

**Use of Bayesian multivariate meta-analysis to estimate HAQ for mapping onto EQ-5D in rheumatoid arthritis**

## Abstract

**Background:** In health technology assessment (HTA), decisions about reimbursement for new health technologies are largely based on effectiveness estimates. However sometimes the target effectiveness estimates are not readily available. This may be because many alternative instruments measuring these outcomes are being used (and not all always reported) or an extended follow-up time of clinical trials is needed to evaluate long-term endpoints leading to the limited data on the target clinical outcome. In the areas of highest priority in healthcare, decisions are required to be made on a short time scale. Therefore alternative clinical outcomes, including surrogate endpoints, are increasingly being considered for use in evidence synthesis as part of economic evaluation. **Methods:** Bayesian multivariate meta-analysis (MVMA) has been used to synthesize data on correlated outcomes in rheumatoid arthritis (RA) and to incorporate external data in the model in the form of informative prior distributions. Estimates of Health Assessment Questionnaire (HAQ) were then mapped onto the health-related quality of life measure EQ-5D and the effect was compared to mapping the HAQ obtained from the univariate approach. **Results:** The use of multivariate meta-analysis can lead to reduced uncertainty around the effectiveness parameter and ultimately uncertainty around the utility. **Conclusions:** By allowing all the relevant data to be incorporated in estimating clinical effectiveness outcomes, multivariate meta-analysis improves the estimation of health utilities estimated through mapping methods. While reduced uncertainty may impact on decisions based on economic evaluation of new health technologies, the use of short-term surrogate endpoints can allow for early decisions .

## Introduction

In health technology assessment (HTA), decisions are made about reimbursement of new health technologies. Such decisions are largely based on effectiveness outcomes, alongside cost and quality of life estimates. Effectiveness measures are usually estimated by meta-analysing outcomes from randomized controlled trials (RCTs). However, the outcomes of interest (target outcomes) are sometimes

not readily available. It may prove too costly to obtain the target clinical endpoint, or take an extended period of time to measure these outcomes, both of which conflict with the need for efficient and timely assessment of new technologies especially in areas of healthcare that are considered a priority in society such as for example cancer. Another reason for the target outcome not always being readily available is the wide diversity of alternative instruments measuring disease activity or response to treatment, as is the case for example in rheumatoid arthritis (RA), leading to variation in the choice of an endpoint for a primary outcome in RCTs. Taking into account only the target outcome measure can lead to the amount of data included in the meta-analysis being radically reduced. In circumstances of absence of data on the primary clinical outcome, surrogate endpoints are sometimes used in lieu of that outcome (7). Alternatively, in some circumstances data on target endpoint may exist but, in addition to the primary outcome, is also available either on alternative scales or measured by surrogate outcome. Some manufacturers may choose a biomarker as a primary outcome, for example when such endpoint can be measured early leading to a shorter time to licensing and hence quicker profit from the new pharmaceutical product. When drug is conditionally licensed based on a biomarker, data needed for HTA, such as overall survival in cancer treatments, is not available from the clinical trial evaluating this product. Multivariate approach to meta-analysis can allow to include the data on biomarker by combining it with data on a target outcome from other RCTs. Synthesizing the data from all sources of evidence can lead to a reduced bias (8) and potentially also the uncertainty (9).

Multivariate meta-analytic methods have already been developed to synthesize simultaneously evidence from correlated outcomes (10; 11). In a recent review, Jackson et al describe advances in the development of the methodology of multivariate meta-analysis and discuss the advantages and disadvantages of the use of these methods (12). **Bujkiewicz et al.** have recently developed a Bayesian multivariate meta-analysis (MVMA) framework for combining data from multiple correlated outcomes (9). One of the advantages of the multivariate approach that was exploited in this model was the incorporation of additional data by the inclusion of studies reporting multiple outcomes. In addition to this, the Bayesian aspect of this approach made it possible to incorporate external data in the MVMA model in the form of informative prior distributions, especially on correlations. This method, developed using an example in RA

for combining data from studies reporting the Health Assessment Questionnaire (HAQ) and the Disease Activity Score (DAS-28) measures, and the American College of Rheumatology (ACR) response criteria, allowed for extending the range of data that could be included in the analysis, thus reducing possible biases (8) and also led to a reduced uncertainty around the outcome of interest (HAQ in this case). The aim of this work was to apply such multivariate meta-analytic framework which allows evidence to be combined from studies that are a mixture of those reporting the target clinical outcome, an alternative clinical endpoint or both (rather than reducing the number of studies to those reporting the clinical outcome of interest or including the surrogate endpoint alone in the economic model) to jointly estimate effect of treatment on HAQ. This is a flexible approach which permits the use of multiple outcomes and hence prevents valuable data from clinical trials from being discarded and potentially can lead to reduced uncertainty around the clinical outcome of interest.

Clinical effectiveness outcomes are often used to derive health-related quality of life or utility estimates, as part of the economic evaluation of the new health technologies. Increasingly, mapping methods are being developed to predict health state utility values from preference-based measures, such as EuroQoL (EQ-5D) values, using disease-specific outcomes. For example, mapping methods have been applied to predict EQ-5D values from (i) the St. George's Respiratory Questionnaire in patients with chronic obstructive pulmonary disease (1), (ii) functional assessment of cancer therapy-melanoma (2), (iii) disease-specific quality-of-life scores in patients with migraine (3), (iv) the Western Ontario and McMaster Universities Osteoarthritis Index in patients with knee osteoarthritis (4) and, finally, (v) using HAQ in rheumatoid arthritis (RA) (5; 6). In this article we explore how reducing uncertainty around the target clinical outcome, in this case HAQ, by the use of multivariate meta-analysis can ultimately lead to better estimation of health utility such as EQ-5D obtained by a mapping technique.

## **Methods**

### *Sources of evidence*

As our motivating example, we used a systematic review and meta-analysis which was carried out by Lloyd et al (14) to investigate the effectiveness of the TNF- $\alpha$  inhibitors, etanercept, infliximab and adalimumab used sequentially (as a second line treatment used sequentially in patients who switched from one anti-TNF to another due to lack of response or adverse effects on the first) in the treatment of RA. The data from this systematic review is single arm (i.e. uncontrolled) due to the lack of comparative studies such as randomised controlled trials evaluating these treatment as second line. Outcome measures included in the analysis in this paper were amongst the standard instruments for measuring response to treatment in RA: changes from baseline in the HAQ ( $\Delta$ HAQ) and in the DAS-28 ( $\Delta$ DAS-28) measures, and 20% response to treatment measured by the ACR criteria (ACR20). Data collected in this systematic review is used to investigate how multivariate meta-analysis can be applied to incorporate multiple outcomes of effectiveness simultaneously in evidence synthesis. Table 2 gives details of the three outcomes which were reported in each of the studies within the systematic review. We will refer to these data as the "Lloyd data" throughout this paper.

Additional sources of evidence were used to construct prior distributions for some of the parameters of the Bayesian evidence synthesis model. External individual patient data (IPD) was obtained from the British Rheumatoid Outcome Study Group (BROSG) trial which was designed to assess the benefit of aggressive disease-modifying anti-rheumatic drug treatments in patients with established RA (15). The trial, which recruited 466 patients with stable RA, assessed clinical outcomes (i.e. HAQ, DAS-28 and ACR20) in two cohorts of patients managed using either a regime focussed on (i) symptomatic control of pain and stiffness in the shared care setting, or (ii) a more aggressive regime focussing on control of symptoms and joint inflammation in the hospital setting. (The BROSG trial found no difference between the aggressive versus symptomatic treatment arms, therefore this data set was used as a single cohort of 466 patients with established RA.) Another source of evidence, external summary data (ESD), was obtained from a systematic review of the anti TNF- $\alpha$  inhibitors in biologically naïve patients (16). Both external sources of evidence (IPD and ESD) were used to construct prior distributions for correlations between outcomes as described below.

*Evidence synthesis: statistical methods*

In our motivating example, we aimed to model the summary data of the correlated outcomes from the “Lloyd data” using a MVMA in a Bayesian form. We used a trivariate random effects meta-analysis (TRMA), described in much more detail elsewhere (9), in the following form

$$\begin{pmatrix} \Delta HAQ_i \\ \Delta DAS_i \\ ACR20_i \end{pmatrix} \sim N \left( \begin{pmatrix} \mu_{Hi} \\ \mu_{Di} \\ \mu_{Ai} \end{pmatrix}, \begin{pmatrix} \sigma_{Hi}^2 & \sigma_{Hi}\sigma_{Di}\rho_{wi}^{HD} & \sigma_{Hi}\sigma_{Ai}\rho_{wi}^{HA} \\ \sigma_{Hi}\sigma_{Di}\rho_{wi}^{HD} & \sigma_{Di}^2 & \sigma_{Di}\sigma_{Ai}\rho_{wi}^{DA} \\ \sigma_{Hi}\sigma_{Ai}\rho_{wi}^{HA} & \sigma_{Di}\sigma_{Ai}\rho_{wi}^{DA} & \sigma_{Ai}^2 \end{pmatrix} \right) \quad [1]$$

$$\begin{pmatrix} \mu_{Hi} \\ \mu_{Di} \\ \mu_{Ai} \end{pmatrix} \sim N \left( \begin{pmatrix} \beta_H \\ \beta_D \\ \beta_A \end{pmatrix}, \begin{pmatrix} \tau_H^2 & \tau_H\tau_D\rho_b^{HD} & \tau_H\tau_A\rho_b^{HA} \\ \tau_H\tau_D\rho_b^{HD} & \tau_D^2 & \tau_D\tau_A\rho_b^{DA} \\ \tau_H\tau_A\rho_b^{HA} & \tau_D\tau_A\rho_b^{DA} & \tau_A^2 \end{pmatrix} \right) \quad [2]$$

which has a hierarchical structure: the correlated outcomes ( $\Delta HAQ$ ,  $\Delta DAS$  and  $ACR20$ ) in each study  $i$  follow a multivariate normal distribution and they are assumed to be the estimates of underlying correlated effects ( $\mu_{Hi}, \mu_{Di}, \mu_{Ai}$ ) for the individual studies (the within-study models in equation [1]). These study-level effects are assumed to follow a common distribution (the between-study model in equation [2]). This meta-analytic model can be used for a bivariate and univariate analyses by reducing the size of the matrices in the equations. The within-study variances ( $\sigma_{Hi}^2, \sigma_{Di}^2, \sigma_{Ai}^2$ ) are assumed to be known (for studies reporting the relevant outcomes) but prior distributions need to be specified for the within-study correlations ( $\rho_{wi}^{HD}, \rho_{wi}^{HA}, \rho_{wi}^{DA}$ ) as none of the studies in the “Lloyd data” reported them. Prior distributions for the within-study correlations were constructed using the external IPD described above. The elements of the covariance matrix of the between-study model (the between-study variances ( $\tau_H^2, \tau_D^2, \tau_A^2$ ) and the between-study correlations ( $\rho_b^{HD}, \rho_b^{HA}, \rho_b^{DA}$ )) also need prior distributions to be placed on them. In our model, prior distributions on the between-study correlations were constructed by meta-analysing ESD using the trivariate model [1]-[2] but this time with non-informative prior distributions. Resulting posterior between-study correlations were then used as prior distributions to the between-study correlations in meta-analysis of the “Lloyd data”. In order to inform the between-study correlation by the ESD directly, we parameterised the between-study model in the form of the product of univariate conditional distributions (in contrast to

using a multivariate Normal distribution and a prior distribution on the precision matrix, such as Wishart distribution (17) or using a spherical parameterisation of the between-study correlation matrix (18) (19)). Non-informative (half-normal) prior distributions are placed on the standard deviations of these conditional distributions.  $\Delta$ DAS-28 and ACR20 are both conditional on  $\Delta$ HAQ in this model. This parameterisation assumes a conditional independence between  $\Delta$ DAS-28 and ACR20. This assumption may not always be valid, especially in the circumstances when the two variables are strongly correlated. We have carried out a sensitivity analysis on the external data to investigate the impact of the choice of such parameterisation on estimates and in particular the resulting prior distributions for the between-study correlation. Details of this sensitivity analysis are included in the Web Appendix. The advantages of this approach along with all the details of the statistical modelling (including models in different number of outcomes) are reported elsewhere (9).

#### *Mapping the HAQ onto EQ-5D*

A number of methodological approaches to modelling utility estimates in RA for the purpose of economic evaluation have been developed. The estimate of the HAQ, a self-reported measure of physical function, is often used to predict EQ-5D, an estimate of health-related quality of life by mapping the HAQ score onto EQ-5D (5; 6). These methods often focus on the relationship between EQ-5D and HAQ scores, usually expressed in a linear form (5), and have been compared with EQ-5D values based on patient responses in a number of datasets in RA (6). The purpose of this paper was to illustrate the potential impact of reduced uncertainty around the clinical outcome on the utility when estimating it from a multivariate meta-analysis. To do this, we modelled the relationship between the HAQ and the health-related quality of life measure, EQ-5D, in a linear form estimated by Chen et al (16) from data reported in a study by Hurst (20):  $EQ - 5D = a + b \times HAQ$ , where the intercept  $a \sim N(\mu_a, s_a^2)$  and slope  $b \sim N(\mu_b, s_b^2)$ , with means  $\mu_a=0.628$  and  $\mu_b=-0.327$ , and standard deviations  $s_a=0.034$  and  $s_b=0.021$  (note that this approach assumed independence of intercept and slope). This relationship can be assumed to remain the same at any time point, hence the relationship between the change from baseline of EQ-5D and the change from baseline of HAQ can be modelled as  $\Delta EQ - 5D = b \times \Delta HAQ$ .

The change from baseline in HAQ estimate,  $\Delta HAQ$ , used in the above mapping equation, typically comes from a meta-analysis. As can be seen in Table 2, there are only 8 studies reporting the HAQ in the "Lloyd data", while there were many more studies investigating effectiveness of the same treatment in the same type of patients but reporting DAS-28 and/or ACR20. To be able to include all outcome data from all studies we have used the multivariate approach to meta-analysis as described above.

### *Software implementation*

All models were implemented in WinBUGS (17) where the estimates were obtained using MCMC simulation using 100,000 iterations (including 50,000 burn-in). All posterior estimates are presented as means with the 95% highest probability density intervals (HPDI) (21) obtained using the BOA package (22). This evidence synthesis aimed to give an estimate of  $\Delta HAQ$  from all available sources of evidence of the effectiveness (of studies reporting a variety of outcomes) which can then be used to estimate  $\Delta EQ-5D$ .

## **Results**

We have applied the MVMA model to the "Lloyd data" in order to model simultaneously data on the HAQ, DAS-28 and ACR20. To investigate the impact of including more data in the analysis on uncertainty around the HAQ estimate and consequently the EQ-5D, we explored results of the meta-analyses on three levels: using univariate random effects meta-analysis (URMA) of  $\Delta HAQ$ , bivariate random-effects meta-analysis (BRMA) combining the  $\Delta HAQ$  and the  $\Delta DAS-28$  and finally trivariate random-effects meta-analysis (TRMA) by extending the data by including ACR20. The prior correlations between DAS-28 and HAQ obtained from the external data were:  $\rho_{wi}^{HD} = 0.24$  (95% HPDI: 0.13 to 0.35) obtained from the IPD and  $\rho_b^{HD} = 0.86$  (95% HPDI: 0.46 to 0.999) obtained from bivariate meta-analysis of the ESD. For the trivariate case, the prior within-study correlations, constructed using the IPD, were  $\rho_{wi}^{HD} = 0.24$  (95% HPDI: 0.10 to 0.38),  $\rho_{wi}^{HA} = -0.13$  (95% HPDI: -0.29 to 0.0103) and  $\rho_{wi}^{DA} = -0.20$  (95% HPDI: -0.31 to -0.08), and the prior between-study correlations, constructed using the ESD, were  $\rho_b^{HD} = 0.78$  (95% HPDI: 0.27 to

0.998) and  $\rho_b^{HA} = -0.14$  (95% HPDI: -0.80 to 0.56) (only the two correlations were required in the product normal parameterisation of the between-study model).

Table 3 shows results of applying all of the three models. Mapping estimate of  $\Delta$ HAQ of -0.25 (95% HPDI: -0.43 to -0.09), obtained from URMA of the eight studies reporting only the HAQ outcomes, resulted in  $\Delta$ EQ-5D equal 0.08 (95% HPDI: 0.025 to 0.141). Extending the evidence synthesis by ten cohorts reporting the  $\Delta$ DAS-28 (but not  $\Delta$ HAQ), by the use of BRMA, lead to a shift in the estimate of HAQ and the reduction in the uncertainty around the estimate to -0.28 (95% HPDI: -0.41 to -0.14). This has also reduced the uncertainty around the estimate of  $\Delta$ EQ-5D which in this case was 0.09 (95% HPDI: 0.041 to 0.138). This effect can be seen in Figure 1 where horizontal lines represent the 95% HPDIs. The intervals are reduced for the estimates of  $\Delta$ HAQ and  $\Delta$ EQ-5D when using BRMA compared to URMA. As can be seen in Table 3 and also in Figure 1, extending the evidence synthesis to TRMA by adding further three studies reporting the ACR20 did not contribute to further reduction of the uncertainty around the estimates of interest ( $\Delta$ HAQ and  $\Delta$ EQ-5D). This is likely due to increased heterogeneity by introducing the three studies. More detailed description of results (related to the effectiveness outcomes only) are reported elsewhere (9). Results of the bivariate meta-analyses of the HAQ with ACR20 and the DAS-28 with ACR20 are shown in Table 3 for completeness. Bivariate model of the HAQ and ACR20 did not lead to reduced uncertainty around the HAQ similarly as for the trivariate case discussed above.

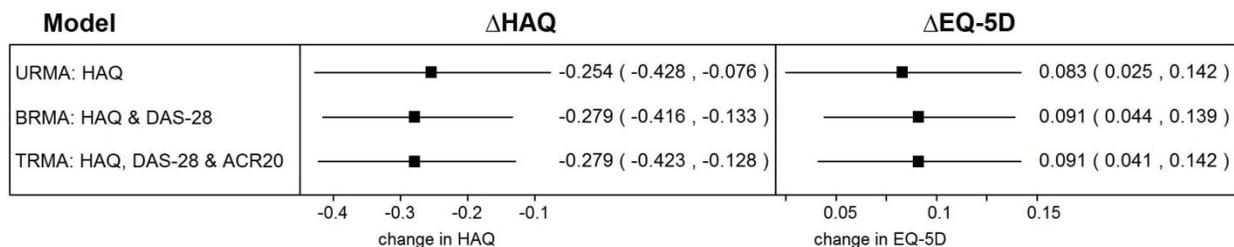


Figure 1. Estimates of change from baseline of HAQ, obtained from meta-analysis of HAQ only, HAQ and DAS-28 and finally HAQ, DAS-28 and ACR20, and resulting estimates of the change from baseline in EQ-5D.

## Discussion

We have applied a Bayesian meta-analytic framework to synthesise diverse sources of evidence and multiple endpoints in RA. Our aim was to increase the use of available data to estimate effectiveness of second line biologics in treatment of RA measured by the change from baseline in HAQ and ultimately the health-related quality of life measured by the change from baseline in EQ-5D. We employed external IPD to construct prior distributions for the within-study correlations and ESD for the between-study correlations. We found that using this Bayesian approach to evidence synthesis of outcomes in RA, by combining the HAQ endpoint with the DAS-28, leads to a 20% reduction in the uncertainty around the HAQ from -0.25 (95% HPDI: -0.43 to -0.09) obtained from URMA of synthesizing data on HAQ alone to -0.28 (95% HPDI: -0.41 to -0.14) obtained from BRMA. The use of BRMA has also led to a reduced uncertainty, by 16%, around the health-related quality of life estimate EQ-5D from 0.083 (95% HPDI: 0.025 to 0.141) when mapping HAQ obtained from URMA to 0.091 (95% HPDI: 0.041 to 0.138) when using the estimate of HAQ from BRMA. Similar results have been obtained when using TRMA of the HAQ, DAS-28 and ACR20. The inclusion of the third outcome, ACR20, did not contribute to the HAQ estimate or further reduction of uncertainty around it. A 20% reduction in uncertainty around the effectiveness outcome may lead to a change of the cost effectiveness outcomes, for example alter the cost-effectiveness acceptability curve, leading to a change in decision made based on such analysis. Joint analysis of multiple outcomes may have advantages in many areas of decision making, for example in risk-benefit analysis, where multiple outcomes contribute to net benefit function and sometimes the risk function as well. For example, in a study on hormone replacement therapy by Minelli et al (23), net benefit was constructed by adding benefits measure by relative risk of hip fracture, menopausal symptoms, colorectal cancer and endometrial cancer, whereas the overall harm was estimated by combining relative risk of breast cancer, coronary heart disease, pulmonary embolism and stroke. Each of these eight outcomes were estimates by a separate univariate meta-analyses. Modelling such outcomes jointly may lead to more precise estimates especially in circumstances where outcomes are measured in randomised controlled trials which are powered with respect to one outcome, but may be underpowered for others. This may also be the case when adverse events are included in the analysis whilst RCTs are not powered

to for such outcomes, therefore there is a lot of uncertainty about them when including them in net-benefit analysis.

When pooling data by means of the Bayesian random-effects meta-analysis, we allow for "borrowing of strength" across studies which leads to "shrinkage" of the estimates from individual studies towards the overall mean (with reduced uncertainty around the estimates) (21). In addition to this, multivariate meta-analysis allows "borrowing of strength" across outcomes. This may lead to reduced uncertainty around the pooled estimates. However, it is difficult to predict the extent of the gain in precision of the estimates. In fact, the inclusion of studies reporting additional outcomes through the multivariate approach can lead to higher heterogeneity in the data set (which is the case in our example in RA when extending the bivariate case by the third outcome) which can in fact lead to increased uncertainty. The issue of the potential gain in precision of the estimates obtained from multivariate models have been discussed previously. Nam et al observed that the gain is more likely to be noticeable when the data sets are small (10). Riley et al showed that the gain in precision is expected unless either the outcome are not correlated or the data is complete on both outcomes which are also reported with equal variance (11). In their recent review, Jackson et al reported that the gain in precision increases with increased within-study correlation, but noted that also it can depend of the between-study variance (12). Whilst the factors driving the uncertainty around the estimates of the multivariate meta-analytic models seem complex and it is difficult to predict the gain in precision or whether its degree will be satisfactory to the analysts, there are further advantages of employing the multivariate scheme. Restricting the analysis to a single outcome may lead to ignoring data from clinical trials which either report treatment effect measured by a different instrument(s) or simply choose to report only positive and/or significant results. In the former case, multivariate approach allows using more diverse sources of evidence and hence prevents relevant data from being discarded. In the latter case, in the presence of outcome reporting bias, the multivariate approach can lead to a more appropriate estimate of the clinical outcome as it has been shown by Kirkham et al that (8). This may apply to the case of RA where, due to the availability of a number of instruments for measuring the disease activity or response to treatment, authors may choose to report only those outcomes whose estimates are positive and/or significant. The pooling of multiple outcomes

through the multivariate approach can obviously be achieved by means of frequentist analysis. The Bayesian approach, however, brings additional advantages of incorporating external evidence in the form of prior distributions. Individual patient data (used in our example to construct the prior distributions for the within-study correlations) is not always available. However, other sources of evidence can be used to inform the parameters of the model. The Bayesian approach, for example, allows elicitation techniques to be employed to take into account experts' opinions in the form of prior distributions in a model. For example, methods for inclusion of expert opinion on bias in the synthesis of studies were developed by Turner et al (24). WinBUGS provides suitable software environment for implementing such Bayesian analysis. Multivariate meta-analysis can be conducted for example in Stata using the `mvmeta` command (25) or in R `mvmeta` package (26) in a frequentist approach. WinBUGS allows for the use of the full Bayesian framework, but can be interfaced with R, which is convenient when decision models are developed in R (27).

In this multivariate approach to evidence synthesis, studies reporting a variety of relevant measures associated with an outcome of interest can be included. Outcomes can include surrogate endpoints, so far mostly used instead of the primary outcome of interest when there is no available evidence on the target outcome (7). In the method described here the surrogate outcomes are used alongside the primary outcome, and therefore the analysis is not restricted to a single primary outcome or a surrogate endpoint. The main interest of this research is in uncertainty (and magnitude of the effect) in terms of the EQ-5D as it is this outcome that is used for the decision making. In our example, none of the studies in the “Lloyd data” reported EQ-5D and hence the linear mapping was chosen to estimate EQ-5D from HAQ. However, when some of studies included in evidence synthesis report EQ-5D, the multivariate meta-analysis can include the EQ-5D as one of the multiple outcomes and the pooled effect is then estimated simultaneously with the clinical outcomes (borrowing strength from data across studies and outcomes). Clinical outcomes can then be considered surrogate endpoints to utility. This approach may be preferable to the linear mapping as often the regression models are developed based on small amount of data. Cooper et al (25) have shown that the data quality (together with quantity, and therefore uncertainty) can be least for utility

estimates (e.g. EQ-5D) especially compared with effectiveness and therefore, in the absence of better direct utility data, the methods considered here may at least reduce uncertainty in the decision problem.

In our view, the approach described here is an important step towards the inclusion of the surrogate endpoints in HTA, especially in the early stages of the drug development process when the target outcome is not readily available. Modelling simultaneously data on surrogate (potentially short-term outcome) and the ultimate endpoint can increase the available evidence allowing analysts and decision makers an early assessment of new technologies. Surrogate endpoints play an increasingly important role in drug development as reported by the European Medicines Agency (26) and HTA and their use will need to increase in HTA as reimbursement decisions are taken closer to regulatory ones (27) and randomized controlled trials become shorter in duration (28).

## References

1. *Predicting EQ-5D values using the SGRQ.* **Starkie HJ, Briggs AH, Chambers MG, Jones P.** 2011, Value Health, Vol. 14, pp. 354-360.
2. *Mapping FACT-melanoma quality-of-life scores to EQ-5D health utility weights.* **Askew RL, Swartz RJ, Xing Y, et al.** 2011, Value Health, Vol. 14, pp. 900-906.
3. *Mapping from disease-specific measures to health-state utility values in individuals with migraine.* **Gillard PJ, Devine B, Varon SF, et al.** 2012, Value Health, pp. 485-494.
4. *Use of a disease-specific instrument in economic evaluations: mapping WOMAC onto the EQ-5D utility index.* **Xie F, Pullenayegum EM, Li SC, et al.** 2010, Value Health, pp. 873-878.
5. *Cost effectiveness of adalimumab in the treatment of patients with moderate to severe rheumatoid arthritis.* **Bansback, NJ, Brennan, A, Ghatnekar, O.** 2005, Annals of Rheum Diseases, Vol. 64, pp. 995-1002.
6. *Exploring the validity of estimating EQ-5D and SF-6D utility values from the health assessment questionnaire in patients with inflammatory arthritis.* **Harrison, M. J., Lunt, M., Verstappen, S. M. M., Watson, K. D., Bansback, N. J. and Symmons, D. P. M.** 21, 2010, Health and Quality of Life Outcomes, Vol. 8.

7. *The use of surrogate outcomes in model-based cost-effectiveness analyses: a survey of UK Health Technology Assessment reports.* **Taylor, R. S. and Elston, J.** 8, 2009, Health Technology Assessment, Vol. 13, pp. iii,ix-xi, 1-50.
8. *A multivariate meta-analysis approach for reducing the impact of outcome reporting bias in systematic reviews.* **Kirkham, J. J., Riley, R. D., and Williamson, P. R.** 2012, Statistics in Medicine, p. published online 25 April 2012.
9. *Multivariate meta-analysis of mixed outcomes: a Bayesian approach.* **Bujkiewicz S, Thompson JR, Sutton AJ, Cooper NJ, Harrison MJ, Symmons DPM, Abrams KR.** 2013, accepted for publication in Statistics in Medicine.
10. *Multivariate meta-analysis.* **Nam I-S, Mengersen K, Garthwaite P.** 2003, Statistics in Medicine, Vol. 22, pp. 2309–2333.
11. *An evaluation of bivariate random-effects meta-analysis for the joint synthesis of two correlated outcomes.* **Riley, R. D., Abrams, K. R., Lambert, P. C., Sutton, A. J., and Thompson, J. R.** 2007, Statistics in Medicine, Vol. 26, pp. 78-97.
12. *Multivariate meta-analysis: Potential and promise.* **Jackson, D., Riley, R. and White I. R.** 2012, Statistics in Medicine, Vol. 30, pp. 2481-2498.
13. *If it works why can't we have it? Sequential anti-TNF therapy in the UK.* **Deighton, C.** 2010, Rheumatology, Vol. 49, pp. 2235-2236.
14. *The effectiveness of anti-TNF-alpha therapies when used sequentially in rheumatoid arthritis patients: a systematic review and meta-analysis.* **Lloyd, S., Bujkiewicz, S., Wailoo, A. J., Sutton, A. J. and Scott, D.** 2010, Rheumatology, Vol. 49, pp. 2313-2321.
15. *Patients with stable long-standing rheumatoid arthritis continue to deteriorate despite intensified treatment with traditional disease modifying anti-rheumatic drugs - results of the British Rheumatoid Outcome Study Group randomized controlled clinical tr.* **Symmons, D. P. M., Tricker, K., Harrison, M. J., Roberts, C., Davis, M., Dawes, P., Hassell, A., Knight, S., Mulherin, D. and Scott, D.** 2006, Rheumatology, Vol. 45, pp. 558-565.
16. *A systematic review of the effectiveness of adalimumab, etanercept and iniximab for the treatment of rheumatoid arthritis in adults and an ec.* **Chen, Y.-F., Jobanputra, P., Barton, B., Jowett, S., Bryan, S., Clark, W., Fry-Smith A., and Burls, A.** 2006.
17. *WinBUGS User Manual.* **Spiegelhalter, D., Thomas, A., Best, N. and Lunn, D.** 2003.
18. *Modelling between-study variance structure in mixed treatment comparisons.* **Lu G, Ades A.** 2009, Biostatistics, pp. 792-805.
19. *Bayesian multivariate meta-analysis with multiple outcomes.* **Wei Y, Higgins PT.** 6 February 2013, Statistics in Medicine, Vols. e-pub ahead of print.

20. *Measuring health-related quality of life in rheumatoid arthritis: validity, responsiveness and reliability of EuroQoL (EQ-5D)*. **Hurst NP, Kind P., Ruta D, Hunter M, Stubbings AI**. 1997, *British Journal of Rheumatology*, Vol. 36, pp. 551-559.
21. **Spiegelhalter DJ, Abrams KR, Myles JP**. *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. s.l. : John Wiley and Sons, Ltd, 2004.
22. *boa: An R Package for MCMC Output Convergence Assessment and Posterior Inference*. **Smith, BJ**. 11, 2007, *Journal of Statistical Software*, Vol. 21.
23. *Benefits and harms associated with hormone replacement therapy: clinical decision analysis*. **Minelli C, Abrams KR, Sutton AJ, Cooper NJ**. 14 February 2004, *BMJ*, Vol. 328.
24. *Bias modelling in evidence synthesis*. **Turner, R. M., Spiegelhalter, D. J., Smith, G. C. S. and Thompson, S. G**. 2009, *Journal of the Royal Statistical Society A*, Vol. 172, pp. 21--47.
25. *Use of evidence in decision models: an appraisal of health technology assessments in the UK since 1997*. **Cooper, N., Coyle, D., Abrams, K., Mugford, M., and Sutton, A**. 2005, *J Health Serv Res Policy*, Vol. 10, pp. 245--250.
26. *Innovative drug development approaches (EMEA/127318/2007)*. **Agency, European Medicines**. 2007, Final report from the EMEA/CHMP-THINK-TANK group on innovative drug development (EMEA, London, UK).
27. *Comparison of anticancer drug coverage decisions in the United States and United Kingdom: does the evidence support rhetoric?* . **Mason, A., Drummond, M. , Ramsey, S., Campbell, J. and Raisch, D**. 2010, *J Clin Onc*, Vol. 28, pp. 3234--3238.
28. *The FDA's critical path initiative: a brief introduction*. **Coons, S. J**. 2009, *Clinical Therapeutics*, Vol. 31, pp. 2572--2573.
29. *Modified Disease-Activity Scores That Include 28-Joint Counts - Development and Validation in A Prospective Longitudinal-Study of Patients with Rheumatoid-Arthritis*. **Prevo MLL, Vanthof MA, Kuper HH, Vanleeuwen MA, Vandeputte LBA, Vanriel PLCM**. 1, 1995, *Arthritis Rheum Jan*, Vol. 38, pp. 44-8.
30. *American College of Rheumatology preliminary definition of improvement in rheumatoid arthritis*. **Felson DT, Anderson JJ, Boers M, Bombardier C, Furst D, Goldsmith C, et al**. 6, 1995, *Arthritis Rheum* , Vol. 38, pp. 727-35.
31. *Stanford Health Assessment Questionnaire modified to assess disability in British patients with rheumatoid arthritis*. **Kirwan JR, Reeback JS**. 2, 1986, *Br J Rheumatol*, Vol. 25, pp. 206-9.

Table 1 Relevant instruments measuring disease activity and/or response to treatment in patients with RA.

Measure	Description	Range (best to worst)	Form used in this paper
Disease Activity Score (DAS-28)	Composite of disease activity based on 28 swollen & tender joint count, ESR (or CRP), and patient global assessment (15).	0-9.3	Change from baseline (ΔDAS-28)
American College of Rheumatology criteria (ACR)	Indicated by % improvement in tender or swollen joint counts and three out of the following five measures: acute phase reactant (e.g. ESR/CRP), patient global assessment, physician global assessment, pain scale (e.g. pain VAS), disability/function (e.g. HAQ) (16).	1, 0 (Binary outcome)	20% improvement (ACR20)
Health Assessment Questionnaire (HAQ)	Functional disability assessed using on 8 domains (dressing and grooming, rising, eating, walking, hygiene, reach, grip, activities) (17).	0 -3	Change from baseline (ΔHAQ)

Table 2 Studies and outcomes included in the "Lloyd data".

Study	ΔHAQ	ΔDAS28	ACR20
	Mean (se)	Mean (se)	r/n
Bennett (2005)	-0.31 (0.13)	-1.7 (0.25)	–

Bingham (2009)	-0.35 (0.05)	-1.6 (0.1)	85/188
Bombardieri (2007)	-0.48 (0.02)	-1.9 (0.05)	486/810
Haraoui (2004)	-0.45 (0.14)	–	14/22
Hyrich (2008)	-0.12 (0.03)	–	–
Iannone (2009)	0.15 (0.13)	–	–
Navarro-Sarabia (2009)	-0.21 (0.07)	-1.1 (0.18)	–
Van der Bijl (2005)	-0.21 (0.08)	-1.5 (0.25)	19/41
Buch (2007)	–	-1.47 (0.18)	55/72
Cohen (2005)	–	-1.87 (0.24)	–
Di Poi (2007)	–	-2.1 (0.29)	–
Finckh (2007)	–	-0.98 (0.18)	–
Hjardem (2007)	–	-1 (0.11)	–
Laas (2008) InTol	–	-1.17 (0.66)	–
Laas (2008) InEff	–	-1.26 (0.35)	–
Nikas (2006)	–	-2.4 (0.16)	18/24
Wick (2005) EA	–	-1.9 (0.22)	7/9
Wick (2005)IA	–	-1.3 (0.28)	19/27
Buch (2005)	–	–	18/25

Karlsson (2008)	–	–	172/337
Van Vollenhoven (2003)	–	–	12/18

**Table 3 Results of evidence synthesis from the three models: URMA, BRMA and TRMA and resulting health-related quality of life estimates.**

Model	Number of cohorts/studies	$\Delta$ HAQ	$\Delta$ DAS-28	ACR20	$\Delta$ EQ-5D
HAQ only	8/8	-0.25 (-0.43, -0.09)	–	–	0.083 (0.025, 0.142)
DAS-28 only	15/13	–	-1.57 (-1.84, -1.31)	–	–
ACR20 only	11/10	–	–	62% (53%, 71%)	–
HAQ & DAS-28	18/16	-0.28 (-0.41, -0.14)	-1.51 (-1.67, -1.35)	–	0.091 (0.044, 0.139)
HAQ & ACR20	15/14	-0.31 (-0.5, -0.13)	–	60% (51%, 69%)	0.101 (0.039, 0.163)
DAS & ACR20	19	–	-1.56 (-1.82, 1.30)	62% (52%, 71%)	–
HAQ , DAS-28 & ACR20	21/19	-0.28 (-0.42, -0.13)	-1.51 (-1.70, -1.33)	61% (52%, 71%)	0.091 (0.041, 0.142)