, A., Shaw, PC., Van Herp, G. (2001) The Cochrane Library, The Cochrane Collaboration., 2000.

Thorp, J. (1998) The Information Paradox: Realizing the Business Benefits of Information Technology. http://www.webopedia.com. Cited: Jan 13, 2003.

Tilley, AR. (2001) The Measure of Man and Woman. Revised Ed. Henry Dreyfuss Associates.

Tokumura, F., Ohyama, K., Fujisawa, H., Matsuda, T., Kitazaki, Y.(1997) Conformability and irritancy of adhesive tapes on the skin. Contact Dermatitis. 37 (4): 173-8.

Tomsich, DA., Nitz, AJ., Threlkeld, AJ. (1996) Shapiro R. Patellofemoral alignment: reliability. Journal of Orthopaedic & Sports Physical Therapy. 23 (3): 200-8.

Trent, PS., Walker, PS., Wolf, B. (1976) Ligament length patterns, strength and rotational axes of the knee joint. Clinical Orthopaedics & Related Research. 263-270.

Vander Linden, DW., Carlson, SJ., Hubbard, RL. (1992) Reproducibility and Accuracy of Angle Measurement Obtained Under Static Conditions with the Motion Analysis Video System. Physical Therapy. 72 (4): 300-5.

Walker, JM., Sue, D., Miles-Elkousy, N., Ford, G., Trevelyan, H. (1984) Active Mobility of the Extremities in Older Subjects. Physical Therapy. (6)64. 919-923.

Wallen, MA., O'Flaherty, SJ., Waugh, MA. (2004) Functional outcomes of intramuscular botulinum toxin type A in the upper limbs of children with cerebral palsy: a phase II trial. Archives of Physical Medicine and Rehabilitation. 85(2):192-200, 351-2.

Walmsley, RP., Tyler, KA.(1996) The application and interpretation of the intraclass correlation in the assessment of reliability in isokinetic dynamometry. Isokinetics and Exercise Science. (6): 117-124.

Watkins, MA., Riddle, DL., Lamb, RL., Personius, WJ. (1991) Reliability of goniometric measurements and visual estimates of knee range of motion obtained in a clinical setting. Physical Therapy. 71(2): 90-6; discussion 96-7.

Whatis.com (2000). http://www.whatis.com. Cited: 03-03-2001.

Wiliams, JG., Callaghan, M. (1990) Comparison of Visual Estimation and Goniometry in Determination of a Shoulder Joint Angle. Physiotherapy. 76(10), 655-657.

Wilson, DJ., Smith, BK., Gibson, JK. (1997) Accuracy of Reconstructed Angular Estimates Obtained with the Ariel Performance Analysis System. Physical Therapy. 77 (12): 1741-6.

Xudong, Z., Sang-Wook, L., Braido, P. (2003) Determining finger segmental centres of rotation in flexion-extension based on surface marker measurement. Journal of Biomechanics. (36) 1097-1102

Youdas, JW., Bogard, CL., Suman, VJ. (1993) Reliability of goniometric measurements and visual estimates of ankle joint active range of motion obtained in a clinical setting. Archives of Physical Medicine & Rehabilitation. 74 (10):1113-8.

Zar, S. (1984) Biostsistical Analysis. 2nd Edition. Englewood Cliffs, New Jersey: Prentice Hall.