NIH Public Access
Author Manuscript
Value Health. Author manuscript; available in PMC 2013 October 29.

Published in final edited form as:
Value Health. 2011 ; 14(6): . doi:10.1016/j.jval.2011.03.002.

The Impact of a Revised EQ-5D Population Scoring on Preference-based Utility Scores in an Inflammatory Arthritis cohort

R Adams[1], **BM Craig**[2], **D Veale**[4], **B Bresnihan**[4], **C Walsh**[1,3], **O FitzGerald**[4], and **M Barry**[1]

[1]National Centre for Pharmacoeconomics, St. James Hospital, Dublin 8, Ireland [2]Health Outcomes & Behavior Program, Moffitt Cancer Center, Tampa, Florida [3]Department of Statistics, Trinity College Dublin, Dublin 2, Ireland [4]St. Vincents University Hospital, Dublin Academic Healthcare, Dublin 4, Ireland

Abstract

It is well established that there are problems with the EQ-5D. This is in part due to the scoring methods used for the original UK EQ-5D (TTO) and in particular how negative-Time Trade Off (TTO) values were treated. [1–3] A revised scoring method has been published and this paper has used this to calculate utility estimates in an inflammatory arthritis cohort.

Objective—To examine the impact of a revised scoring system for the EQ-5D(UK) TTO on the utility estimates and compare these to estimates produced using the original scoring and the SF-6D. In the case of RA, to explore the impact of using different utility metrics on the on the ICER results of an economic model.

Methods—504 patients with inflammatory arthritis were rescored using a revised scoring system for the EQ-5D, which uses an episodic random utility model to deal with negative TTO values. Differences in utility scores were compared and the mapping coefficients were used in an economic model, to examine the impact on the ICER estimate.

Results—In rheumatoid arthritis the overall change is less for the revised EQ-5D scoring than with the original EQ-5D (TTO) but greater than the SF-6D: EQ-5D UK −0.22 (95% CI −0.30, −0.15), revised EQ-5D UK −0.16 (95% CI −0.21, −0.10) and SF-6D −0.08 (95% CI −0.11, −0.05). A similar trend is seen in the psoriatic arthritis group. The economic model produced different ICERS, when different utility measures were used; EQ-5D (TTO) €42,402, SF-6D €111,788 and revised EQ-5D (TTO) €57,747.

Conclusion—In the context of inflammatory arthritis this article demonstrates that choice of utility measure may impact significantly on the output of the economic model and the subsequent reimbursement decision. In order to examine the heterogeneity between utility measures it may be useful to refit a cost effectiveness model using multiple metrics and produce a range of ICER estimates, to explore the uncertainty due to the choice of utility measure used.

Introduction

Rheumatoid arthritis (RA) and psoriatic arthritis (PsA) are part of a group of conditions described as inflammatory arthritis. These are a chronic, progressive group of conditions that place a substantial burden on patients, their caregivers and the health service. The most common of these conditions are RA and PsA. RA, in particular, has a negative effect on quality of life (QOL), including physical, psychological and social functioning and is associated with premature mortality.[4] PsA presentation ranges from mild impairment of function that can be adequately treated with mild therapeutic interventions to severe disease

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

with erosive arthropathy that may result in significant functional disability and increased mortality.[5]

Measurement of patients' response to treatment via QOL instruments is now one of the recommended methods for quantifying the effectiveness of a technology in economic evaluations around the world. [6–9] QOL is quantified using a single index figure anchored between 0 and 1, with 0 indicating immediate "death" and 1 equating to "perfect health". This figure has been termed a *utility value*. QOL can be measured using direct and indirect methods. Direct measurement asks a person to value directly their own health or a relevant disease-specific state using a valuation task such as time trade off (TTO) or standard gamble (SG). Indirect measurement involves QOL questionnaires, for example EQ-5D or SF-6D, where the patient completes questions on their current health state and these responses are scored using weights or preferences obtained from the general population. There are documented differences in the utility values produced in these ways with indirect methods giving consistently lower levels of utility than direct methods.[10] Indirect generic measures such as the EQ-5D[11] and the SF-6D[12] are the most commonly used and a single index measurement of quality of life can be derived.

A variety of instruments have been used to measure both disease severity in RA and PsA and the impact of this severity on QOL, particularly in inflammatory arthritis trials. [13–16]These measures include clinical instruments that measure disease activity, such as the European League Against Arthritis (EULAR) disease activity score (DAS)[17], disease-specific instruments such as the Health Assessment Questionnaire (HAQ),[18] and generic instruments such as the EQ-5D[11] or SF-36.[19]

All instruments display some shortcomings in assessing health-related QOL (HR-QOL) in inflammatory arthritis. [20, 21] While using generic measures should theoretically allow us to compare results for a variety of different conditions, disparities have been shown to exist in the utilities derived from the EQ-5D and SF-6D, and this is attributed to the different descriptive systems, the valuations attached to the health states or a combination of these. [22] This has important implications for economic analyses of treatments such as biological therapy in inflammatory arthritis, which is more likely to be used for patients in severe health states than those in mild health states. If we cannot measure the change within these categories accurately, the full potential of the treatment may not be adequately measured.

It is well established that the EQ-5D and the SF-6D produce different utility values in the same cohort.[23, 24] This is in part due to different definitions of perfect health. According to the 1995 Health Survey of England, the EQ-5D considers over half of the population to be in perfect health, while the SF-6D considers less than 3% to be in perfect health. [25] Therefore the SF-6D has a different criterion for perfect health than the EQ-5D. This presents decision makers with a challenge in comparing results of economic evaluations which have used different methods to calculate utility. In a RA cohort the utility gain produced by the EQ-5D is twice that produced by the SF-6D[23, 26]. This discrepancy between the measures has been the subject of a number of recent publications which highlight [1, 23, 27–30] the methodology of the original scoring of the EQ-5D(UK) and the manner in which *worse-than-dead* (WTD) values were adjusted.[31] Using an inflammatory arthritis cohort, a recent study found that the floor utility value measured by the EQ-5D was -0.43 and that 17% of utility values in this cohort were WTD. The prevalence of negative values had a profound effect on the burden of disease estimates in this paper.[23] The lowest measured value in the group for SF-6D was 0.29.

In an attempt to lessen the heterogeneity between these two measures, we have used a revised scoring method for the EQ-5D and rescored a rheumatology cohort receiving

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

biological therapy. We compared the scores, to the utility values produced by the original scoring methods for both RA and PsA.[29] We then used the mapping co-efficients calculated for RA in a rheumatoid arthritis model and compared the ICER results using different measures. The aim of our analysis is to examine the impact of a revised scoring system on the utility estimates and in the case of RA, on the results of an economic model.

Methods

Data Source

Utility data were derived from a database of 504 patients from a tertiary referral centre in the Irish healthcare setting, which records the clinical and QOL outcomes of patients receiving biological therapy for RA and PsA. Patients included in this study have a diagnosis of either RA according to the ACR criteria or PsA according to CASPAR (ClASsification criteria for Psoriatic Arthritis) criteria and commenced biological therapy (either anti-tumour necrosis factor [TNF]-α monoclonal antibodies, B-cell antagonists or T-cell modulators).

Instruments Used

The QOL instruments used were the paper versions of the EQ-5D, SF-36 (version 1) and the modified HAQ. The EQ-5D, SF-36 and HAQ were collected as part of normal clinical practice for monitoring the impact of treatment on QOL. All questionnaires were measured at baseline prior to the commencement of therapy and at follow-up, at 12 months.

The DAS 28 was collected as one of the clinical outcomes in monitoring response to treatment and disease activity.[17] This instrument incorporates the number of both swollen and tender joints, a laboratory measure of inflammation such as the erythrocyte sedimentation rate (ESR) or C-reactive protein (CRP), and a patient-assessed global disease impact measure. The DAS 28 score first developed for RA incorporates just 28 of the 68 joints. The DAS 28 is an index that can assess disease activity and also be used to derive a response measure.[32]

The HAQ is the physical disability scale (Modified HAQ), which measures function in relation to the degree of difficulty experienced in performing activities of daily living such as dressing, rising, personal hygiene, walking, eating and ability to carry out chores. The HAQ contains 20 items across eight domains, which are scored from 0 (no difficulty) to 3 (unable to do). [18] The HAQ disability index (HAQDI) is the single index score derived from scoring the HAQ.

The SF-6D is derived from the SF-36 and uses 11 items from the 36 items of the RAND Medical Outcomes Study short form health survey (SF-36)[12, 19]. Scoring data for the SF-6D were collected using the Standard Gamble (SG) valuation technique on a random sample (n=836) of the general population in the UK.[12] The SF-6D scoring algorithm was revised in 2007 using non-parametric Bayesian analysis.[33] The Bayesian version overcomes some of the bias of the original regression models when assigning values to the worst health states (e.g. it yields a value of 0.203 for the worst SF-6D state compared with 0.301 using the original parametric algorithm). The Bayesian estimates of the SF-6D utilities were calculated in Microsoft® Excel (further details of the methodology are availablefrom the University of Sheffield).[34]

Methods Used to Calculate Utilities

The EQ-5D index is a preference-based index measure, in which an individual provides an assessment of each component of his/her health status according to a structured health-status classification system, and a single preference-based score is derived for each individual

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

based on societal preferences.[35] It is used extensively to measure QOL in inflammatory arthritis.[11, 36–38] The EQ-5D has 5 dimensions, each with three levels of severity (Level 1=no problems to level 3= extreme problems). Therefore the instrument can produce 243 health states, 3^5. An additional two health states are included: “dead” and “unconscious”.

Scoring Method for the UK TTO

The preferences for the scoring function were measured using the time trade-off (TTO) technique on a random sample of 2997 adults of the UK population (Measurement of Health (MVH) study).[31] Dolan et al. devised a scoring method which assigned a single index utility value for each health state described.[31] Forty five of the health states were scored directly from the population using TTO valuation and the values of the remaining states were predicted using regression estimates.

To anchor the scale, perfect health and dead were assigned scores of 1 and 0, respectively. For states described as *better-than-dead* (BTD) (>0) on the TTO, scores were calculated using the formula $^{x}/10$ where x is the number of years spent in perfect health equal to 10 years in the health state. For states scored as WTD (<0) the formula given is $^{-x}/(10\text{-}x)$ where immediate death equates to a scenario of x years in perfect health followed by (10-x) years in the health state. For states BTD, the ratio range from 1 to 0 but ratios for WTD states lie between 0 and −39 (the WTD x has an upper bound at 9.75 years). The asymmetry seen between the positive and negative ratios seem to inflate the influence of the WTD responses; therefore Dolan transformed the negative ratios to −x/10, replacing 34% of the TTO responses.[39] By bounding the negative ratio at −1, the influence of these WTD responses on the mean slope lessened and improved face validity of mean ratio estimates.

Revised Scoring Method for the EQ-5D UK

In order to provide an alternative method for handling the challenges posed in valuing the WTD states in the Measurement of Health (MVH) study, Craig *et al.* re-examined the original data using a episodic regression model instead of a ratio regression model.[29] The health state valuations have been published and these are provided in Appendix 1. [2] The theoretical basis for both models was presented in a previous published paper. [40] The utility of a health state, *j*, over time, *t*, for an individual, *i, is* random and can be represented by:

$$U_{ij}(t)=\begin{cases} \mu_j t+\varepsilon_{ij} & Episodic\ RUM \\ (\mu_j+\varepsilon_{ij})\,t=\mu_j t+\varepsilon_{ij} t & Instant\ RUM \end{cases}$$

The main distinguishing factor between these models is how the WTD TTO responses are interpreted and this differs greatly between the episodic and the instant. In the episodic random utility model (RUM), the error represents variability in the value of an episode (error associated with time). In the instant RUM the error represents variability of an instantaneous state, not the episode, and suggests a random slope with respect to time. The regression model of the episodic RUM treats the time in perfect health as the dependent variable and the time in the health state as the independent variable. The coefficient is the value estimator, not a mean ratio. The central advantage of the episodic RUM over the original approach is that this procedure does not involve arbitrary transformations of the WTD responses.

Statistical Analysis—Descriptive statistics were used to describe the baseline demographics; mean values, range and standard deviation (SD) are given. A paired sample t-test was used to compare the mean utility at baseline and at follow-up and the mean change

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

measured by the original EQ-5D UK TTO, the revised EQ-5D UK and the SF-6D. Confidence intervals (95%) are presented around the change in utility and the ICER estimates. Statistical analysis was completed using Statistical Package for Social Sciences (SPSS) Version 16.

Statistical Models Used—General linear models were fitted for each of the measures with HAQDI and DAS28.. Quadratic and higher dimensional models were examined but did not have a statistically significantly improved fit than the linear model. For each of the regression models standard errors, 95% confidence intervals and R^2 are shown.

ICER calculation

In order to calculate an ICER for the RA group we used the use the mapped co-efficients of the three dependent variables to populate a RA model, which is based on changes to patients' HAQDI scores. The model incorporates a linear equation to model the utility change as mapped from the HAQDI. While newer models now use a quadratic equation to describe the relationship between utility and HAQDI in this case there was no significant statistical difference between the quadratic model and the linear model.[41] The model was populated using Irish cost data. Our objective was to examine the change to the ICER when the utility estimates are changed.

Results

Patient Demographics

At baseline, the mean age at inclusion was 54 years for the RA cohort and 45 years for the PsA group (table 1). The average disease duration was similar (RA=12 years, PsA=11 years). The mean DAS 28 score was significantly higher in the RA group (5.39 [95% CI 5.16, 5.43]) than in the PsA group (4.91 [95% CI 4.65, 5.05]), as was the mean HAQDI (RA 1.3 [95% CI 1.26, 1.46] vs PsA 0.96 [95% CI 0.81, 1.08]).

The mean utility scores and standard deviations (SD) are provided for each of the three methods (Table 2). In the RA group, the overall change was less for the revised EQ-5D scoring than the original EQ-5D (TTO) but greater than the SF-6D.(table 2). The change was greater in the PsA group across all three methods and a similar trend between the scoring methods was seen as that in the RA group; greatest change produced when using the EQ-5D, less so with the revised method and considerably less so with the SF-6D.

In order to describe the relationship between the measures we fitted regression lines between each of the QOL instruments and the HAQDI and DAS28. The equations for the mapping, including the coefficients of the regression, are presented in Appendix 2. We present both a linear and a quadratic equation for each of the methods and HAQDI. As there was no statistical difference between the models we have used the more parsimonious linear model. We plotted the regression lines for the measures in order to investigate the relationship between them (figure 1).

The revised scoring for the EQ-5D lessens the gap between the SF-6D and the original EQ-5D (figure 1). The slope produced by the relationship between the HAQDI and the revised scoring is less steep than that produced by the HAQDI with the original scoring. The distribution of the utility score produced by the two methods of scoring for the EQ-5D differs. The marginal distribution of each of these measures is provided in figure 2. The distribution of the scores was narrower with the revised scoring system (range −0.143, 0.995) than the original EQ-5D scoring method (range −0.429, 1.0) in this cohort.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

We have plotted the results of both measures on a scatter plot. The line is the line of equality. The methods produce results for utility scores less than approximately and the magnitude of this difference is 0.5 with the original EQ-5D method. The difference in the values is approximately 0.25 for scores below 0.5. (figure 3)

Impact on incremental cost effectiveness ratio

The ICER for a biological agent is presented for each of the utility measures. The original EQ-5D mapping produced the lowest ICER, €42,402 with a 95% confidence interval of €36,837 to €52,061. The SF-6D mapping produced the highest ICER, €111,788, and lies outside the acceptable willingness to pay range for most decision makers. The 95% confidence interval was €105,154 to €141,665. The EQ-5D mapping using the revised scoring method fell between these two measures, €57,747, with a 95% confidence interval of €52,032 to €72,845.

Discussion

There is a substantial burden of evidence highlighting the problems associated with the EQ-5D; this evidence is primarily referring not to the instrument itself but to the preference based scoring method which was used to assign population weighted values to the raw TTO scores. [1, 2] The area of most concern with the original scoring method is how the WTD states were handled. Recent papers propose an alternative method to handle WTD states but to date it has not been demonstrated how this method could in practice alter the utility estimates and ultimately the results of an economic model. [2, 29]

In this paper, we present the utility estimates from a large cohort of rheumatology patients scored using this alternative scoring method for the EQ-5D (TTO). In doing so we provide a practical application of this revised method in an observational cohort of patients and demonstrate that the mean utility estimates produced by each of these methods differ considerably which in turn influence the estimates of the economic model (table 2).

The revised method used here handles the raw scores differently and as a result the distribution of scores observed when using the revised EQ-5D scoring method was narrower than the original method. [29] The lowest score in this cohort with the revised method is −0.143 and with the original method is −0.43.

The range of ICERs estimated using three different methods of utility measurement highlights the impact that utility has on the overall result in the case presented here. While a pragmatic approach may be to recommend that one utility measure (either directly measured using questionnaires or via mapping) is used for all economic evaluations, this may restrict our ability to explore uncertainty associated with this parameter. Probabilistic sensitivity analysis only explores uncertainty within the limits of the instrument measured. In order to examine the heterogeneity between utility measures it may be useful to refit a cost effectiveness model using multiple metrics and produce a range of ICER estimates.

This paper also compares the utility estimates of the SF-6D to those of the EQ-5D using both scoring methods. One of the main areas of concern for economic evaluations is the lack of concordance between the two main generic QOL instruments: the EQ-5D and the SF-6D. While there are well documented reasons for this, the problem remains that they produce quite different results in QALY estimation. [27, 42–44] It is unlikely that these instruments will ever produce a similar result because of their differing descriptive systems. According to the Health Survey of England, perfect health is largely absent from the SF-6D system and prevalent in the EQ-5D system.[45] A reasonable assumption is that the true estimate may lie somewhere between both measures. We can see that the revised scoring produces an

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

estimate that lies between the slope of the original EQ-5D and the SF-6D (figure 1) and produces a change that that is less than the original EQ-5D and more than the SF-6D (table 2).

This alternative scoring method is useful in corroborating some of the reasons for the discrepancies in the scoring of the EQ-5D that have been presented by other authors[29, 46]. Dolan et al. replaced the negative slopes with −x/10 while Shaw et al. (US valuation) divided the negative slopes by a constant (i.e., 39). [31, 47] The episodic RUM reduces dependence on these arbitrary adjustments that have been made to deal with WTD valuations and provides a more robust coefficient estimator.

It is noted that the proportion of states considered WTD by the original UK EQ-5D values is much higher in this population than in others[46] (17% WTD at baseline in the Irish cohort). The impact on the ICER is significant and it would be of benefit to establish if such change would be seen in other less severe disease states. Although the methodology for this rescoring is relatively new, the revised results are more concordant with the SF-6D predictions, suggesting convergent validity. While we can see how the approach changes the results in this cohort it would be of interest to examine the effect across a number of different geographical populations and for different diseases.

Finally the SF-6D valuation study examined standard gamble responses without WTD responses, except for one state (i.e. pits). The worse imaginable health state, “pits” was valued in a similar manner to the valuations in the Dolan paper[31]. In future research, it would be prudent to rescore the SF-6D values and compare the difference in results overall.

Conclusion and Recommendations

This paper presents the results, from a large cohort, of using an alternative method for scoring the EQ-5D, and examines the relationship between both the revised and original generic measures (EQ-5D and SF-6D) and disease measures (HAQDI and DAS28) in inflammatory arthritis.

Decision makers who are using utility measures should be aware of the impact that the instrument and its scoring has on the ICER. In the context of inflammatory arthritis this article demonstrates that choice of utility measure may have a significant effect on the ICER which may therefore impact on the reimbursement decision. In choosing just one QOL measure to produce a single ICER estimate we may be restricting our ability to fully explore the uncertainty within the final estimate of a cost effectiveness analysis.

References

1. Harrison MJ, Davies LM, Bansback NJ, Melanie JM, et al. Why Do Patients with Inflammatory Arthritis Often Score States “Worse than Death” on the EQ-5D? An Investigation of the EQ-5D Classification System. Value in Health. 2009; 12(6):1026–1034. [PubMed: 19473335]
2. Craig BM, Busschbach JJV. Toward a more universal approach in health valuation. Health Economics. 2010 Published online 02 Aug 2010.
3. Dolan P, Gudex C, Kind P, Williams A. The time trade-off method: results from a general population study. Health Econ. 1996; 5(2):141–54. [PubMed: 8733106]
4. Yelin E, Callahan L. National Arthritis Data Work Group. The economic cost and social psychological impact of musculoskeletal conditions. Arthritis Rheum. 1995; 38:1351–6. [PubMed: 7575685]
5. Ackermann C, Kavanaugh A. Economic burden of psoriatic arthritis. Pharmacoeconomics. 2008; 26(2):121–9. [PubMed: 18198932]
6. NICE. Guide to the Methods of Technology Appraisal. London: 2008.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

7. Canadian Agency for Drugs and Technologies in Health (CADTH). Guidelines for the Economic Evaluation of Health Technologies. Canada: 2006 Mar. 2009
8. The Norwegian Medicines Control Authority. Norwegian guidelines for pharmacoeconomic analysis in connection with applications for reimbursement. 2000. cited 2009 March Available from
9. Rovira J, Antoñanzas F. Economic analysis of health technologies and programmes. A Spanish proposal for methodological standardisation. Pharmacoeconomics. 1995; 8(3):245–52. [PubMed: 10155620]
10. Arnold D, Girling A, Stevens A, Lilford R. Comparison of direct and indirect methods of estimating health state utilities for resource allocation: review and empirical analysis. BMJ. 2009; 339(jul20_3):b2688. [PubMed: 22128393]
11. Hurst N, Kind P, Ruta D, Hunter M, Stubbings A. Measuring health-related quality of life in rheumatoid arthritis: validity, responsiveness and reliability of EuroQol (EQ-5D). Rheumatology. 1997; 36(5):551–559.
12. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. J Health Econ. 2002; 21(2):271–92. [PubMed: 11939242]
13. Mease PJ, Gladman DD, Ritchlin CT, Ruderman EM, Steinfeld SD, Choy EHS, Sharp JT, Ory PA, Perdok RJ, Weinberg MA. Adalimumab Effectiveness in Psoriatic Arthritis Trial Study Group. Adalimumab for the treatment of patients with moderately to severely active psoriatic arthritis: Results of a double-blind, randomized, placebo-controlled trial. Arthritis & Rheumatism. 2005; 52(10):3279–3289. [PubMed: 16200601]
14. Van Der Heijde D, Klareskog L, Singh A, Tornero J, et al. Patient reported outcomes in a trial of combination therapy with etanercept and methotrexate forrheumatoid arthritis: The TEMPO trial. Annals of the Rheumatic Diseases. 2006; 65(3):328. [PubMed: 16079172]
15. Heiberg MS, Kaufmann C, Rodevand E, Mikkelsen K, Koldingsnes W, Mowinckel P, Kvien TK. The comparative effectiveness of anti-TNF therapy and methotrexate in patients with psoriatic arthritis: 6 month results from a longitudinal, observational, multicentre study . Ann Rheum Dis. 2007; 66(8):1038–1042. [PubMed: 17213251]
16. Kimel M, Cifaldi M, Chen N, Revicki D. Adalimumab plus methotrexate improved SF-36 scores and reduced the effect of rheumatoid arthritis (RA) on work activity for patients with early RA. Journal of Rheumatology. 2008; 35(2):206. [PubMed: 18085743]
17. van Gestel AM, Prevoo MLL, van't Hof MA, van Rijswijk MH, et al. Development and validation of the european league against rheumatism response criteria for rheumatoid arthritis: Comparison with the preliminary american college of rheumatology and the world health organization/ international league against rheumatism criteria. Arthritis & Rheumatism. 1996; 39(1):34–40. [PubMed: 8546736]
18. Fries J, Spitz P, Young D. The dimensions of health outcomes: the health assessment questionnaire, disability and pain scales. J Rheumatol. 1982; 9(5):789–93. [PubMed: 7175852]
19. Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care. 1992; 30:473–483. [PubMed: 1593914]
20. Scott DL, Khoshaba B, Choy EH, Kingsley GH. Limited correlation between the Health Assessment Questionnaire (HAQ) and EuroQol in rheumatoid arthritis: questionable validity of deriving quality adjusted life years from HAQ. Ann Rheum Dis. 2007; 66(11):1534–1537. [PubMed: 17660223]
21. Marra CA, Rashidi AA, Guh D, Kopec JA, Abrahamowicz M, Esdaile JM, Brazier JE, Fortin PR, Anis AH. Are indirect utility measures reliable and responsive in rheumatoid arthritis patients? Quality of Life Research. 2005; 14(5):1333. [PubMed: 16047508]
22. Bryan S, Longworth L. Measuring health-related utility: why the disparity between EQ-5D and SF-6D? Eur J Health Econ. 2005; 6(3):253–60. [PubMed: 15968563]
23. Adams R, Walsh C, Veale DJ, Bresnihan B, et al. Understanding the relationship between the EQ-5D, SF-6D, HAQ and disease activity in inflammatory arthritis. Pharmacoeconomics. 2010; 28(6):477–487. [PubMed: 20465316]
24. Sach T, Barton G, Jenkinson C, Doherty M, et al. Comparing cost-utility estimates: does the choice of EQ-5D or SF-6D matter? Med Care. 2009; 47(8):889–94. [PubMed: 19584759]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

25. Williams, A. Discussion Paper. York: Centre for Health Economics, York Health Economics Consortium, NHS Centre for Reviews & Dissemination, University of York; 1995. A measurement and valuation of health: a chronicle.
26. Brennan A, Bansback N, Nixon R, Madan J, et al. Modelling the cost effectiveness of TNF-a inhibitors in the management of rheumatoid arthritis: Results form the British Society for Rheumatology Biologics Registry - An Independent Report. Rheumatology. 2007; 46:1345–1354. [PubMed: 17562686]
27. Harrison, MJ.; Lunt, M.; Bansback, N.; Verstappen, S., et al. The Validity of Estimating EQ-5D and SF-6D Utility Values from the HAQ. American College of Rheumatology Annual Meeting; San Francisco. October 24–29; 2008; 2008.
28. Barton G, Sach T, Doherty M, Avery A, et al. An assessment of the discriminative ability of the EQ-5Dindex, SF-6D, and EQ VAS, using sociodemographic factors and clinical conditions. The European Journal of Health Economics. 2008; 9(3):237. [PubMed: 17605057]
29. Craig B, Busschbach J. The episodic random utility model unifies time tradeoff and discrete choice approaches in health state valuation. Population Health Metrics. 2009; 7(1):3. [PubMed: 19144115]
30. Marra CA, Marion SA, Guh DP, Najafzadeh M, et al. Not all “quality-adjusted life years” are equal. Journal of Clinical Epidemiology. 2007; 60(6):616. [PubMed: 17493521]
31. Dolan P. Modeling Valuations for EuroQol Health States. Medical Care. 1997; 35(11):1095. [PubMed: 9366889]
32. van Riel P, van Gestel A, van de Putte L. Development and validation of response criteria in rheumatoid arthritis: steps towards an international consensus on prognostic markers. Br J Rheumatol. 1996; 35 (Suppl 2):4–7. [PubMed: 8810683]
33. Kharroubi S, Brazier JE, O’Hagan A. Modelling covariates for the SF-6D standard gamble health state preference data using a nonparametric Bayesian method. Soc Sci Med. 2007; 64(6):1242. [PubMed: 17157971]
34. University of Sheffield Health Economics and Decision Science. [Accessed 2010 June 1] Bayesian programme [online]. Available from URL: http://www.shef.ac.uk/scharr/sections/heds/mvh/sf-6d/bayesian.html
35. Hurst N, Jobanputra P, Hunter M, Lambert M, et al. Validity Of EuroQoL--A Generic Health Status Instrument--In Patients With Rheumatoid Arthritis: Economic And Health Outcomes Research Group. Rheumatology. 1994; 33(7):655–662.
36. Kekow J, Moots RJ, Emery P, Durez P, et al. Patient-reported outcomes improve with etanercept plus methotrexate in active early rheumatoid arthritis and the improvement is strongly associated with remission: The COMET trial. Ann Rheum Dis. 2009 Mar 16. Epub ahead of print.
37. Lindgren P, Geborek P, Kobelt G. Modelling the cost-effectiveness of treatment of rheumatoid arthritis with rituximab using registry data from Southern Sweden. Int J Technol Assess Health Care. 2009; 25(02):181–189. [PubMed: 19331709]
38. Marra C, Esdaile JM, Guh D, Kopec JA, Brazier JE, Koehler BE, Chalmers A, Anis AH. A comparison of four indirect methods of assessing utility values in rheumatoid arthritis. Med Care. 2004; 42(11):1125–31. [PubMed: 15586840]
39. Patrick DL, Starks HE, Cain KC, Uhlmann RF, Pearlman RA. Measuring preferences for health states worse than death. Med Decis Making. 1994; 14:9–18. [PubMed: 8152361]
40. Craig BM, Busschbach JJV, Salomon JA. Modeling Ranking, Time TradeOff, and Visual Analog Scale Values for EQ-5D Health States: A Review and Comparison of Methods. Medical Care. 2009; 47(6):634–641.10.1097/MLR.0b013e31819432ba [PubMed: 19433996]
41. National Institute for Clinical Excellence. Final appraisal determination – Adalimumab, etanercept, infliximab, rituximab and abatacept for the treatment of rheumatoid arthritis after the failure of a TNF inhibitor (part review of NICE technology appraisal guidance 36, review of NICE technology appraisal guidance 126 and 141. 2010. Available from: http://www.nice.org.uk/nicemedia/live/12135/49473/49473.pdf
42. Harrison, M.; Davies, L.; Bansback, N.; McCoy, M., et al. Quality of Life Research. 2009. The comparative responsiveness of the EQ-5D and SF-6D to change in patients with inflammatory arthritis.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

43. Brazier J, Roberts J, Tsuchiya A, Busschbach J. A comparison of the EQ-5D and SF-6D across seven patient groups. Health Economics. 2004; 13(9):873–884. [PubMed: 15362179]

44. Marra CA, Woolcott JC, Kopec JA, Shojania K, et al. A comparison of generic, indirect utility measures (the HUI2, HUI3, SF-6D, and the EQ-5D) and disease-specific instruments (the RAQoL and the HAQ) in rheumatoid arthritis. Social Science & Medicine. 2005; 60(7):1571. [PubMed: 15652688]

45. Prescott-Clarke, P.; Primatesta, P., editors. Vol 1: Findings; Vol 11: Survey methodology and documentation. London: The Stationary Office; 1998. Health Survey for England 1996.

46. Harrison MJ, Davies LM, Bansback NJ, McCoy MJ, et al. Why Do Patients with Inflammatory Arthritis Often Score States “Worse than Death” on the EQ-5D? An Investigation of the EQ-5D Classification System. Value Health. 2009 Epub 15 May 2009.

47. Shaw JW, Johnson JA, Coons SJ. US valuation of the EQ-5D health states: development and testing of the D1 valuation model. Med Care. 2005; 43:203–220. [PubMed: 15725977]

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

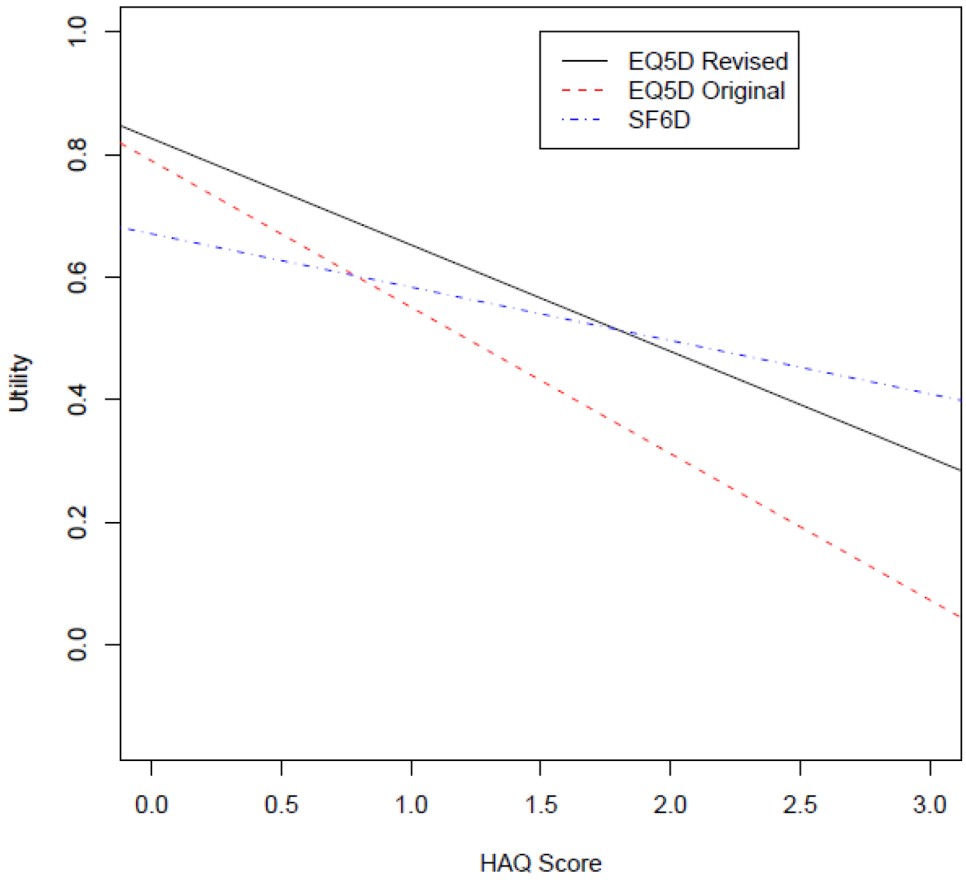

Fig 1.
Comparison between the fitted lines associating mapped utility from the revised EQ-5D UK scoring, original EQ-5D UK and the SF-6D with the HAQDI score.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

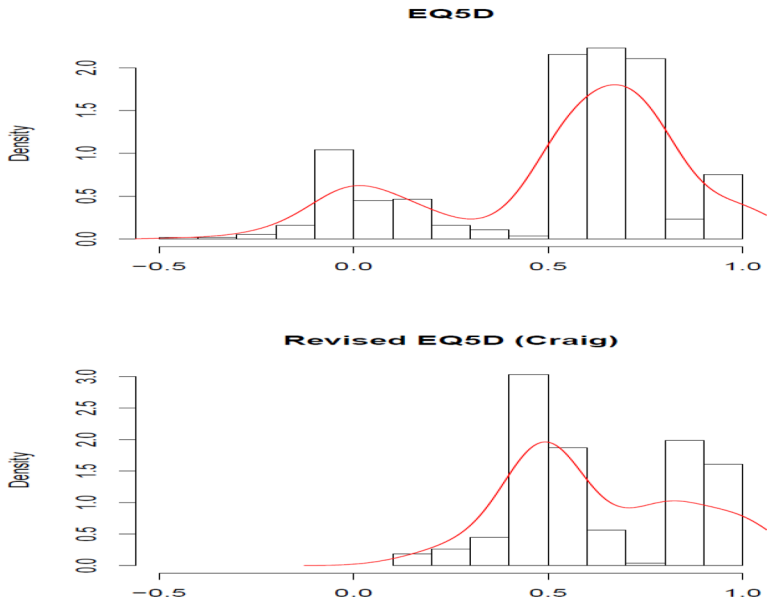

Fig. 2.

Histograms of the EQ-5D and revised EQ-5D showing the marginal distribution for each of these measures for this cohort. Of note in the revised version is the impact on individuals with values less than 0.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

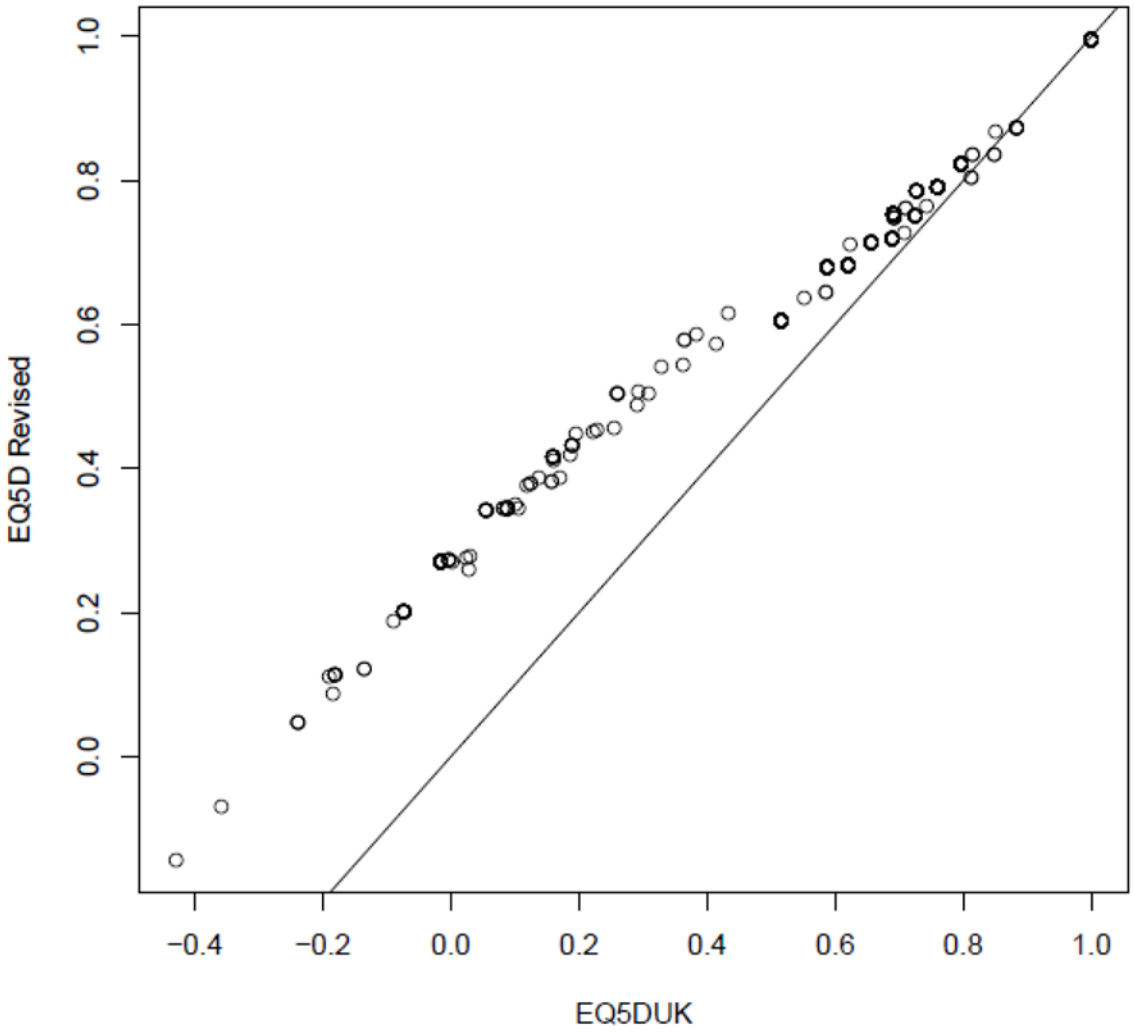

Fig 3.
Scatter plot of utilities derived from the revised EQ-5D scoring and original EQ-5D scoring. A line of equality is fitted.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 1

Baseline Demographics

Characteristic	RA (n = 345)	PsA (n = 159)
	mean ± SD (range)[a]	mean ± SD (range)[a]
Female sex (%)	245 (71%)	82 (52%)
Age at inclusion (y)	54 ± 12.9 (17, 85)	45 ± 12.8 (15, 77)
Duration of disease (y)	12 ± 9.4 (0, 42)	11 ± 10.1 (0, 45)
ESR	35 ± 25.8 (2, 140)	22 ± 21.1 (1, 120)
CRP	29 ± 29.5 (2, 158)	18 ± 22.7 (0, 149)
DAS 28: CRP	5.39 ± 1.18 (1, 9)	4.91 ± 1.0 (1, 7)
Patient Global Assessment (10 cm VAS)	6 ± 2.3 (0, 10)	5.5 ± 2.3 (0, 10)
Pain (10 cm VAS)	6 ± 2.3 (0, 10)	5 ± 2.3 (0, 10)
Tender joint count (range 0–28)	5 ± 2.3 (0, 28)	8 ± 6 (0, 28)
Fatigue (10 cm VAS)	6 ± 2.4 (0, 10)	6 ± 2.6 (0, 10)
Swollen joint count (range 0–28)	10 ± 6.6 (0, 25)	7 ± 6 (0, 28)
Tender joint count (range 0–66)	10 ± 6.0 (1, 10)	12 ± 9 (0, 43)
Concomitant methotrexate (n)	220 (64%)	56 (35%)
Previous DMARDs (n)	292 *	118 **
HAQDI (0–3)	1.3 ± 0.7 (0, 3)	0.96 ± 0.7 (0, 2.5)
SF-36 PCS (0–100)	30 ± 8.5 (12, 57)	34 ± 9.5 (13, 58)
SF-36 MCS (0–100)	45 ± 10.4 (17, 72)	46 ± 12.2 (20, 66)
SF-6D utility	0.54 ± 0.09 (0.3, 0.7)	0.57 ± 0.12 (0.25, 0.80)
EQ-5D UK TTO utility	0.43 ± 0.32 (−0.43, 1)	0.53 ± 0.32 (−0.24, 1)
Revised EQ-5D UK TTO utility	0.576 ± 0.22 (−0.14, 0.9954)	0.638 ± 0.19 (0.046, 0.9954)

[a] Unless otherwise indicated.

CRP = C-reactive protein; **DAS 28** = Disease Activity Score (28 joint); **DMARDs** = disease-modifying anti-rheumatic drugs; **ESR** = erythrocyte sedimentation rate; **HAQ** = Health Assessment Questionnaire; **HAQDI** = Health Assessment Questionnaire Disease Index; **MCS** = mental component summary; **PCS** = physical component summary; **PsA** = psoriatic arthritis; **RA** = rheumatoid arthritis; **VAS** = visual analogue scale;

* indicates missing data (n = 18 patients);

** indicates missing data (n = 4).

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 2

Mean Utility scores at baseline and follow-up

	RA			PsA		
	Baseline [mean ± SD (range)	**12 month [mean ± SD (range)**	**change in utility (CI 95%)]**	**Baseline [mean ± SD (range)**	**12 month [mean ± SD (range)**	**change in utility (CI 95%)]**
	0.54 ± 0.09 (0.29, 0.75)	0.62 ± 0.077 (0.44, 0.83)	0.08 (0.106, 0.049)	0.57 ± 0.12 (0.25, 0.79)	0.66 ± 0.12 (0.32, 0.89)	0.09 (0.123, 0.052)
EQ-5D	0.43 ± 0.32 (−0.43, 1.0)	0.65 ± 0.28 (−0.18, 1.0)	0.22 (0.302, 0.145)	0.49 ± 0.32 (−0.24, 1.0)	0.77 ± 0.28 (−0.24, 1.0)	0.28–0.360, 0.200)
Revised EQ-5D	0.56 ± 0.22 (−0.14,0.9954)	0.72 ± 0.2 (0.88,0.9954)	0.16 (0.214, 0.102)	0.62 ± 0.21 (−0.14,0.9954)	0.84 ± 0.17 (0.046, 0.9954)	0.22 (0.281, 0.167)