

Model specification in hierarchical meta analysis

C. Walsh*, K. Mengersen†

Abstract

We highlight the fact that careful consideration has to be given to the specification of the order of the hierarchy in hierarchical meta analysis. In practical applications, it can often be unclear as to which is the ‘correct’ way to specify these, and thus an arbitrary choice may be made when fitting the model. However, depending on how the model is specified, there may be important differences in the interpretation of the analysis. The points raised are highlighted using an example involving the estimation of bone density in an elderly population in a clinical setting.

1 INTRODUCTION

The general question of choice of a statistical model has, of course, consumed an immense literature. In an ideal world, a best model arises for philosophical, physical or biological reasons, or is a natural consequence of the study design. Often, however, the model structure is a pragmatic or empirical decision: a regression analysis of an observational study often involves the selection of a subset of available variables, and a multi-level analysis increases the dimensionality of the decision-making to include the number and composition of hierarchies. The choice thus depends on, and influences, the relative size and precision of the corresponding point estimates at the different hierarchical levels of the model [1, 2, 3] and the associative and causal inferences that are based on the analysis [4, 5].

The choice of the order of the hierarchies in a multi-level model is similarly influenced by a combination of theoretical, physical, design and empirical arguments. In the situation considered in this paper, it is assumed that the first two of these are not sufficiently informative to completely determine the model structure. Moreover, it is assumed that the influence of design, while partially informative, is diluted due to the multiplicity of studies involved in a meta-analysis, the more complex covariate structure arising from the meta-analysis framework and missing data that often arise in multivariate meta-analyses. Thus the importance of being able to estimate the covariance structure at the lower levels

*Trinity College Dublin, Dublin 2, Ireland

†School of Mathematical Sciences, Queensland University of Technology, GPO Box 2434, Brisbane, QLD 4001, Australia

of a hierarchical model, and the consequent concerns about bias and precision of estimation at these levels that have been raised by previous authors, [1, 4, 5], are applicable here.

As a general rule, in the absence of other information about model structure, greater precision for all estimates is obtained if the model is structured in such a way that the variance between objects within subgroups decreases as the hierarchy decreases. Thus, the first-level hierarchy should comprise those subgroups between which there is greatest variance, relative to the subsequent nested subgroups. At the other end, the objects (nested individuals within subgroups) at the lowest level of the hierarchy should be the most homogeneous of those in any of the higher-level subgroups. Similarly, a nested structure is only advantageous over a non-nested (block) structure if the variances of the corresponding estimates are smaller. This will occur if there is within-group correlation and relatively little missing data, especially relating to the model structure. Other authors [4, 6] discuss this issue in some detail.

Meta-analysis is now a standard and often imperative tool in comprehensive statistical analysis. The formal quantitative combination of results of interest from multiple sources can be employed for the purposes of study design or justification, motivation, literature review or contextual placement of new results.

A convenient framework for meta-analysis is an hierarchical model structure. Very often, estimates of interest are naturally nested within studies, which in turn may be nested within homogeneous subgroups of studies, and so on. The nesting within a hierarchy reflects ‘exchangeability’ of components within subgroups and ‘non-exchangeability’ across subgroups. For example, study populations might be considered to be more similar within regional groups or according to some other important covariate. Similarly, the hierarchies themselves reflect ‘nonexchangeability’; for example, if geography is important in its own right or as a surrogate for other differentiating descriptors, studies might be grouped within regions, which are in turn grouped within countries, which are in turn grouped together.

An important computational and inferential advantage of such an hierarchical structure is that it induces conditional independence. Thus elements within a group are considered to be independent conditional on their membership of that group, and estimation depends only on the connecting hierarchies. More general methods of path analysis, networks and graphical models rely on the same principle.

In many cases, however, the ordering of the hierarchy in a meta-analysis is not obvious. For example, suppose that multiple studies report multiple effects of interest; possibly, the effects are measured on different subgroups of subjects within each study. Thus there is correlation between measures within studies and there is also correlation between measures across studies. Hence possible hierarchical structures could be “measures combined within studies, then studies combined”, or “studies combined within measures, then measures combined”. In both cases, the full multivariate correlation structure is reduced, but the induced conditional independence structure is different and, as we show in this paper, the resultant estimates of effect are different as a consequence.

Note that in situations like the above, it might be preferable to consider a full multivariate meta-analysis instead of the suggested hierarchical structure. This approach has been taken in other situations [7]. However, the full multivariate analysis requires that sufficient information about correlation of estimates within and between hierarchies is reported. Typically, this information is not available in reported information available from published studies. Hence this option is not pursued further in this paper.

Without loss of generality, we consider here the problem described above, of combining multiple measurements across studies, where there is some uncertainty about the order of the hierarchy in the meta-analysis model. Features of the model are discussed in the next section, followed by a detailed illustration of the potential impact of the choice of hierarchical structure in a real example.

2 MODELS

Information about a particular condition or disease may be measured in different ways in different studies. Combining this information to estimate an overall effect within a given population of relevance is the function of meta-analysis.

Key features of relevance in this paper are that the type of measurement obtained within a given study may be slightly different to that in another study. Indeed it may be the case that two measurements of essentially the same underlying property are obtained within a single study. For example, in measuring cognitive function for a cohort, either the Mini Mental State Examination (MMSE) or Cambridge Neuropsychological Test Automated Battery (CANTAB) can be seen as a useful tool.

On occasion, it may be the case that two (or more) of the measures of a property of interest may be obtained within a single study, albeit in a situation where there are two subgroups or indeed it is not clear to what extent the groups overlapped within a study.

In combining information from multiple sources, the analyst has to make a number of choices. For example, in a frequentist framework, one choice the analyst would make is to decide between a fixed effects or random effects model, depending on whether one wishes to allow for heterogeneity between studies. In a Bayesian framework it is natural to consider fitting a hierarchical model that easily accommodates these and similar decisions.

Of specific interest here, is the situation where multiple measurements of different types, across different studies are going to be combined. In such a case, the structure of the model may not be obvious.

If it is the case that there is overlap between measurements, with many individuals being measured using multiple different techniques, then it would be natural to consider the measurement types being nested within studies. Indeed, in such a case, it would be possible to estimate the association between measures and continue along the multivariate meta analysis route as in [7].

If, however, the question of interest is about a standardised outcome measure, where one of a number of tools can be used in the data collection phase,

then it may be unclear whether measures should be combined firstly within studies, or whether the measures of the same type should be combined across studies before combining measures together at a higher level. This uncertainty in model specification, and the fact that it impacts materially on the conclusions of an analysis is the topic here.

2.1 Notation

Here we consider a situation where the mean size of a quantity of interest y_{ij} is measured for measurement type j from individuals within a study i . The precision of y_{ij} is available and denoted τ_{ij} . Thus y_{ij} is an estimate of the underlying measurable property of interest θ_{ij} .

2.2 Hierarchical Model

A hierarchical meta-analysis model might first combine effects within a given study to give an overall estimate for the mean size of the property of interest within a study θ_i , so that

$$\theta_{ij} \sim \theta_i, \tau_{sm}.$$

Similarly the studies may be combined in order to give an estimate of the overall mean size of the property of interest across studies, so that

$$\theta_i \sim \theta, \tau.$$

The specification of the model in this fashion allows for heterogeneity between *studies* to be modelled at the highest level, and the heterogeneity between the *measures* within a single study to be modelled at a lower level.

Alternatively the model may be described with the studies combined for each of the measures to give an overall estimate for the mean size of the measure across studies, θ_j .

Thus

$$\theta_{ij} \sim \theta_j, \tau_{ij}$$

and similarly the top level measure effects can be combined in order to give an estimate of the overall mean size of the property of interest. Thus

$$\theta_j \sim \theta, \tau.$$

The specification of the model in this fashion allows for heterogeneity between *measures* to be modelled at the highest level, and the heterogeneity between the *studies* for each of the measures to be modelled at a lower level.

It is noted that the models are different, that there is a different total number of parameters. In particular it may readily be shown (see Appendix) that the likelihoods for θ are, in general, different for each.

2.3 Multivariate model for multiple outcomes

In modelling a situation as described, where there are multiple outcomes, it may be desirable to specifically allow the measures to be related to one another (as opposed to conditionally independent of one another.)

This may be done by specifying a multivariate hierarchical model. In this case;

$$\begin{pmatrix} \theta_{i1} \\ \theta_{i2} \\ \theta_{i3} \end{pmatrix} \sim MVN(\Theta, \mathbf{D})$$

We note that the full specification of the model requires some understanding of the association between the measures that have been used. This information is not, in general, available, and for this reason the approach is not pursued in depth here. Work in this area has been done previously [5, 7, 8].

It is noted that by choosing a fully hierarchical model, rather than a multivariate one, the off diagonal elements of the matrix, \mathbf{D} , are set to zero. The impact of this in a practical sense is to assume exchangeability of the measures within each of the patient groupings; this is a reasonable assumption in this case, but is something that requires consideration for each situation.

3 EXAMPLE

3.1 Data

In order to prevent fractures in the elderly population, bone density screening is carried out for the at risk population in St James's Hospital, Dublin.

Three different types of measurement technique are of interest here.

The first technique is the estimation of bone density from DXA (dual energy x-ray absorptiometry) scans of the hip. This is a standard technique and requires that a patient undergo a low dose x-ray of their hips, the scan being carried out by an experienced operator.

The second technique is the use of a portable ultrasound machine to estimate the density of bones by scanning the heel of the individual. These devices are faster to use and scans may be carried out on patients in the ward without needing to transfer them to where a DXA machine is located.

The third method of interest is the estimate of bone density that may be obtained from an individual using a DXA scan of their spine. This scan involves capturing an image from a different angle and estimates are sensitive to features that may be apparent in the individual such as vertebral fractures or other anomalies.

In the data that have been recorded, the information has been obtained from a mixture of patients who attended as day visitors to the hospital and those who are inpatients, with the reason for attendance or admission unrelated to their bone density.

Thus, it may be considered that there are two study groups here; namely Group 1 being day-hospital attendees and Group 2 being inpatients.

Table 1: Summary statistics standardised mean (sem) from each measure for each study.

Group	Hip	Heel	Spine
1	-2.58 (0.14)	-3.07 (0.14)	-2.88 (0.17)
2	-1.87 (0.15)	-3.38 (0.45)	-1.81 (0.20)

There are three measures. Measure 1 is Hip DXA; Measure 2 is Heel US and Measure 3 is Spine DXA. Each of these measures returns T-Scores which are adjusted and standardised indications of the relative bone density of the individual.

Since the method of measurement and standardisation already allows for the fact that patients are of different ages and genders, it is not necessary to include these covariates in the models that we examine.

Summary statistics in the form of mean T-Score and standard error are available for each of the measures for each of the studies. However, within a given study, the measurements have not been carried out on the same group of individuals. The summary data are shown in table 1.

3.2 Model H1

We first considered a two level hierarchical random effects model. In this case, we posit that within each of the two groups, the three measures provide us with an estimate of average bone density within the group. Between measure variability is allowed for with a precision term at this level.

At the group level, the estimates for each group are combined to give an estimate of overall mean bone density for the hospital population. Variability between groups is allowed for at this level.

A schematic representation of the model is shown in Figure 1. This shows the ordering of the hierarchy. The actual parameterisation and the numbers of parameters at each level matches the example used for the case of the simulation.

Thus, the model may be written down in terms of the following conditional densities:

$$\begin{aligned}
 y_{ij} &\sim N(\theta_{ij}, \tau_{ij}), \\
 \theta_{ij} &\sim N(\theta_i, \tau_i), \\
 \theta_i &\sim N(\theta, \tau).
 \end{aligned}$$

The priors placed on $\tau, \tau_i, \tau_{ij}, \mu$, were non-informative, in order that the impact of model specification could be investigated.

3.3 Model H2

Here, the model may be written down in terms of the conditional densities in a fashion similar to the previous section, but with the nesting in terms of ij within j ;

$$\begin{aligned}y_{ij} &\sim N(\theta_{ij}, \tau_{ij}) \\ \theta_{ij} &\sim N(\theta_j, \tau_j) \\ \theta_j &\sim N(\theta, \tau)\end{aligned}$$

The priors placed on τ , τ_i , τ_{ij} , μ , were once again chosen to be non-informative.

3.4 Simulation Study

In order to examine the impact of model specification in situations where a larger number of studies is included, we carried out a simulation study.

In the simulation, a similar setup was used, where there are three measures which can be used in estimating the property of interest. For simplicity, we will assume that these are measures of bone density in a simulation setup. The number of studies examined in this instance was ten. We chose this size as it is compatible with the number of studies used in published meta analyses.

We set the overall mean at -2.5 (compatible with osteoporosis and observed values in the case of our data analysed in H1 and H2.) The mean effect for each of the studies was then simulated to be normally distributed about this, with precision 1.0. The mean size of the bone densities for each of the 3 measures was then simulated as normally distributed about these values with precision 1.0.

Having simulated a complete set of observations, the model was fit using WinBUGS 1.4.1., and the mean bone density and standard error (based on the standard deviation of the sampled values) were recorded.

This process was then repeated ten times in order to examine the impact of model specification allowing for sampling variability.

3.5 Analysis

The models were fitted using WinBUGS 1.4.1 [9] on an Intel Pentium 4, 3.0GHz with 1GB of RAM, running Windows XP Service Pack 2. Processing was done using R 2.1.1 [10] and the R2WinBUGS package [11]. In fitting these models, three chains were used, with a burn in of 2000 iterations and thinning of every second iteration. Convergence of chains was assessed using Coda [12].

Variable	Mean Posterior	2.5%	50.0%	97.5%
μ	-2.6	-4.9	-2.6	-0.3
τ	133.1	0.0	13.7	1089.0
μ_1	-2.7	-3.5	-2.7	-1.9
μ_2	-2.4	-3.2	-2.7	-1.7
μ_{11}	-2.6	-3.1	-2.6	-2.0
μ_{12}	-2.8	-3.3	-2.9	-2.1
μ_{13}	-3.0	-3.8	-3.2	-1.9
μ_{21}	-2.9	-3.3	-3.0	-2.1
μ_{22}	-2.0	-2.9	-1.9	-1.6
μ_{23}	-2.0	-2.9	-1.9	-1.5

Table 2: Summary statistics, point estimate and quantiles for sampled values of mean BMD for model H1.

4 RESULTS

4.1 Model H1

Summary statistics for the posterior distributions at the top and second level are provided in table 2.

As is the case in most meta-analyses, the key statistics of interest are those at the top level; in this case the mean bone density for the population.

Based on the quantiles from the posterior sample, a 95% credible interval (CI) for the mean BMD is (-4.9,-0.3). This is a rather wide interval; but is to be expected since the uncertainty in the overall estimate depends somewhat on the number of groups at the next level down.

The values of the estimates for the next level down have been included. We therefore have the mean of each of the measures for each of the groups in this case.

It is possible to obtain p_D and the DIC [13]. These are 168.8 and 161.9 respectively.

4.2 Model H2

Summaries from the posterior distributions are provided in table 3.

In this instance a 95% confidence interval for the mean BMD is (-3.5,-1.6). This interval is noticeably narrower, although centred at the same location as H1.

In this case, p_D was 162.1 with a corresponding DIC of 158.8.

The DIC is a little lower in this case, but the difference between the fits is small. One possible approach that could be taken is to weight the models according to the DIC. This is a standard approach one might take in the context of model averaging. In this instance, one may use $\exp(-textDIC/2)$ as the weight for each of the models [18], which is a practical approach for the purposes of prediction.

Variable	Mean Posterior	2.5%	50.0%	97.5%
μ	-2.5	-3.5	-2.5	-1.6
τ	179.1	0.3	34.8	1290.0
μ_1	-2.6	-3.4	-2.6	-1.8
μ_2	-2.5	-3.3	-2.5	-1.7
μ_3	-2.5	-3.4	-2.5	-1.6
μ_{11}	-2.6	-3.2	-2.6	-2.0
μ_{12}	-2.9	-3.4	-3.0	-2.1
μ_{21}	-2.7	-3.3	-2.8	-1.9
μ_{22}	-2.1	-3.0	-1.9	-1.6
μ_{31}	-2.9	-3.7	-3.0	-1.6
μ_{32}	-2.1	-3.1	-1.9	-1.4

Table 3: Summary statistics, point estimate and quantiles for sampled values of mean BMD for model H2.

Replicate	Mean Model 1	CI q.width	Mean Model 2	CI q.width
1	-2.0	0.34	-2.0	1.25
2	-2.5	0.33	-2.5	1.31
3	-2.0	0.32	-1.9	1.28
4	-2.8	0.31	-2.8	1.18
5	-2.5	0.34	-2.5	1.29
6	-2.2	0.35	-2.2	1.34
7	-2.5	0.29	-2.5	1.33
8	-2.7	0.47	-2.7	1.37
9	-2.3	0.32	-2.3	1.67
10	-2.2	0.33	-2.1	1.25

Table 4: Summary statistics, point estimate and CI quarter width for mean BMD for 10 replicates.

In practice, clinicians are interested in the ‘cutoff’ value of -1.5, since this is the level that represents osteopenia, a state of abnormally low bone density. Thus, depending on whether H1 or H2 is fitted, their qualitative interpretation of the findings will differ.

4.3 Simulation Study

In table 4, the summaries that are reported are the mean and CI quarter width; this q.width can be interpreted in the same way as the SEM is used in a classical sense.

It is clearly the case here that the CI width (although not its location) is affected by the ordering in the hierarchy. And, as mentioned above, the qualitative interpretation in a clinical context may be altered, depending on which of the two models is fitted.

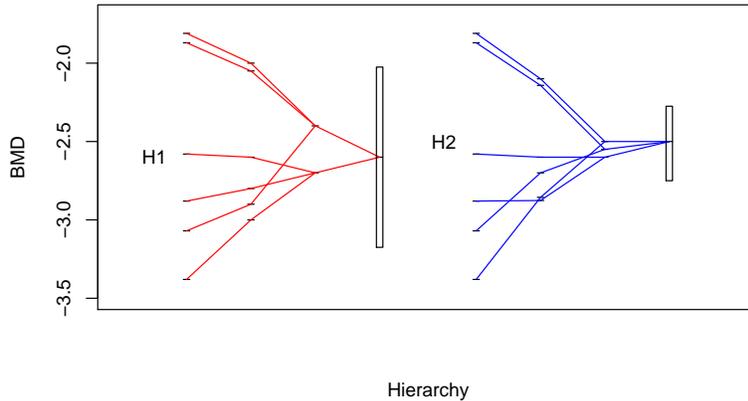


Figure 3: Plot showing the shrinkage from data, through each level to the estimate of overall mean.

5 DISCUSSION

This work demonstrates that in the case of a real example, and an associated simulation study, the way in which the model is specified is of material importance. Thus, if there is any confusion about which way the hierarchy should be arranged in these models, then care should be taken in the interpretation of the results.

Overall, the feature of note is that the hierarchical models that are fitted are not the same. The difference between the specification of the models are the assumptions about which parameters are exchangeable.

The estimate of the overall mean within the group of interest is the same, independent of the model that has been fitted. However, uncertainty in this estimate is different for the two models. The reason for this difference is simple - the two models are quite different; and the parameter being estimated at the top level is related to the next level down. Where the amount of ‘information’ at that level is different, then the uncertainty changes.

The impact of ordering is clear from figure 3. The box on the right hand side of each plot is proportional to the confidence interval width. In model H1, this is influenced by the fact that at the next level in the hierarchy (on the left), there are two parameters being estimated, and there is a modest difference between them. In H2, however, there is less variability between these parameters and there are more of them. As the results in table 3 show, the posterior means for μ_2 and μ_3 happen to be equal and thus are overplotted on the graph.

The example used involves multiple measurements in a single hospital. However, it could just as easily be the case that these ‘Groups’ are presented in

different studies, or some measurements are presented in one paper and other measurements in another. Thus the situation is exactly analogous to any situation where we wish to combine information contained in multiple measures over multiple studies.

5.1 Implication of Ordering

The results for H1 and H2 indicate that in the first case there is substantial uncertainty in the overall mean BMD. This uncertainty is largely due to the fact that the ‘population’ of patient groups is being estimated by just two groups in the sample. When a mean is estimated by a sample of two, it is well appreciated that there is substantial residual uncertainty in the point estimate.

In the second case, the estimate of mean BMD suffers only from the fact that there is between measure variability; the fact that there are three measures provides a better estimate of the ‘population’ of measures.

Again, it is emphasised that these are different models; however, when estimating overall mean BMD, the summaries from either model may plausibly be presented.

This fact is mirrored in the simulation study, where the point estimates are the same (up to monte carlo error and rounding), but the uncertainty in the estimates differs.

5.2 Recommendations

The aim of this article is to highlight the fact that the question of model specification involves an element of choice. In the particular case of hierarchical models, the order of the hierarchies is one such choice.

As outlined at the start of the paper, the model used arises for a mixture of philosophical, physical or biological reasons. It may be chosen because of the way in which the study has been designed. It may be that there is a natural ordering for the hierarchy in a hierarchical model. However, the fact that the ordering of the hierarchy makes a difference to the results highlights the fact that a choice has to be made.

If the analyst is examining a collection of models, with a view to allowing the data decide on a model, this is a question of data driven model choice. In this setting the ‘best’ model is chosen because it has a better fit to the data. This is the case, for example with stepwise regression models.

The general question of how a model is chosen is something that has engaged minds across the discipline for many years. Lindley, [19] paraphrasing and exaggerating de Finetti, suggests ‘*think* when constructing the model; with it, do not think but leave it to the computer.’ If thinking leaves one confused, then perhaps some combination of models is what is required.

In the absence of any compelling reason to choose one ordering over another, then consideration may be given to model averaging. Such methods may naturally be applied in a Bayesian context, as outlined in [20].

In any case, it is essential, therefore, that authors are clear about how the model was specified, why a particular specification was chosen, and what is being reported in the interval estimates.

If there is any doubt about the order of the specification, then consideration should be given to a clear strategy for model choice, to averaging across models, or even to simply reporting the results from multiple models that could reasonably be fitted.

Acknowledgements

The authors are grateful to the research and clinical staff at St James's Hospital Dublin for providing summary data on bone density.

Appendix

The likelihood for H1 may be written;

$$\begin{aligned} f(\mathbf{y}|\theta) &= \int_V \prod_{i=1}^2 [f(\mathbf{y}|\theta_i) f(\theta_i|\theta) d\theta_i] \\ &= \int_V \prod_{i=1}^2 \left[\prod_{j=1}^3 \{f(\mathbf{y}|\theta_{ij}) f(\theta_{ij}|\theta_i) f(\theta_i|\theta) d\theta_i d\theta_{ij}\} \right] \end{aligned}$$

The likelihood for H2 may be written;

$$\begin{aligned} f(\mathbf{y}|\theta) &= \int_V \prod_{i=1}^2 [f(\mathbf{y}|\theta_j) f(\theta_j|\theta) d\theta_j] \\ &= \int_V \prod_{i=1}^2 \left[\prod_{j=1}^3 \{f(\mathbf{y}|\theta_{ij}) f(\theta_{ij}|\theta_j) f(\theta_j|\theta) d\theta_j d\theta_{ij}\} \right] \end{aligned}$$

Thus, it is noted that the likelihoods differ since they are convolutions of gaussians of different dimension.

In the simple case of the fixed effects model, it is noted that;

$$\begin{aligned} f(\theta_{ij}|\theta_i) &= \delta(\theta_{ij}, \theta_i) \\ f(\theta_i|\theta) &= \delta(\theta_i, \theta) \end{aligned}$$

and so the convolution simplifies the likelihood for a model of the form of H1 and H2 to the same thing, namely;

$$\prod_{i=1}^2 \left[\prod_{j=1}^3 \{f(y_{ij}|\theta)\} \right]$$

a straightforward product of Gaussians.

References

- [1] Haile, A., Webster, J.T. (1975) A comparison of four designs for estimating the variances of random effects in a two-way classification. *Technometrics* 17, 495-500.
- [2] Witte, J.S., Greenland, S. (1996) Simulation study of hierarchical regression. *Statist. Med.* 15, 1161-1170.
- [3] Arends, L.R. (2006) *Multivariate Meta-Analysis: Modelling the Heterogeneity*. PhD Thesis, Erasmus University Rotterdam.
- [4] Greenland, S. (2000) Principles of multilevel modelling. *Int. J. Epidemiol.* 29, 158-167.
- [5] Riley, R.D., Abrams, K.R., Lambert, P.C., Sutton, A.J., Thompson, J.R. An evaluation of bivariate random-effects meta-analysis for the joint synthesis of two correlated outcomes. *Statistics in Medicine* 2007; **26**(1):78–97.
- [6] Montgomery, D. *Design and analysis of experiments* (6th edn). Hoboken, NJ : John Wiley & Sons (2005)
- [7] Nam I, Mengersen K, Garthwaite P. Multivariate meta-analysis. *Statistics in Medicine* 2003; **22**(14):2309–2333.
- [8] Riley, R.D., Evidence synthesis of prognostic marker studies. University of Leicester: PhD Thesis. (2004)
- [9] Spiegelhalter, DJ, Thomas A, Best NG, Lunn D. *WINBUGS version 1.4.1: User Manual*. Cambridge: Medical Research Council Biostatistics Unit. (2004)
- [10] R Development Core Team *R: A Language and Environment for Statistical Computing* R Foundation for Statistical Computing, Vienna, Austria, 2005. <http://www.R-project.org/> [16 September 2006].
- [11] Sturtz S, Ligges U, Gelman A R2WinBUGS: A package for running WinBUGS from R *Journal of Statistical Software* 2005. **12**(3):1–16. <http://www.jstatsoft.org/v12/i03/v12i03.pdf> [16 October 2006].

- [12] Plummer M, Best NG, Cowles K, Vines K. *CODA: Output Analysis and Diagnostics for MCMC. R package version 0.9-1* (2004) URL <http://www-fis.iarc.fr/coda/>. [12 February 2007].
- [13] David J. Spiegelhalter, Nicola G. Best, Bradley P. Carlin and Angelika van der Linde. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2002; **64**(4):583–639.
- [14] Helfenstein, U. (2002) Data *and* models determine treatment proposals - an illustration from meta-analysis. *Postgraduate Medical J.* 78, 131-134.
- [15] Van Houwelingen, H. Zwinderman, K., Stijnen, T. (1993) A bivariate approach to meta-analysis. *Statistics in Medicine* 12, 2272-2284.
- [16] vanHouwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. *Statistics in Medicine* 2002; **21**(4):589–624.
- [17] DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clinical Trials* 1986; **7**(4):177–188.
- [18] DIC Page *The BUGS Project - DIC* The WinBUGS team. <http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/dicpage.shtml> [12 February 2007].
- [19] Dennis V. Lindley The Philosophy of Statistics *Journal of the Royal Statistical Society: Series D (The Statistician)* (2000); **49** (3):293-337
- [20] Jennifer A. Hoeting, David Madigan, Adrian E. Raftery and Chris T. Volinsky Bayesian model averaging: a tutorial *Statistical Science* (1999); **14** (4):382-417