Use of evidence in cost-effectiveness decision modelling: taxonomy for current practice and methods

Paper ID number: P27

Saramago, Pedro[a,]*, Manca, Andrea[a], Sutton, Alex J.[b]

[a] Centre for Health Economics, University of York, York, U.K.

[b] Department of Health Sciences, University of Leicester, Leicester, U.K.

* Corresponding author.

Centre for Health Economics, The University of York, Alcuin A, York YO10 5DD, UK, Tel: +44 (0)1904 321453;Fax: +44 (0)1904 321402; E-mail: prsg500@york.ac.uk

Summary The evidence base informing economic evaluation models does hardly ever derive from a single source. To reflect the needs of a particular decision problem it is the researcher choice which method best estimates the required parameters, leading to different individuals taking different approaches. This paper proposes a taxonomy and provides guidance on how distinct sources of evidence could be used to inform effectiveness input parameters for economic models. This taxonomy considers the sources of evidence, type or format(s) of this information and whether the estimation is concerned with a single or multiple parameters of effectiveness. A sample of applied and methodological papers was reviewed and an examination of the methods employed to synthesize evidence from distinct sources was performed. This work allowed the (i) categorisation of evidence sources to inform model structure and effectiveness parameter estimates; (ii) assessment of methods used to synthesize or transform evidence from different study designs; and (iii) drawing guidance on these analysis methods.

1. Introduction

Economic evaluations assess the economic and health consequences of healthcare interventions, programmes or services with the aim of informing policy decisions regarding resource provision within healthcare systems operating under a fixed budget.

When conducting an economic evaluation, evidence on total costs incurred and effects is required and may be derived from randomised clinical trials (RCTs), observational

studies, expert opinion and/or secondary analyses (such as meta-analysis) of any of the above [3; 19]. The evidence base required for economic evaluation comes hardly ever from a single data source. Decision-analytic models (DAMs) are therefore needed to synthesise the data to allow the cost-effectiveness (CE) of alternative clinical interventions to be assessed. DAMs offer a means of structuring decision problems, synthesizing all available data and characterizing the uncertainty in the decision [29; 30]. As with any model framework, results obtained from these decision models depend on the suitability of the model structure and the quality of the data inputs [5].

The National Institute for Health and Clinical Excellence (NICE) in the UK is an example of an institution that recognises the value of DAMs to inform the assessment of whether technologies are value for money. The guidelines for methods of technology appraisal published by NICE [24] recommend that 'all relevant evidence must be identified', 'quality assessed and, where appropriate, pooled using explicit criteria and use justifiable and reproducible methods' and that 'explicit criteria by which studies are included and excluded' should be presented.

One issue frequently faced by health economics modellers is how to progress when multiple sources of evidence are available to inform certain parameters, particularly on relative effectiveness. Measures derived from the statistical synthesis of aggregated data (e.g. meta-analysis or mixed treatment comparison) have been increasingly used in CE analysis. Although evidence may exhibit heterogeneity, be more prone to bias or may be indirect in various ways, if it is clearly relevant it should not be excluded from the analysis. Nonetheless, the analysis undertaken need to appropriately account for these issues and results should reflect the uncertainty present in the input evidence [2].

As important as the appropriate use of evidence to inform parameter estimates, is the identification of relevant evidence (for the different model parameters) and the assessment of the quality of different sources of evidence. These and other issues have raised a number of important methodological questions on the use of evidence to inform decision making. The recent statistical developments in the methods for synthesis and analysis of evidence that extended the use of complex methods alongside DAMs may considerably change 'current practice'.

Two very interesting published papers that approach both practical and methodological issues surrounding evidence synthesis in economic decision models are the Cooper et al (2007) [5] and the Ades et al (2006) [2]. While the later focus on the methods for

evidence synthesis within a Bayesian framework, the former outlines the 'state of play' regarding the use of evidence in decision models pointing out challenges related to identifying, combining and reporting evidence to inform DAMs parameters and structure.

This paper proposes a taxonomy that aims to (comprehensively) identify and describe the use (and methods of analysis) of evidence to derive input effectiveness parameters for CE decision modelling. We identify possible developments in methods and needs for future research.

2. A taxonomy for the use of evidence in cost-effectiveness models

The purpose of the present paper is to focus on the use of evidence on parameters related to relative effectiveness, used to inform a DAM evaluating the value for money of health technologies. In this context, evidence is commonly searched retrospectively. In this sense, the way to proceed with the analysis may depend on the sources of evidence, type or format(s) of this information and whether the estimation is concerned with a single or multiple parameters of effectiveness.

A taxonomy of possible scenarios faced by the analyst based on a combination of the above mentioned factors is presented in Tables 1a and 1b, where each grid cell is related to its neighbours by altering only one factor.

Single source of evidence

		Parameter	
		Single	Multiple
Data available	AD	A1	B1
	IPD	C1	D1

Multiple sources of evidence

		Parameter	
		Single	Multiple
Data available	AD	A2	B2
	IPD	C2	D2
	Mixture (IPD & AD)	E2	F2

Table 1a (left) and 1b (right) –A gallery of scenarios arising when using clinical evidence in cost-effectiveness models

The first dimension to consider relates to the number of *sources of evidence*. If only one source of evidence is available to contribute to the required analysis, then Table 1a is to

be considered. On the other hand, if more than one source of evidence is used to inform the estimation of the parameter(s) of interest, then the sub-categorizations in use by Table 1b should be considered.

The format in which data is available is the second dimension to consider here. That is whether the evidence base is available in aggregate form only (aggregate data, AD), is available at individual patient-level (IPD) only or whether some studies contribute with IPD and others only give AD. Caution must be taken over the definitions of AD and IPD. Here, IPD evidence is defined as data collected within one study (one source of evidence) available to the analyst as raw data for each study participant. In this sense, IPD is the most informative format of data collected within a study. By reducing IPD, information is lost and evidence is now in AD form. Commonly, AD is also denominated summary data, and examples may be the proportion of patients who experienced a clinical event among those at risk, or median time to wound healing. In particular cases, aggregate and individual-patient level data may convey the same level of information, e.g. if only two binary random variables were collected, say occurrence of death and group allocation, a two-by-two table contains all information, and may be used to estimate a measure of effect size like the odds ratio (OR). In this case, both the two-by-two table and a per patient database will be denominated IPD. On the contrary, if other variables, such as age, were collected within the same study, information is being lost by presenting only the two-by-two table on gender and group allocation. In this case, it is considered the two-by-two table as AD.

The third dimension to take into account is the *number of parameters* the synthesis model estimates. The proposed taxonomy makes a distinction between informing a single effectiveness parameter and informing multiple effectiveness parameters, since methods of analysis may differ. By informing multiple effectiveness parameters we imply the circumstances when, for instance, more than two treatments are being assessed, or when subgroup analysis is conducted or when one require effectiveness parameter estimates at different time points.

2.1 Single source of evidence

The simplest situations are when only a single source of evidence is available from which to derive the parameter(s) of interest. These types of situations are here described and analysed.

2.1.1 - Aggregate data to inform a single parameter (A1)

In the search for information, AD is extremely abundant in comparison to IPD. An example of this situation is when there is only a single source of evidence and this is available in aggregate format to inform the input effectiveness parameter in the decision model. This situation may occur in decision models, either because no additional studies evaluating the same treatment were conducted or because other existing studies are excluded due to the presence of unmanageable heterogeneity (e.g. distinct populations being evaluated).

The main issue of using a single AD source is that information may be lost in relation to what was collected. As an example, we may need to further adjust for confounders. It would be impossible to derive an adjusted treatment estimate if this was not done as part of the process of generating the AD estimate. As a consequence, this may compromise the robustness of the effectiveness estimates inputted in the decision model.

2.1.2 - Aggregate data to inform multiple parameters (B1)

There are often several parameters of interests one may need to extract when a single source of (aggregate) data is available. While situation A1 focused on the scenario where one may have more than one parameter but only one parameter is required, here one wants to inform multiple parameters that are needed for the decision model.

In a cardiovascular study for instance, one may be interested in the median time to myocardial infarction and median time to stroke. Other example is when multiple comparators are of interest and results from one three armed trial comparing placebo (P), drug 1 (D_1) and drug 2 (D_2) are used to extract odds ratios: OR_{PvsD1} and OR_{PvsD2}. Given the study framework, these aggregate measures are inherently correlated, however, one may ignore this or be forced to ignore the correlation between these parameters from the same study unless this is published and/or accounted for in the analysis. Another possible situation is when the CE analysis aims at studying subgroups. In such case, results of a study reporting summary statistics on treatment effect and an interaction effect with subgroup may be used. Also, CE models may consider effectiveness of treatments in more than one dimension, and consider that the treatment may affect, for example, both mortality and time to event. Potential pitfalls of the analysis described in the previous section (A1) also apply to the current situation.

2.1.3 Individual patient data to inform a single parameter (C1)

In those instances where there is only a single source of evidence for the parameter of interest it is possible that the analyst has access to IPD. In this case data can be analysed to estimate the parameter of interest (e.g. median time to myocardial infarction). It is widely acknowledged that access to IPD allows the re-analysis of the data, inclusion of further explanatory covariates and conduction of more in-depth analyses than is possible from summary data extracted from published reports. IPD may be considered the most flexible way to explore and answer clinical and economical research questions.

In recent years the development of statistical methods in the economic evaluation area has strengthened the advantages of analysing IPD in CE analysis compared to the use of AD. However one should have in mind that when conducting economic evaluation alongside clinical trial there are various methodological problems that could limit the relevance of the study result for policy making. Firstly, it may lack external validity which raises problems when interpreting results; secondly, the shortness or inadequacy of the follow-up periods used; thirdly, the use of surrogate endpoints and the frequent presence of incomplete and inappropriate comparators; and fourthly, the analysis of the study can be highly time consuming and costly to be carried out [5].

We will not expand on methods of analysis of IPD data as these relate to the standard statistical literature.

2.1.4 Individual patient data to inform multiple parameters (D1)

One of the major advantages from having available evidence in IPD format is the potential wealth of information that can be drawn from it. In fact, not only it is possible to use the trial IPD to derive multiple effectiveness parameters but also to estimate the correlation between these parameters, facilitating a better reflection of the parameter uncertainty in decision model. For instance, the economic analysis of the third Randomized Intervention Trial of unstable Angina (RITA 3) [10] used several regression models to estimate rates of cardiovascular death and myocardial infarction (also costs and health-related quality of life). Within each effectiveness parameter estimated the correlation structure was maintained by using the Cholesky decomposition of the variance-covariance matrix method. Additional information on standard mortality rates were incorporated in the decision model.

2.2 Multiple sources of evidence

More complicated are those instances where there are multiple sources of evidence from which to derive the parameter(s) of interest. These are discussed in turn.

2.2.1 Aggregate data to inform a single parameter (A2)

Let us start with the scenario in which several studies report results on the same parameter of interest and one wants to combine these into a single quantitative estimate. The statistical methods commonly used to achieve this synthesis fall within the meta-analysis family.

In meta-analysis, each of *k* studies estimates the parameter of interest θ_i (*i=1, …, k*). In meta-analysis of clinical trials, for instance, the parameter of interest is usually some measure of relative efficacy between the treatment arms. The most popular choice is the log-odds ratio, but this could also be the risk or rate difference or the risk or rate ratio for dichotomous outcome or similar measures for continuous outcomes or survival data. All studies report an estimate, $\hat{\theta}_i$, of the true θ and a measure of sampling uncertainty, most commonly the standard error, s_i.

An analysis under the assumption of between-study homogeneity of the treatment effect makes the assumption that the unknown parameter is exactly the same in all studies, that is, $\theta_1 = \theta_2 = ... = \theta_n = \theta$. More often though one can observe between-study heterogeneity in the estimation of the treatment effect. As well as investigating the presence of heterogeneity it is also necessary to consider its underlying cause. Heterogeneity may be due to chance, or may be spurious due to the scale used to measure the treatment effect. Study level features, such as characteristics of the treatment, may also lead to heterogeneity and these can be investigated by adopting a meta-regression approach. When patient characteristics are contributing to the heterogeneity, these can only potentially be accounted for using meta-regression, but is should be pointed out that there is always the risk of ecological fallacy. If we only have one source of evidence then it is impossible to quantify and explore heterogeneity.

If none of the above seems to account for the existing heterogeneity, then the latter may be unexplainable. In this case, a judgement has to be made on whether it is appropriate

to statistically combine the results and if so with what model and what conclusions can be drawn from it. After defining a model for the data (accounting or not for known sources of heterogeneity) we can choose to disregard, or not, possible unexplained sources of heterogeneity. It is possible to account for unexplained heterogeneity by assuming the various estimates of the parameter of interest from distinct sources differ but are 'related' to each other. The latter is achieved by assuming the study's parameter estimate is a specific realisation from a common distribution, perhaps Gaussian, with unknown mean and variance (these must then be estimated). This approach is known as random-effect meta-analysis [6; 33].

When patient's baseline covariates mean estimates at the study-level are available and we suspect these may be important in explaining heterogeneity, meta-regression techniques can and have been used in a wide diversity of situations. Despite the fact that such analyses have been shown to have low statistical power when using aggregate patient-level covariates [20].

In CE analysis, the use of estimates of relative treatment effects derived from meta-analysis is common [22]. An example of its use within an economic decision model is on the prevention and treatment of influenza A and B (Turner et al (2003) [39]). Following a systematic review, a meta-analysis of randomised evidence was undertaken to investigate the effectiveness of Oseltamivir and Zanamivir for the treatment and prophylaxis of influenza A and B. For different risk groups, separate meta-analysis were conducted to evaluate time to symptoms alleviated and time to return to normal activities. Informed by the meta-analytic outputs, economic decision models were then constructed to examine the CE of the alternative strategies for treating and preventing influenza A and/or B.

2.2.2 Aggregate data to inform multiple parameters (B2)

Meta-analysis can be used to achieve more complex forms of synthesis, including multiple comparison problems and combinations of evidence on multiple or surrogate end points. Much of the published work on these complex methods of synthesis has been undertaken within a Bayesian framework, mainly for computational reasons but also because of its coherent link to decision-making. To designate these extended methods of synthesis the term 'multi-parameter evidence synthesis' adapted from Hasselblad et al (1995) [9] has been used and they are the focus of situation B2.

When multiple outcomes are of interest, a multivariate meta-analysis model allows estimating jointly these endpoints (considering possible correlation). Often the advantage of a multivariate random-effects meta-analysis lies in its ability to use the between-study correlation of the multiple endpoints of interest. For example, Reitsma et al (2005) [25] have suggested applying a bivariate random-effects meta-analysis (BRMA) to jointly synthesise logit-sensitivity and logit-specificity values from diagnostic studies. Even if we are only interested in one of the individual outcome estimates, it is advantageous to use BRMA as 'strength' is borrowed across outcomes and thus estimates might be more precise.

A common feature of available evidence is the absence of head-to-head trials comparing all relevant options being considered in a CE analysis. When more than two treatments must be compared and the evidence base contains different randomized pair-wise or multi-arm comparisons, the most common technique used in the decision making context is mixed treatment comparisons (MTC) [21]. MTCs can be recognized as an example of multi-parameter evidence synthesis, in which parameters are related to one another by a definable structure [1].

In MTC one may be interested in performing fixed-effects or random-effects analysis, based on the assumptions on between trial heterogeneity as discussed in A2 [1; 11]. Of course, in the multi-parameter case the assumption adopted holds across the entire set of trials, irrespective of which treatments were actually evaluated.

Chain of evidence structures are also very common in the medical literature. Here, the essential feature is that evidence may form several links in a chain of evidence, that is, for instance, the effect of treatment on cardiovascular disease and the relation between disease and mortality. Put together, these two types of information could be used to predict five-year survival. Also, a third type of evidence could tell us about five-year survival directly and could, in principle, confirm or reject hypotheses about how the first two types of information fit together [1].

A good example of applying the multi-parameter evidence approach is a recent work by Welton et al (2008) [40] that was originally developed using data from the earlier appraisal of antiviral treatment by Turner et al (2003) [39] referred to in A2. Welton et al (2008) [40] present a Bayesian MTC of antivirals for treatment of influenza, where some trials reported summary measures on at least one of two events: alleviation of fever and alleviation of symptoms. The variety of summary measures available - mean

or median time to alleviation of symptoms and proportion free of event at the end of follow-up - was also responsible for further complexity in the synthesis. The authors have evaluated the contribution of the different evidence sources to the residual deviance. A Weibull model with exchangeable treatment effects (independent for each outcome) returned the best fit, summarising treatment effect on two outcomes in a single measure. The results from this Bayesian synthesis were then combined with other relevant clinical effectiveness parameters (i.e. the relative treatment effect for complications) [4] as well as other data related to costs and quality of life, in order to assess the cost-effectiveness of oseltamivir and zanamivir for the treatment of influenza.

The multi-parameter evidence synthesis approach may therefore provide a number of potential advantages for decision making purposes. This approach addresses some of the problems caused by variation in reporting summary statistics and in combining evidence on multiple parameters related to the summary outcome of interest, ensuring the consistency of the data sources. It ensures also that the appropriate summary outputs (for example mean durations) are available for the economic model with the potential to increase their precision. However, one should also consider the additional assumptions required in this type of analysis. If assumptions do not hold, then the subsequent results are subject to potential bias. Given that the assumptions employed may not be statistically verifiable, results need to be considered in the light of both clinical and epidemiological judgement.

2.2.3 Individual patient data to inform a single parameter (C2)

Meta-analysis of IPD or 'mega-analysis', where raw data from each study is obtained and synthesized to inform a single measure, is analogous to the AD approach referred to in scenario A2, and is considered the evidence synthesis 'gold-standard' because of its numerous advantages [12; 32]. As mentioned by Stewart et al (1993) [35], IPD avoids the biases of published AD, allows one to obtain information not available from published reports and allows consistent inclusion/exclusion criteria to be used across studies [17]. IPD also allows a longer follow-up time to be assessed and patient-level covariates to be directly modelled. An increase in statistical power to detect true patient treatment relationships is gained when compared to meta-regression of AD [20], which only assesses treatment in relation to group-level summaries [5]. The Cochrane Working Group on meta-analysis using IPD has recently published updated guidelines [13; 36].

Surprisingly there isn't much published literature concerning methodologies for meta-analysing IPD. A review of IPD meta-analyses of trials has been recently published by Simmonds et al (2005) [31] suggesting that most methods used are straightforward. The review shows that the majority of applications use a 'two-stage process', where each IPD dataset is analysed separately, summary data derived for each study and then combined using a standard fixed effect meta-analysis model for aggregate-level data. This type of approach may be considered a simplification towards the techniques explored over scenario A2. Alternatively, one could use random effects multilevel models synthesizing IPD as developed to deal with binary [38], ordinal [41], continuous [12] and longitudinal outcomes [7; 18]. However, this approach appears to be hardly ever used in practice.

As mentioned, one motivation to use IPD is that covariates measured at the level of the individual patient may be incorporated. Several options are available in practice, requiring choices of whether to allow for random coefficients and between no, fixed or random effects for an interaction with treatment. An important consideration is the danger of inappropriate complexity and over-fitting of the data.

2.2.4 Individual patient data to inform multiple parameters (D2)

One of the scenarios where multiple IPD sources can really make the difference is when one the analyst needs to estimate multiple parameters for the decision model.

Recent research suggests that a BRMA of sensitivity and specificity is an appropriate method to use [27]. Despite the potential benefits of IPD for meta-analysis of diagnostic studies, there has been only little consideration of how to synthesize IPD from diagnostic studies, and how to assess accuracy-covariate effects. The recently published work by Riley et al (2008) [27] describes how the meta-analysis of IPD from diagnostic studies can lead to clinical results more tailored to the individual patient. The authors developed models for IPD that extend the BRMA framework to include study-level covariates, which help explain the between-study heterogeneity, and also patient-level covariates, which allow one to assess the effect of patient characteristics on test accuracy.

As mentioned in scenario B2, when AD is available, MTC analysis adds value by combining randomized evidence from networks of trials. Given that we may have indirect or a mix of direct and indirect evidence, the MTC framework can be extended

to include evidence in IPD format. This type of complex multi-parameter synthesis of evidence has not yet been put in to practice, but it would certainly share the benefits and the possible pitfalls of MTC analysis. Importantly, it would allow patient-level covariate adjustment to be extensively facilitated.

2.2.5 Mixture of individual and aggregate-level data to inform a single parameter (E2)

The issue of statistically synthesising IPD and AD to inform clinical effectiveness parameters for economic evaluation modelling is of substantial importance. Limiting the analysis to those studies where only IPD or only AD is available may result in a non-representative sample of studies and thus bias the decision model results [26; 27], contradicting NICE's recommendation that 'all relevant evidence must be identified' [24]. The underlying objective is to better inform model input parameter estimates, accurately reflect their uncertainty illustrating its consequences on the cost-effectiveness outputs of interest. As yet, this issue has received little attention in the economics literature.

Results of IPD-only meta-analysis may be biased if unavailability of IPD is related to the study results [37]. Therefore it is important to increment the available IPD with AD for those studies where IPD are not available. An IPD and AD meta-analysis can be conducted through a two-stage or a one-stage approach [26]. In the former, available IPD are first reduced to AD in each study, and then these AD (from the IPD studies) are combined with the existing AD (from the AD studies) using 'standard' meta-analytic methods of AD techniques (see situation A2 for brief description of 'standard' meta-analysis methods). The two stage approach can be conducted analogously to scenario C2. One should emphasize that this method best suits cases where only the overall pooled (treatment) effect is of interest and not its relation to study-level covariates. One stage methods are mainly frequentist multilevel or hierarchical modelling [8; 38] and Bayesian hierarchical related regression [16; 34].

Recently, a systematic review of the methods used to combine IPD and AD evidence related to clinical effectiveness has been published by Riley et al (2007) [26]. This review identified 8 methodological papers and 33 applied papers. The majority of the applied studies (80%) used the 'two-stage' method; the remaining used partially

reconstruction of IPD methods (e.g. Hukkelhoven et al (2003) [14]) or other unclear methods.

The partially reconstruction of IPD (from AD studies) has been suggested in the literature [38; 41]. This method entails recreating IPD from aggregate information, in binary or ordinal outcome format, and has received little validation so far. These simulated data may be then combined with the already available IPD using methods of meta-analysis of IPD, as described in scenario C2.

While Turner et al (2000) [38] multilevel framework focused on binary outcomes, Goldstein et al. (2000) [8] suggested a multilevel model for combining IPD and AD where continuous data responses are of interest. This allows AD studies to be incorporated alongside IPD studies by including a dummy variable in the model that distinguishes AD responses from IPD responses and allows one to estimate the effect of interest in relation to both study-level covariates and patient-level covariates. Multilevel modelling of both IPD and AD is thus potentially advantageous, but as yet it has received little empirical assessment or validation. Consideration of the approach in more clinical and economical settings is required.

In the context of ecological studies, Jackson et al. (2005) [15] considers the use of individual-level data to improve inferences at the aggregate-level, and uses a Bayesian approach termed hierarchical related regression (HRR). In this model two related regression models are simultaneously estimated -- the IPD-only model and the AD-only model -- where common parameters assess the exposure-outcome relationship. Both IPD and AD are informing the common parameters, avoiding the issues associated with analyzing IPD (low power) or AD (ecological bias) separately. Two potential limitations of applying Bayesian HRR [16] are the need for a correct specification of the IPD-only model and the definition of appropriate prior distributions. Further assessment and validation of these models are necessary, especially in clinical areas.

A closely related contribution was published by Sutton et al (2008) [34]. Here, the 'one stage' synthesis of data was performed with studies from distinct designs – individually and cluster-allocated – and where distinct data formats where available – IPD and AD. The model differs from the Bayesian HRR given that it aims at improving efficiency by using all available data at the level that it is available at.

2.2.6 Mixture of individual and aggregate-level data to inform multiple parameters (F2)

An important aspect referred to in E2 is how to combine IPD and AD studies, which is important as all relevant IPD information are not always available. If patient-level covariates are to be assessed, modelling needs to accomplish this by using the IPD studies to provide the necessary patient-level covariate information. When only study-level parameters are modelled, the AD and IPD studies can be combined reasonably easily by converting, for instance, the AD to IPD. Complex models using a mixture of multiple IPD and AD information with the purpose of informing multiple parameters are possibly the most intricate framework that the present taxonomy will refer to and is the focus of scenario F2.

Multi-parameter analysis like these remains under-researched when one considers that these evidence structures may be actually far more common than others explored in the present taxonomy. Although methods are apparently not fully developed yet, one may think of extending multi-parameter analysis structures, like the ones explored by Ades et al (2003) [1] or Lu et al (2007) [21]. As an example, this later paper reports a series of Bayesian hierarchical models based on piece-wise exponential hazards that borrowed strength across the MTC networks and also across time points. Extending this framework could allow incorporating IPD alongside the existing AD evidence in the MTC structure at more than one follow-up time, and where information from different sources may relate to different points in time.

A recently published work by Riley et al (2008) [27] describes the use of BRMA to model outcomes of diagnostic studies (referred to in B2). The authors extended the analysis to include a mixture of IPD and AD. The model built, combining IPD and AD, allows all studies to estimate the sensitivity and specificity across studies, the impact of study-level covariates, and also the across-study effects. However, only IPD studies were used to estimate the within-study effects. Thus, both IPD studies and AD studies are contributing valuable information. The authors acknowledge that sometimes there is little or no IPD or perhaps only small variation of covariate values within IPD studies, such that the across-study effects may then be the major source of information available. In such situations, before interpreting the across-study effects consideration must be given to whether ecological bias may exist and whether all potential study-level

confounders have been adjusted for. With this in mind, one may wish to assess the sensitivity of meta-analysis results to the inclusion of AD studies, and explore any differences between IPD and AD studies.

Also, Molitor et al (2009) [23] published recently a modelling framework that accounts for multiple biases simultaneously and gives more accurate parameter estimates than standard approaches. By concluding that multiple sources of data are often necessary to identify the biases and to inform about different aspects of the research question they use a Bayesian graphical models that provides a coherent way to connect a series of local sub-models, based on different data sets, into a global unified analysis.

3 Discussion

This taxonomic grid identifies and describes scenarios related to the use of effectiveness evidence in cost-effectiveness analysis. In particular, it addresses the question of how to proceed depending on the availability of sources of evidence, type or format(s) of this information and when single or multiple parameters are being informed by the available evidence. The solutions presented are mainly based on statistical models; however their role is key to best inform decision makers. Some of the synthesis options herein discussed are already well described and established in the literature, for example, meta-analysis and meta-regression. Other options need further research before being used to address decision making within the health care framework.

This work could be extended to include extra dimensions – e.g. extrapolation of model estimates. Also, the evidence synthesis methodologies and attached applied studies described throughout the paper are not the result of a comprehensive systematic review, but just a sample of the literature in this area of research. It is our intention to use a comprehensive review of the literature to obtain an all-inclusive taxonomy.

Guidelines on methods for technology assessment published by NICE [24] emphasize that 'all relevant evidence must be identified' and incorporated in decision analysis. Obeying to this pre-requisite, primary evidence from trials to expert opinion must be identified and adequately filtered to inform the structure and input parameters of the decision model. However, the guidelines lack specific practical assistance and provide no clear guidance principle over the definition of 'relevant' evidence. Cooper et al (2007) discusses thoroughly this issue of evidence relevancy interacting it with the concept of 'sufficient' evidence. They mention that usually evidence identification and

selection strategies are both neither uniform nor transparent. Actuality, the underlying message is that researchers use the evidence that they perceive to be the most relevant or convenient for each model input parameter rather than the recommended 'all relevant evidence available'. In fact, commercial interests lead to pharmaceutical companies, for instance, having IPD from their trials but only AD from their competitors.

Most decision models have been 'populated' by having independent data sources for each parameter. But this is a marvellous coincidence: whatever the model, the number of data sources available to inform it usually equals the number of parameters. The truth is that 'additional' sources of evidence, such as register evidence, tend to be ignored because analysts have been unfamiliar with recently developed methods for multi-parameter evidence synthesis.

Sculpher et al (2000) [28] recommend that model structure should be as simple as possible, consistent with the stated decision problem and not constrained by data availability. Model structure influences the approach to the analysis, in particular the parameters required by the decision model. Pragmatically, of course, the data available and any possible pre-existing evidence synthesis can also inform the parameterisation of the decision model, e.g. the choice of whether to use two separate parameters for the effectiveness of treatment, or the relative or absolute difference between those two values. Thus, complex statistical modelling used to derive the parameters may generally have impact on decision model specification, as it is sometimes explicitly laid out.

Finally, it is clear that there is a great need for a wider consciousness of the range of techniques and their applicability. Several of these techniques are in the sphere of epidemiology, statistics and operational research, which, in some cases, their possible complexity may be responsible for some of the apparent averseness to analysis methods. However, complexity of methods usually means that the results available are either more robust or more informative, and in either case helping the decision process. The use of the full range of techniques set out in the presented taxonomy and the criteria presented here for an appropriate use of evidence in decision modelling may allow an appropriate identification and assessment of the quality of model inputs, which may reduce dramatically the scepticism sometimes surrounding model outputs.

References

[1] Ades, A. E. (2003). A chain of evidence with mixed comparisons: models for multi-parameter evidence synthesis and consistency of evidence. Statistics in Medicine, 22: 2995-3016;

[2] Ades, A. E., Sculpher, M. J., Sutton, A. J., Abrams, K., Cooper, N., Welton, N. and Lu, G. (2006) Bayesian Methods for Evidence Synthesis in Cost-Effectiveness Analysis. Pharmacoeconomics, 24 (1): 1-19;

[3] Briggs, A., Claxton, K. and Sculpher, M. J. (2006). Decision analytic modelling for the evaluation of health technologies. Handbooks in Health Economic Evaluation Series. Oxford University Press, First edition;

[4] Burch, J., Paulden, M., Conti, S., Stock, C., Corbett, M., Welton, N. J., Ades, A. E., Sutton, A., Cooper, N., Elliot, A., Nicholson, K., Duffy, S., McKenna, C., Stewart, L., Westwood, M. and Palmer, S. [in Press] Antiviral drugs for the treatment of influenza: A Systematic Review and Economic Evaluation. Health Technology Assessment;

[5] Cooper, N. J., Sutton, A. J., Ades, A. E., Paisley, S. and Jones, D. R. (2007). Use of evidence in economic decision models: practical issues and methodological challenges. Health Economics (16), 1277–1286;

[6] DerSimonian, R. and Laird, N. (1986). Meta-analysis in clinical trials. Controlled Clinical Trials (7), 177-188;

[7] Farlow, M. R., Small, G. W., Quarg, P. and Krause, A. (2005). Efficacy of Rivastigmine in Alzheimer's disease patients with rapid disease progression: results of a meta-analysis. Dementia, Geriatrics Cognitive Disorders, 20: 192–197;

[8] Goldstein, H., Yang, M., Omar, R. Z., Turner, R. M. and Thompson, S. G. (2000). Meta-analysis using multilevel models with an application to the study of class size effects. Annals of Applied Statistics (49), 399-412;

[9] Hasselblad, V. and McCrory, D. C. (1995). Meta-analytic tools for medical decision making: a practical guide. Medical Decision Making, 15: 81-96;

[10] Henriksson, M., Epstein, D., Palmer, S., Sculpher, M., Clayton, T., Pocock, S., Henderson, R., Buxton, M. and Fox, K. (2008). The cost-effectiveness of an early interventional strategy in non-ST-elevation acute coronary syndrome based on the RITA 3 trial. British Medical Journal - Heart, (94), 717-723;

[11] Higgins, J. P. T. and Whitehead, J. (1996). Borrowing strength from external trials in meta-analysis. Statistics in Medicine, 15: 2733-2749;

[12] Higgins, J. P. T., Whitehead, A. and Turner, R. M. (2001). Meta-analysis of continuous outcome data from individual patients. Statistics in Medicine (20), 2219-2241;

[13] Higgins, J. P. T. and Green, S. Cochrane Handbook for Systematic Reviews of Interventions. Version 5.0.1 (September, 2008). The Cochrane Collaboration, 2008. Available online from: www.cochrane-handbook.org;

[14] Hukkelhoven, C. W., Steyerberg, E. W., Rampen, A. J., Farace, E., Habbema, J. D. and Marshall, L. F., (2003). Patient age and outcome following severe traumatic brain injury: an analysis of 5600 patients. Journal of Neurosurgery, 99: 666-673;

[15] Jackson, C., Best, N. and Richardson, S. (2005). Improving ecological inference using individual-level data. Statistics in Medicine, 25: 2136–2159;

[16] Jackson, C., Best, N. and Richardson, S. (2008). Hierarchical related regression for combining aggregate and survey data in studies of socio-economic disease risk factors. Journal of the Royal Statistics Society (171, part 1), 159-178;

[17] Jeng, G. T., Scott, J. R. and Burmeister, L. F. (1995). A comparison of meta-analytic results using literature vs individual patient data. Journal of the American Medical Association (274), 830–836;

[18] Jones, A. P., Riley, R. D., Williamson, P. R. and Whitehead, A. (2009). Meta-analysis of individual patient data vs aggregate data from longitudinal clinical trials. Clinical Trials, 6: 16-27;

[19] Kuntz, K. M. and Weinstein, M. C. (2001). Modelling in economic evaluation. In Economic evaluation in health care: merging theory with practice. Drummond, M., Mcguire, A. Oxford University Press. Oxford, Second edition;

[20] Lambert, P. C., Sutton, A. J., Abrams, K. R. (2002). A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis. Journal of Clinical Epidemiology (55), 86-94;

[21] Lu, G., Ades, A. E., Sutton, A. J., Cooper, N. J., Briggs, A. H. and Caldwell, D. M. (2007). Meta-analysis of mixed treatment comparisons at multiple follow-up times. Statistics in medicine, 26(20): 3681-99;

[22] Mandelblatt, J. S., Fryback, D. G. and Weinstein, M. C. (1996). Assessing the effectiveness of health interventions. In: Gold MR, Siegel JE, Russell LB, et al., editors. Cost-effectiveness in health and medicine. New York: Oxford University Press, 1996: 135-175;

[23] Molitor, N. T., Best, N., Jackson, C. and Richardson, S. (2009). Using Bayesian graphical models to model biases in observational studies and to combine multiple sources of data: application to low birth weight and water disinfection by-products. Journal of the Royal Statistics Society A, 172(3): 615–637;

[24] National Institute for Health and Clinical Excellence (2008). Briefing paper for methods review workshop on exploring uncertainty. Institute's Decision Support Unit, London: National Institute for Health and Clinical Excellence (online at: http://www.nice.org.uk/aboutnice/howwework/devnicetech/technologyappraisalprocess guides/guidetothemethodsoftechnologyappraisal.jsp);

[25] Reitsma, J. B., Glas, A. S., Rutjes, A. W., Scholten, R. J., Bossuyt, P. M. and Zwinderman, A. H. (2005). Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. Journal of Clinical Epidemiology, 58(10): 982-990;

[26] Riley, R. D., Simmonds, M. C. and Look, M. P. (2007). Evidence synthesis combining individual patient data and aggregate data: a systematic review identified current practice and possible methods. Journal of Clinical Epidemiology (60), 431–439;

[27] Riley, R. D., Lambert, P. C., Staessen, J. A., Wang, J., Gueyffier, F., Thijs, L. and Boutitie, F. (2008) Meta-analysis of diagnostic test studies using individual patient data and aggregate data. Statistics in Medicine (27), 1870-1893;

[28] Sculpher, M., Fenwick, E. and Claxton, K. (2000). Assessing quality in decision analytic cost-effectiveness models. Pharmacoeconomics,17: 461–477;

[29] Sculpher, M., Claxton, K. (2005). Establishing the cost-effectiveness of new pharmaceuticals under conditions of uncertainty – when is there sufficient evidence? Value in Health (8), 433–446;

[30] Sculpher, M. J., Claxton, K. and Akehurst, R. (2005). It's Just evaluation for decision making: recent developments in, and challenges for, cost-effectiveness research. In Smith, P., Ginnelly, L., Sculpher, M., Health Policy and Economics – Opportunities and Challenges. Open University Press – State of Health Series;

[31] Simmonds, M. C., Higgins, J. P. T., Stewart, L. A., Tierney, J. F., Clarke, M. J. and Thompson, S. G. (2005). Meta-analysis of individual patient data from randomized trials: a review of methods used in practice. Clinical Trials (2), 209–217;

[32] Sutton, A. J., Abrams, K. R., Jones, D. R., Sheldon, T. A. and Song, F. (2000). Methods for meta-analysis in medical research. Wiley series in probability and statistics. London, Wiley, First edition;

[33] Sutton, A. J. and Abrams, K. R. (2001). Bayesian methods in meta-analysis and evidence synthesis. Statistics and Methods for Medical Research, 10: 277-303;

[34] Sutton, A. J., Kendrick, D. and Coupland, C. A. C. (2008). Meta-analysis of individual- and aggregate-level data. Statistics in Medicine, 27: 651–669;

[35] Stewart, L. A. and Parmar, M. K. (1993). Meta-analysis of the literature or of individual patient data: is there a difference? Lancet (22), 341-418;

[36] Stewart, L. A. and Clarke, M. J. (1995). Practical methodology of meta-analysis (overviews) using updated individual patient data. Cochrane Working Group. Statistics in Medicine (14), 2057-2079;

[37] Stewart, L. A. and Tierney, J. F. (2002). To IPD or not to IPD? Advantages and disadvantages of systematic reviews using individual patient data. Evaluation and the Health Professionals (25), 76–97;

[38] Turner, R. M., Omar, R. Z., Yang, M., Goldstein, H. and Thompson, S. G. (2000). A multilevel model framework for meta-analysis of clinical trials with binary outcomes. Statistics in Medicine, 19:3417–3432;

[39] Turner, D., Wailoo, A., Nicholson, K., Cooper, N., Sutton, A. and Abrams, K. (2003) Systematic review and economic decision modelling for the prevention and treatment of influenza A and B. Health Technology Assessment, 7:1-182;

[40] Welton, N. J., Cooper, N. J., Ades, A. E. and Sutton, A. J. (2008). Mixed treatment comparison with multiple outcomes reported inconsistently across trials: evaluation of antivirals for treatment of influenza A and B. Statistics in Medicine, 27(27): 5620-5639;

[41] Whitehead, A., Omar, R. Z., Higgins, J. P., Savaluny, E., Turner, R. M. and Thompson, S. G. (2001). Meta-analysis of ordinal outcomes using individual patient data. Statistics in Medicine (20), 2243-2260.