Paper Number P35

How valuable are multiple treatment comparison methods in economic decision modelling?

Nicola Cooper, Alex Sutton, Monica Lai, Keith Abrams, Sylwia Bujkiewicz.
Department of Health Sciences, University of Leicester.

Jaime Peters
Peninsula Technology Assessment Group (PenTAG), Exeter.

David Spiegelhalter
MRC Biostatistics Unit, Cambridge

Peter Juni, Simon Wandel
University of Berne, Switzerland

Steve Palmer, Mike Paulden
Centre for Health Economics, University of York

Stefano Conti
Health Protection Agency, London

Nicky Welton
Academic Unit of Primary Health Care, University of Bristol

Contact details

Nicola Cooper: Department of Health Sciences, Adrian Building, University of Leicester, University Road, Leicester LE1 7RH, UK. E-mail: njc21@le.ac.uk

ABSTRACT

AIMS: To explore how multiple treatment comparison (MTC) synthesis models compared to the use of standard pair-wise meta-analysis methods impact on economic decision models with respect to the i) assumptions the models make, ii) empirical estimates produced, and iii) conclusions drawn.

METHODS: In three topic areas, we compare the use of pair-wise meta-analysis methods to more complex MTC synthesis approaches for estimating the relative effectiveness of alternative treatment regimes. The first analysis considers the clinical effectiveness of interventions for preventing strokes in people with atrial fibrillation. Secondly, we consider the use of drug-eluting stents in percutaneous coronary intervention in patients with coronary artery disease. Here we re-evaluate an economic decision model using an MTC incorporating data at multiple time points compared to a standard pair-wise approach. Finally, we focus on the use of neuraminidase inhibitors in the treatment of influenza adopting a MTC that allows for censored time to event data which informs an economic decision model.

RESULTS & CONCLUSIONS: In all three examples, more evidence was used in the estimation of effectiveness parameters using MTC methods. This was due to the inclusion of further RCTs and consideration of information at multiple time-points. The use of further RCTs generally increases the precision of the MTC estimates, while the less restrictive distributional assumptions used in the influenza analysis resulted in less precise effectiveness estimates. The difference between point estimates of effectiveness produced by the two approaches was generally unpredictable – sometimes agreeing closely while in other instances differing considerably in either direction.

INTRODUCTION:

A core component of health technology assessments (HTAs) is to estimate the effectiveness and cost-effectiveness of alternative healthcare interventions. Ideally, effectiveness data are obtained from well conducted randomised controlled trials (RCTs); the most reliable source. Where multiple relevant RCTs exist, appropriate evidence synthesis methods should be used. Comparisons are often made between two specific interventions and summarized through the use of pairwise meta-analysis methods(Sutton et al., 2000). These pooled results are often used to inform the associated economic analyses. However, there may be interest in comparing more than two competing healthcare interventions to answer policy relevant questions. For example, compared to current usual practice, two new competing interventions may need evaluating. Unfortunately, there is usually a lack of trial data available that compares *all* the interventions of interest directly (e.g. it is unlikely that all trials compare all three interventions of interest in the above situation).

A logical extension to more established meta-analysis methods, that allows the comparison of more than two interventions, is multiple (mixed) treatment comparison (MTC) approaches(Lumley, 2002) (Lu and Ades, 2004, Caldwell et al., 2005). These approaches allow the simultaneous estimation of the comparative effectiveness of multiple treatments using an evidence base of trials which individually do not compare all treatment options. A detailed discussion of the use of MTC methods in HTA is available elsewhere(Sutton et al., 2008). Briefly, the main assumptions of an MTC analysis are that:

i) The trials to be synthesised form a connected network. Examples of diagrammatic representations of trial networks are given in figures 1, 3 and 6. In each of these, there are no treatments which are isolated and not compared to at least one of the other treatments in the network (i.e. all the treatment nodes are connected by lines indicating that a randomised comparison exists leaving no isolated nodes without a connection).

ii) There is a consistency across the evidence base. Consider a three treatment network with treatments labelled A, B and C. The method assumes that, if 2-arm trials comparing B v C exist, then if such trials had a third, A, arm, then they would produce an estimate of A v C and A v B that was consistent [1] with any A v C and A v B trials which may actually exist.

A further feature of MTC is that networks can be extended to include RCTs for which only one, or even none, of the treatments relevant to the decision question of interest are evaluated. Although such evidence may not initially seem relevant to the decision of interest, they can reduce uncertainty

[1] Consistency here implies the underlying effects are assumed to be identical or exchangeable depending whether fixed or random effects are assumed in the synthesis model.

in the comparisons of interest (as well as providing an opportunity for assessing the consistency of the evidence). Therefore, when MTC methods are used, issues relating to the structure and scope of the network require careful consideration.

Currently there is much deliberation regarding the use of MTC methods for HTA with NICE (who provide guidance for England and Wales) now advising they can be used, but not as the base case in their Methods Guide(NICE, 2008). Similar institutions in Australia and Canada are also cautious about their use(Wells et al., 2009, Committee, 2009).

In their simplest form, MTC methods can be viewed as a type of regression model(Lumley, 2002) (Lu and Ades, 2004, Caldwell et al., 2005). As such, it is possible to fit fixed effect versions quite straightforwardly in standard statistical packages, whilst random effect models are probably possible but more challenging to fit. However, many of the methods papers and most practitioners have implemented the models using the (Bayesian Markov Chain Monte Carlo (MCMC)) software WinBUGS(Spiegelhalter et al., 2003). When the MTC is used to provide multiple estimates of effectiveness for a stochastic decision model (e.g. estimates for different treatments and/or time-points etc) care is needed since such estimates will be correlated. In order to maintain such correlations into the decision model, two approaches have been used:

i) A simulation based method of estimating the effectiveness parameters. This could either be achieved by using MCMC or bootstrapping etc. The parameter estimates from each simulation could then be used to inform a decision model evaluated using Monte Carlo (MC) methods taking care to ensure the values from iterations are used simultaneously (since it is this which maintains their correlations).

ii) Evaluate the decision model within the same modelling framework as the synthesis. That is, it is possible to fit both the MTC and the decision model within one single WinBUGS program(Cooper et al., 2004). In this way the correlations between parameters are automatically respected.

Extended MTC methods, which account for outcomes reported at multiple time points (Lu et al., 2007) and allowing for censored data(Welton et al., 2008) have also been developed and implemented using the WinBUGS software and are considered further in this paper. As well as modelling the network of multiple comparisons, they account for other complexities in the data. These potentially allow more data to be included, and assumptions to be made less strong, when compared to standard pair-wise meta-analysis models. However, there is nothing uniquely 'MTC' about these aspects of the more sophisticated modelling, i.e. they could just have easily been applied in a pair-wise context.

The aim of this paper is to compare the use of MTC synthesis models with standard pair-wise meta-analysis models to estimate effectiveness and to inform decision modelling with respect to the i) assumptions the models make, ii) empirical estimates produced, and iii) conclusions drawn. Three case studies are considered: 1) Use of aspirin to prevent stroke in individuals with atrial fibrillation. This case study considers issues surrounding the expansion of the evidence network used with respect to estimates of effectiveness; 2) Use of drug-eluting stents in the percutaneous coronary intervention in patients with coronary artery disease. This example considers the impact of using an MTC allowing for estimates at multiple time-points on an economic decision model; and 3) Use of neuraminidase inhibitors in the treatment of influenza. In this case study, information of censoring is incorporated into a simple evidence network to inform a decision model.

CASE STUDIES:

1) Use of aspirin to prevent stroke in individuals with atrial fibrillation

There have been a number of evidence syntheses conducted to estimate the effect of different interventions for the prevention of stroke in individuals with atrial fibrillation (Singer, 1993, Green et al., 1997, Hart et al., 1999, Segal et al., 2000, Taylor and Ebrahim, 2001, van Walraven et al., 2002, Cooper et al., 2006, Hart et al., 2007). All of these reviews, except one(Cooper et al., 2006), have used standard pairwise meta-analyses to obtain pooled estimates of effectiveness. Cooper et al. used a random effect MTC to estimate effectiveness. The network diagram from this analysis, displaying the interventions that have been considered in RCTs together with the number of times each intervention has been compared to another intervention, is presented in Figure 1 (Cooper et al., 2006).

Using the evidence base associated with the network compiled by Cooper et al., we now explore the implications of using different approaches to the estimation of effectiveness for a single pair-wise comparison.

Assume we are interested in investigating the clinical effectiveness of aspirin compared to placebo. Currently, the standard approach in HTA would be to search and identify the (four) RCTs that directly address this question (i.e. those represented by the solid line in figure 1). A standard random effects meta-analysis of these 4 RCTs results in a pooled estimate of the rate ratio of 0.8 with a 95% confidence interval of 0.1 to 6.6; thus providing a non-conclusive result as to whether aspirin lowers the rate of stroke in individuals with atrial fibrillation compared to placebo (at the 5% statistical significance level).

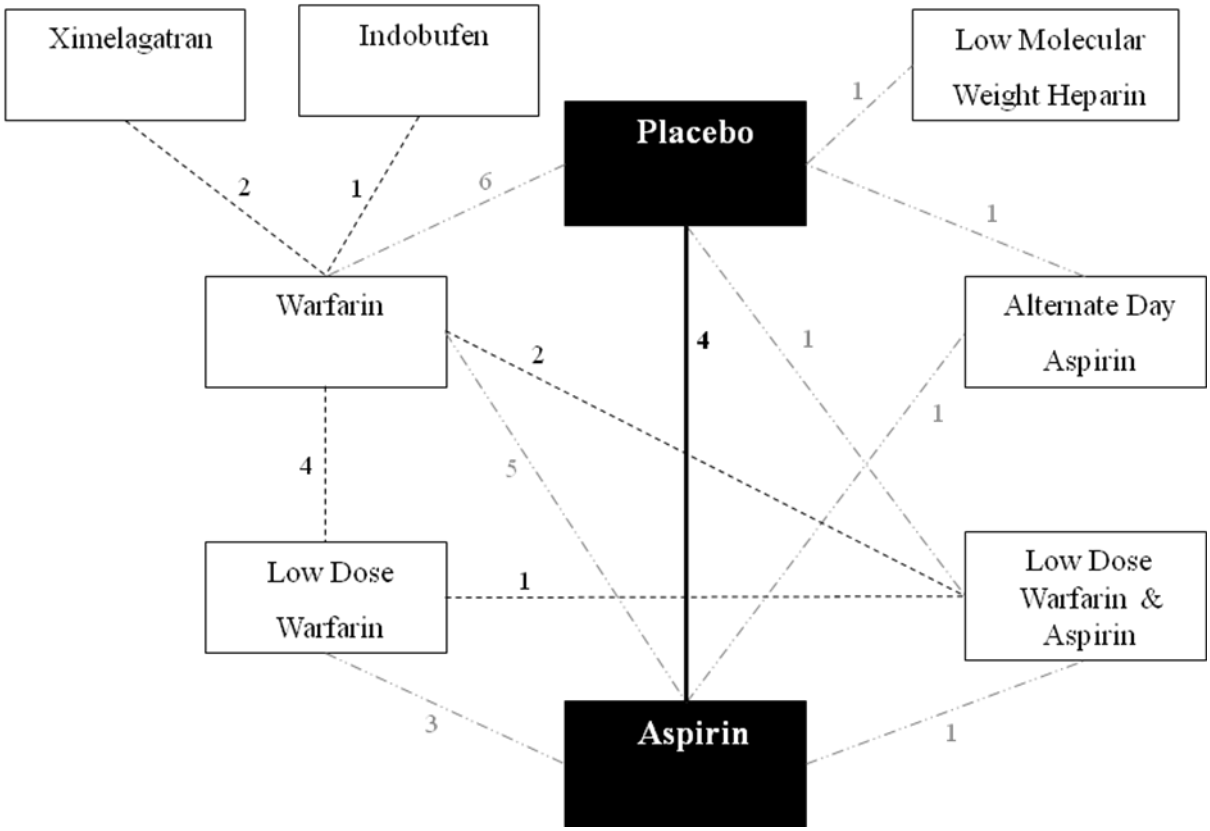


Figure 1: Network diagram of stroke prevention treatments for individuals with non-rheumatic atrial fibrillation. Each treatment strategy is a node in the network. The links between the nodes are trials or pairs of trial arms. The numbers along the link lines indicate the number of trials or pairs of trial arms for that link in the network.

The question we then considered is: How would the estimate and its associated uncertainty change if a broader evidence base were used and MTC methods utilised? The initial network we now consider extends the evidence base to the RCTs in atrial fibrillation which include arms randomising to either of the two treatments of interest, i.e. placebo or aspirin. This introduces the comparisons represented by dot-dashed lines in Figure 1. This extends the network to include 19 further randomised comparisons, which are derived from 12 further RCTs (i.e. some trials included more than two arms and hence made multiple comparisons although this information is not represented on the figure) introducing a further five treatment nodes. Notice that some of these new comparisons form alternative indirect "routes" for comparing placebo to aspirin (i.e. indirect routes via warfarin, alternate day aspirin and low dose warfarin & aspirin all now exist).

Including these 12 RCTs into the evidence synthesis (specific details of the random effects MTC model used, which is based on Poisson regression, is available elsewhere(Cooper et al., 2006)) results in a pooled estimate of the rate ratio of 0.5 with a 95% confidence interval of 0.3 to 0.6. Thus the incorporation of this extra "indirect" evidence has greatly reduced the uncertainty with this

analysis suggesting that aspirin is associated with a statistically significant lower rate of stroke compared to placebo.

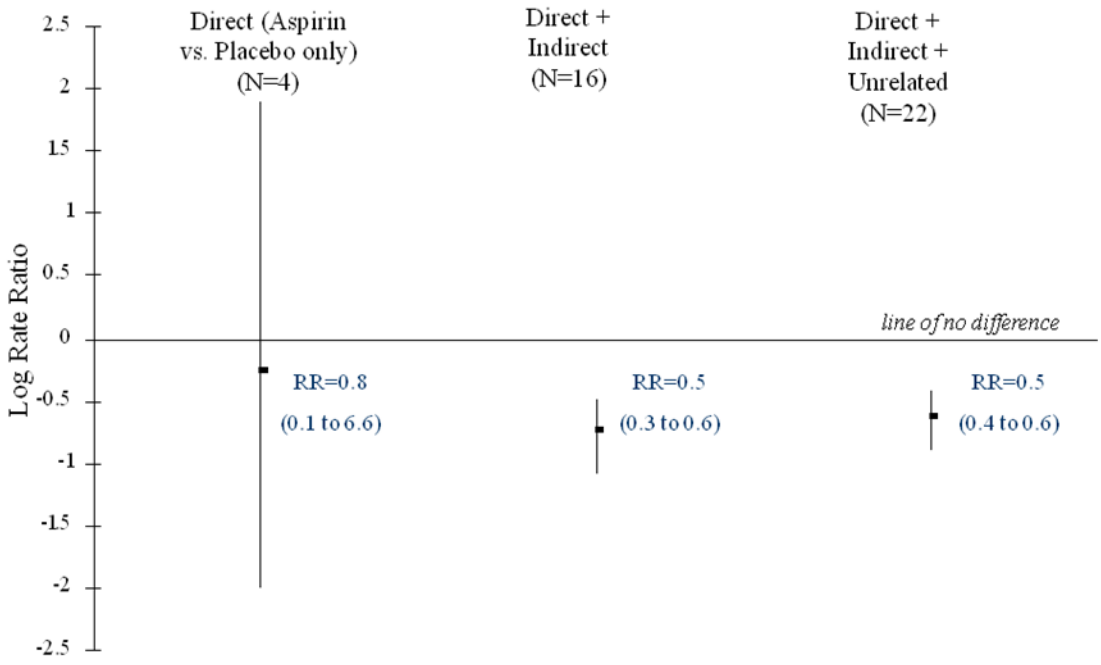


Figure 2: Pooled results obtained from 3 different analyses combining: i) direct evidence only, ii) direct plus indirect evidence, and iii) direct plus indirect plus 'unrelated' evidence.

Extending the network to all RCTs of anticoagulant and antiplatlet therapies that were available when the original analysis was published, by including RCTs which did not consider either placebo or aspirin therapy, adds in the comparisons denoted by dotted lines in Figure 1 resulting in the "full" network. This adds in a further ten comparisons from 6 RCTs and introduces a further two treatment nodes. This further increases the indirect "routes" which connect placebo to aspirin. The pooled rate ratio for the analysis of the whole network is 0.5 with 95% confidence interval 0.4 to 0.6); which has reduced the uncertainty fractionally from the previous analysis. The results from these 3 different analyses are displayed in Figure 2 for easy comparison.

This example clearly illustrates that use of MTC methods can have a large impact on estimates of effectiveness. The inclusion of indirect evidence will generally reduce the uncertainty of an estimate. Concern has been raised regarding whether such estimates are reliable (i.e. unbiased), and this depends on whether the assumptions of the model hold. An initial assessment of the full model(Cooper et al., 2006) suggested that it fit the data well.

For clarity, we have focused on the results of one comparison here. Of course, one of the advantages of MTC is that it can compare all treatments in a connected network simultaneously, and even obtain probabilities that each treatment is best. Case studies 2 and 3 consider decision

problems with more than two alternatives. Clearly, situations such as the one considered above, in which the effectiveness parameter changes considerably with synthesis model could have important implications for any decision model, and subsequent policy decisions, which is sensitive to such parameters. We also consider the impact on the results of the synthesis on the results of decision models in Case studies 2 and 3.

2) Use of drug-eluting stents in the percutaneous coronary intervention in patients with coronary artery disease

In 2007 an NHS Health Technology Assessment was published that assessed the effectiveness and cost-effectiveness of using drug-eluting stents in percutaneous coronary intervention in patients with coronary artery disease(Hill et al.). Pairwise meta-analyses were carried out in the systematic review component of the report for different drug-eluting stent 'designs' compared to bare-metal stents, for a range of outcomes (e.g. mortality, myocardial infarction events, revascularisation), and time points. However, the economic evaluation only considered bare-metal stents versus drug-eluting stents regardless of the 'design' and the outcome of target lesion revascularisation at one year assuming all other outcomes to be equal. The model was evaluated deterministically.

Since the HTA report in 2007, Stettler et al.(Stettler et al., 2007) published a MTC of outcomes associated with drug-eluting stents (Cypher and Taxus) and bare-metal stents. To incorporate data from the range of follow-up times reported by the different studies, their analysis used a random walk model based on piece-wise constant hazards.(Lu et al., 2007) As well as incorporating data from multiple time points, this random walk MTC also allowed estimates of effectiveness to be more similar to estimates at adjacent time points than estimates at 'far-off' time points. It also allowed the two types of drug eluting stents (Cypher and Taxus) to be more similar to each other than to bare metal stents. Figure 3 shows the network diagram for this MTC analysis but note that the numbers on the diagram only relate to the number of trials reporting target lesion revascularisation at one year, as this number will vary across time points.

Here we consider the following three evidence synthesis models to estimate the pooled odds ratio for target lesion revascularisation in individuals with Cypher or Taxus drug-eluting stents versus bare-metal stents:

i) Pairwise random effects meta-analysis of Taxus or Cypher vs. Bare-metal stents (**M-A**)

ii) Random effects MTC of Taxus vs. Cypher vs. Bare-metal stents (**MTC**)

iii) Random walk MTC (using data from multiple time points) of Taxus vs. Cypher vs. Bare-metal stents (**Random walk MTC**)(Stettler et al., 2007)

All of the above analyses include the same trial data used in the Stettler et al. analysis (Stettler et al., 2007). Pooled estimates are then input as distributions into the economic decision model, developed as part of the 2007 HTA(Hill et al., 2007), to assess the effect of using different evidence synthesis models, if any, on the overall cost-effectiveness result. For the first analysis listed above (M-A), the meta-analysis was fitted in Stata and then the results input into a spreadsheet model. This two-stage approach was adopted to emulate conventional modelling practice. In the second and third analyses both the meta-analysis and the decision model were evaluated in WinBUGS within a single model. Unlike the 2007 HTA model, for all analyses the decision model is evaluated stochastically (i.e. all model parameters are expressed as distributions).

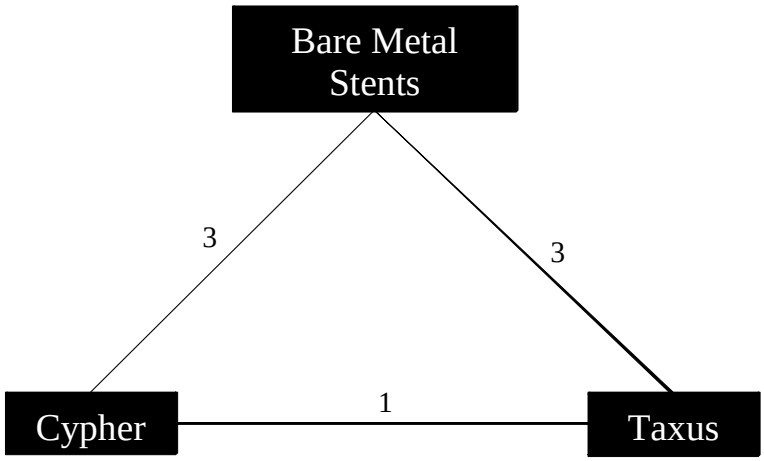
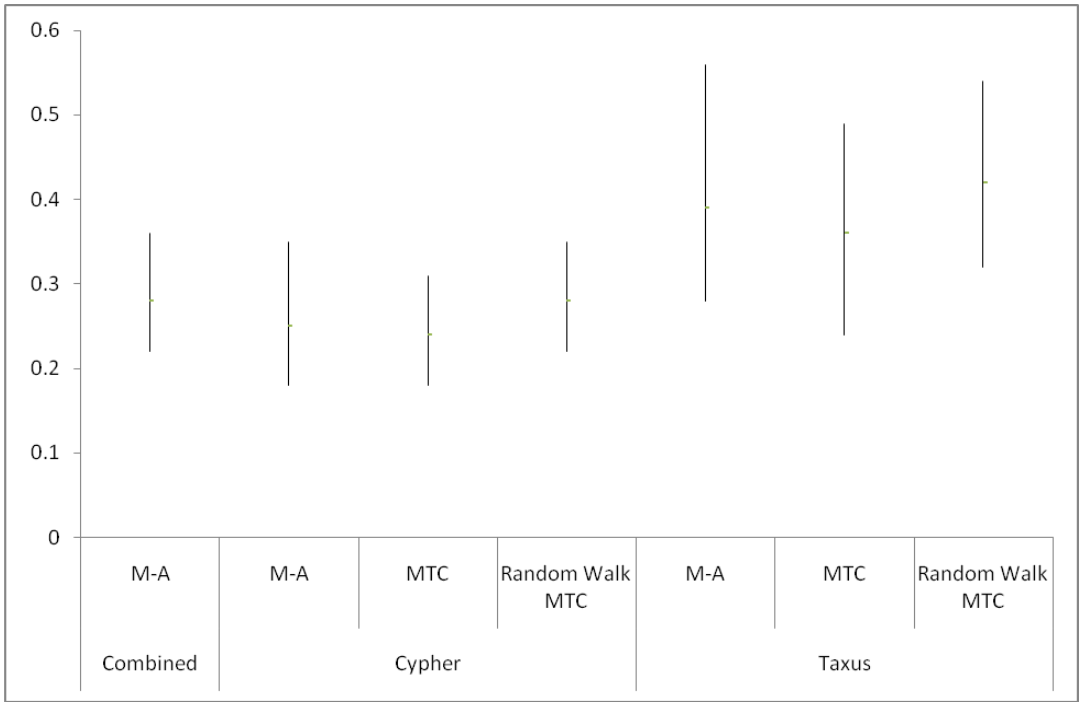


Figure 3: Network diagram of stents used in percutaneous coronary intervention in patients with coronary artery disease. Each treatment strategy is a node in the network. The links between the nodes are trials or pairs of trial arms. The numbers along the link lines indicate the number of trials or pairs of trial arms for that link in the network with target lesion revascularisation data at one year.

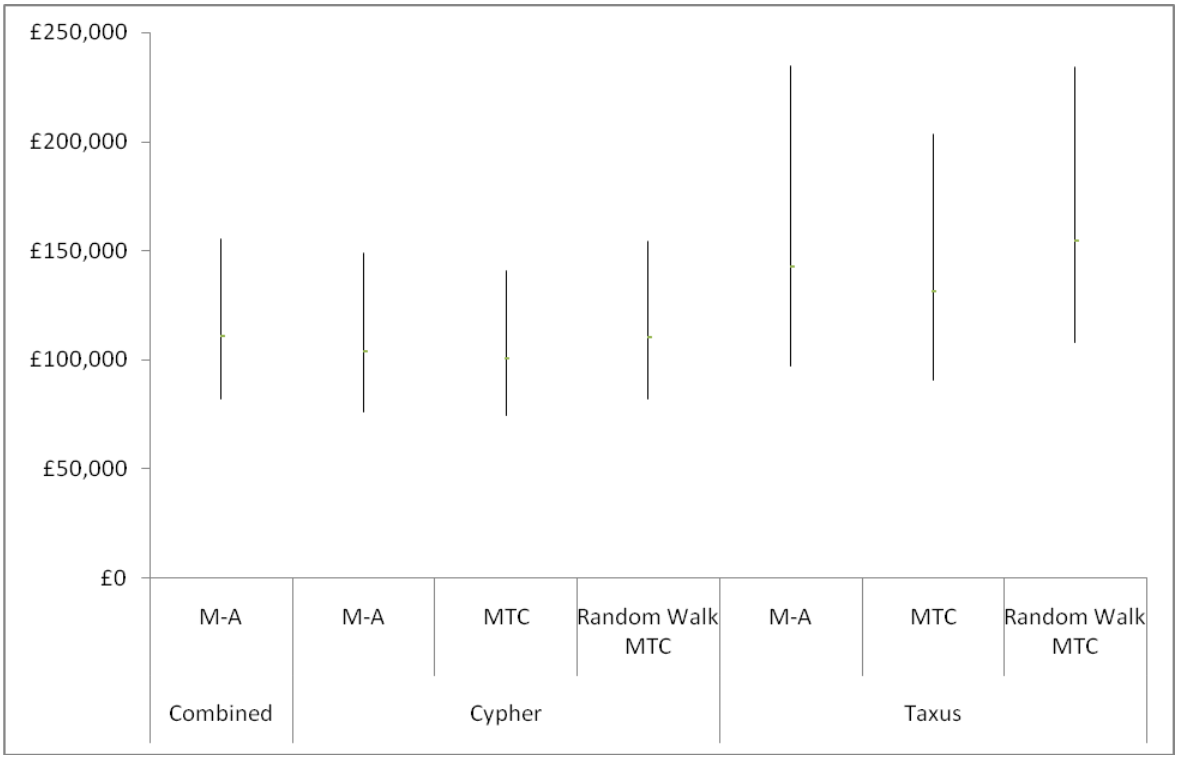
The results from the 3 different analyses are presented in Figure 4. Panel (a) displays the pooled odds ratio for target lesion revascularisation in individuals with Cypher or Taxus drug-eluting stents versus bare-metal stents. It shows the point estimates to be very similar for the *M-A* and *MTC* models with a slight increase estimated for the *random walk MTC* model; however, the uncertainty is reduced in the *MTC* and *random MTC* models compared to *M-A* model (depicted by the narrower confidence/credible intervals). Panel (b) in Figure 4 shows the incremental cost-effectiveness ratios (ICERs) obtained when the different evidence synthesis models are used to inform the effectiveness parameter in the decision model. As above, the point estimates are similar for the *M-A* and *MTC* models with a slight increase estimated for the *random walk MTC* model; however, the uncertainty is reduced slightly in the *MTC* model compared to *M-A* and *random walk MTC* models (depicted by the narrower confidence/credible intervals).

Figure 5 shows the cost-effectiveness acceptability curves (CEACs) for the *MTC* and *random walk MTC* models for Cypher compared to bare metal stents and Taxus compared to bare metal stents. It

can be observed that incorporating the results of the *random walk MTC* into the decision model lowers the CEAC due to the increased uncertainty in the estimation of the effectiveness parameters.



(a) Pooled odds ratio for target lesion revascularisation



(b) Incremental cost-effectiveness ratios

Figure 4: Pooled odds ratio for target lesion revascularisation and incremental cost-effectiveness ratios together with 95% confidence intervals obtained from the 3 different analyses (MA, MTC and Random Walk MTC) for Cypher, Taxus and Combined versus bare metal stents.

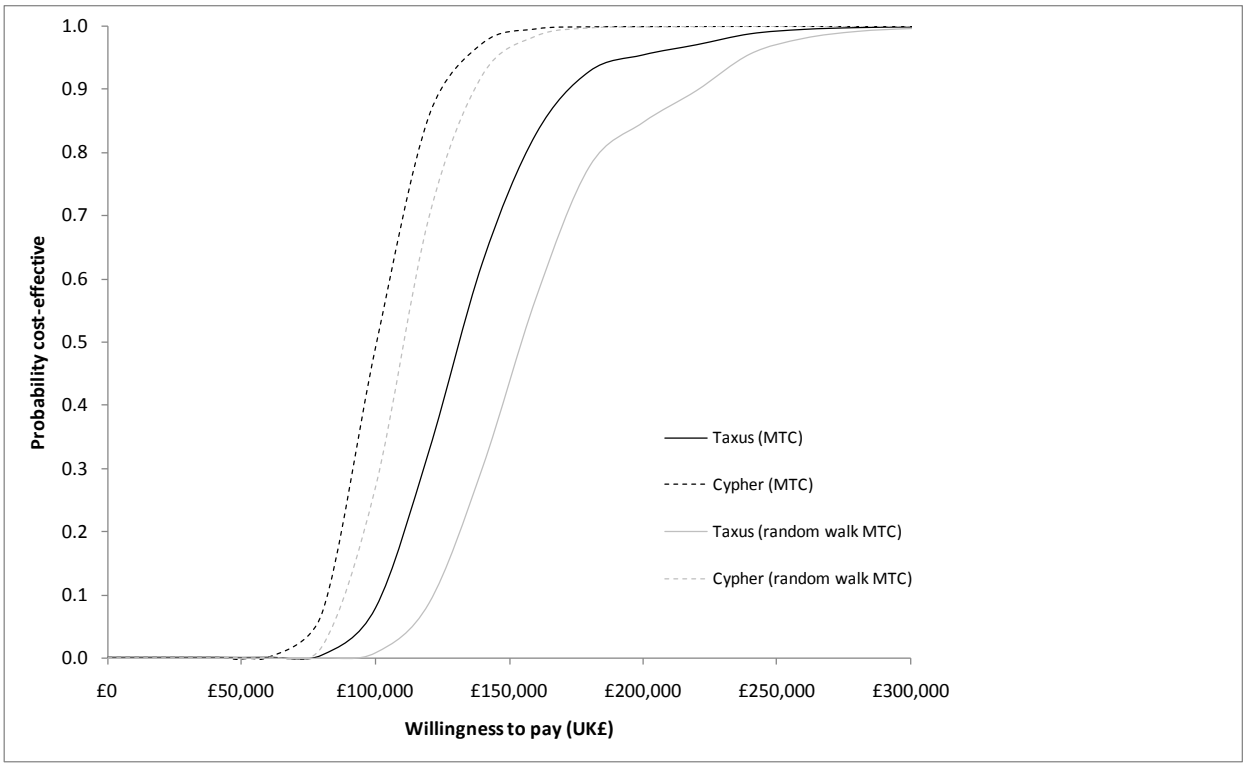


Figure 5: Cost-effectiveness acceptability curves for the stents example

3) Use of neuraminidase inhibitors in the treatment of influenza

The third case study revisits a recent HTA conducted to inform NICE guidance on the use of antiviral drugs for the treatment of influenza(Burch et al., 2009). This assessment was done to provide an update to previous guidance published in 2000(Turner et al., 2003) by incorporating new evidence.

In the recent assessment, two antiviral drugs, both Neuraminidase Inhibitors (NIs) - Zanamivir and Oseltamivir – were appraised and compared to usual care and each other.

The evidence-base considered in both appraisals is relatively straightforward. There existed a number of RCTs of each of the two NIs versus placebo. There were no head-to-head trials of the two NIs nor were there any trials of the NIs versus any other active comparator. Therefore, the network structure was relatively straightforward allowing an indirect comparison to be made between Zanamivir and Oseltamivir. (Figure 6).

In the earlier appraisal(Turner et al., 2003), standard pair-wise meta-analyses were conducted to estimate the effectiveness parameters for the economic model – the mean difference in time to i) the alleviation of symptoms; and ii) return to work. This was unexpectedly challenging because a proportion of patients had not recovered by the end of the RCTs resulting in censoring of data

which meant mean durations were not defined (clinically, focus had been on the difference in *median* times and this was estimable from all trials because censoring was always on less than 50% of patients).

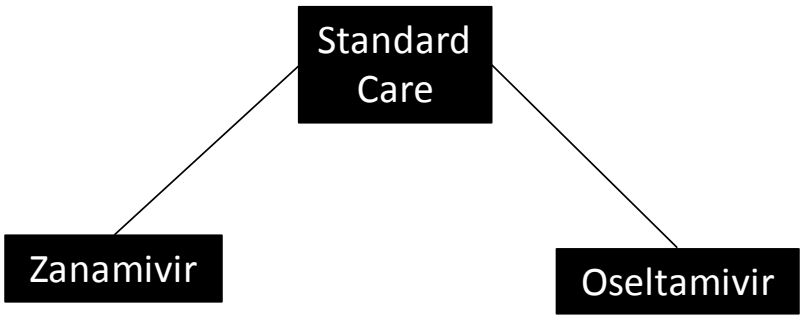


Figure 6: Network diagram of antiviral drugs for the treatment of influenza assessment.

In order to estimate mean durations (and associated standard errors) in each of the arms of the RCTs in the presence of the censoring, it was assumed that the 'survival' curves for times to alleviation of symptoms / return to work followed an exponential distribution(Collett, 2003).

For the updated assessment, a more sophisticated approach was undertaken (with further details available elsewhere(Burch et al., 2009)) building on an extended analysis of the data available in the previous assessment(Welton et al., 2008). This analysis relaxes the assumption that the 'survival' curves are exponential in shape. It does this by fitting the more flexible (i.e. 2-parameter) Weibull distribution to the 'survival' curves for both outcomes using the median data. In doing so, it takes into account further data which was available on the numbers still ill at the end of the reported follow-up since this informs a second 'point' on the time to alleviation of symptoms / return to normal activities 'survival' curve. The analysis models both outcomes simultaneously so information can be 'borrowed' across outcomes for RCTs that do not report both outcomes. Further, four specific patient subgroups were considered distinctly in the economic evaluation - otherwise healthy adults, otherwise healthy children, 'at risk' individuals (i.e. individuals of any age with a concurrent disease severe enough to require regular medical follow-up or hospital care (for example, chronic disorders such as chronic respiratory disease, cardiovascular disease, and pulmonary disorders) plus otherwise healthy elderly individuals aged 65 years and above) and mixed population (the latter was used to include patients in trials where it was not possible to obtain stratified results for the subgroups of interest). These were also simultaneously modelled assuming exchangeability across each of the treatment/subgroup combinations which allows a 'borrowing of strength' which reduces the precision of the subgroup-specific estimates. This model was fitted using WinBUGS and 5000 MCMC samples from the relevant posterior distributions imported into Excel for use in the associated decision model which was evaluated using Monte Carlo methods.

Hence, although this model automatically provides indirect comparison estimates for Zanamivir versus Oseltamivir, given the simple evidence structure, such an estimate could be obtained classically using standard indirect comparison methodology(Bucher et al., 1997), and thus this is not the main advantage of the sophisticated modelling. Rather, it is in the other modelling complexities outlined above which sets this analysis apart from the frequentist approach.

For this case study, we compare the effectiveness and cost-effectiveness results obtained from this recent assessment using the complex Bayesian evidence synthesis model with the simpler frequentist meta-analysis methods for evidence synthesis (i.e. which were very similar to those used in the previous assessment), which assumed an exponential distribution for each 'survival' curve and estimated parameters for each treatment/subgroup combination independently but using the same RCT evidence as the Bayesian analysis (although the data on number of participants still ill at the end of each trial could not be incorporated). For both analyses the structure of the economic model was identical, with only the effectiveness parameter estimates varying between analyses.

Tables 1 and 2 report the estimates of treatment effect for each NI versus placebo and for the 2 NIs head to head for time to alleviation of symptoms and time to return to normal activities respectively. Forest plots for the 'at risk' populations for all four outcome / treatment combinations are provided in the four panels of Figure 7 to give a graphical display of a proportion of the data in the analysis. Here the estimate of effect from the Bayesian analysis is also plotted for easy comparison.

Subgroup	**Treatment**	**CLASSICAL METHOD:** **Mean Difference in symptoms duration (95% CI)**	**BAYESIAN METHOD:** **Mean Difference in symptoms duration (95% CI) (Burch et al.)**
Otherwise Healthy Adults	Zanamivir vs. Placebo	-0.905 (-1.458, -0.353)	-1.30 (-2.96, -0.30)
	Oseltamivir vs. Placebo	-1.309 (-2.029, -0.588)	-2.08 (-4.34, -0.73)
	Zanamivir vs. Oseltamivir	0.404 (-0.504, 1.312)	0.78 (-0.53, 2.51)
At Risk	Zanamivir vs. Placebo	-2.801 (-4.478, -1.124)	-4.70 (-9.44, -1.98)
	Oseltamivir vs. Placebo	-1.020 (-2.236, 0.196)	-1.56 (-4.66, 0.78)
	Zanamivir vs. Oseltamivir	-1.781 (-3.852, 0.29)	-3.14 (-7.73, -0.44)
Otherwise Healthy Children	Zanamivir vs. Placebo	-1.443 (-2.923, -0.038)	-1.77 (-5.10, 0.41)
	Oseltamivir vs. Placebo	-2.164 (-3.519, -0.809)	-2.63 (-6.53, -0.38)
	Zanamivir vs. Oseltamivir	0.721 (-1.286, 2.728)	0.86 (-2.01, 4.47)

Table 1: The mean difference in days (95%CI) for the time to symptoms alleviation in FLU POSITIVE population

Subgroup	Treatment	CLASSICAL METHOD: Mean Difference in the time to normal activities (95% CI)	BAYESIAN METHOD: Mean Difference in the time to normal activities (95% CI) (Burch et al.)
Otherwise Healthy Adults	Zanamivir vs. Placebo	-0.977 (-1.915, -0.039)	-1.65 (-3.94, -0.35)
	Oseltamivir vs. Placebo	-3.506 (-5.478, -1.533)	-2.64 (-5.83, -0.86)
	Zanamivir vs. Oseltamivir	2.529 (0.345, 4.713)	0.99 (-0.66, 3.29)
At Risk	Zanamivir vs. Placebo	-1.155 (-3.408, 1.098)	-5.97 (-12.65, -2.32)
	Oseltamivir vs. Placebo	-1.618 (-3.269, 0.033)	-1.98 (-6.09, 0.99)
	Zanamivir vs. Oseltamivir	0.463 (-2.33, 3.256)	-3.99 (-10.15, -0.55)
Otherwise Healthy Children	Zanamivir vs. Placebo	-0.721 (-2.606, 1.163)	-2.25 (-6.66, 0.52)
	Oseltamivir vs. Placebo	-2.669 (-3.703, -1.635)	-3.34 (-8.60, -0.47)
	Zanamivir vs. Oseltamivir	1.948 (-0.201, 4.097)	1.09 (-2.56, 5.78)

Table 2: The mean difference in days (95%CI) for the time to return to normal activities in FLU POSITIVE population:

Across all the outcomes, the NIs are generally estimated to be associated with a larger treatment difference compared to placebo in the Bayesian synthesis compared to the frequentist meta-analysis. The uncertainty is usually greater in the Bayesian analysis also reflected in the wider, and often asymmetric, credible intervals. The uncertainty was also generally greater in the head-to-head comparisons from the Bayesian synthesis model although the difference between the point estimates from the two models was less predictable. In summary, while unpredictable, the differences between the two analyses are quite considerable for a proportion of the estimates.

Figure 8 presents the acceptability curves for an at-risk adults population resulting from a decision model using the Bayesian synthesis model and the simple pair-wise meta-analysis for effectiveness inputs. It can be seen that for a willingness to pay of £5,000 per QALY gained or above, Zanamivir would appear to be the most cost-effective intervention for this subgroup regardless of approach used to estimate effectiveness, although the Bayesian synthesis model increases the difference between the two NI treatments by approximately 20% (after £5,000 per QALY gained). Clearly, these sorts of differences could impact on model conclusions in situations where acceptability curves for different treatments are closer together.

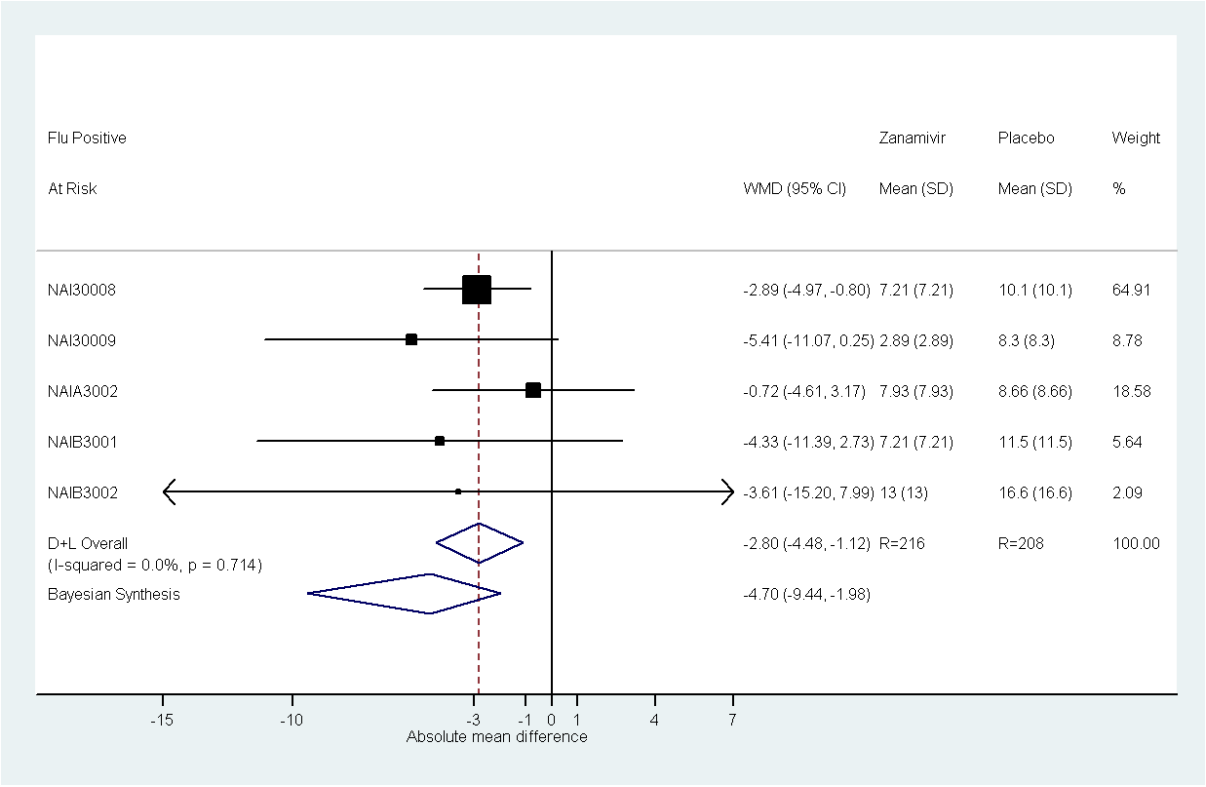


Figure 7a: Time to symptoms alleviation. Flu positive 'At-risk' in the Zanamivir treatment trials

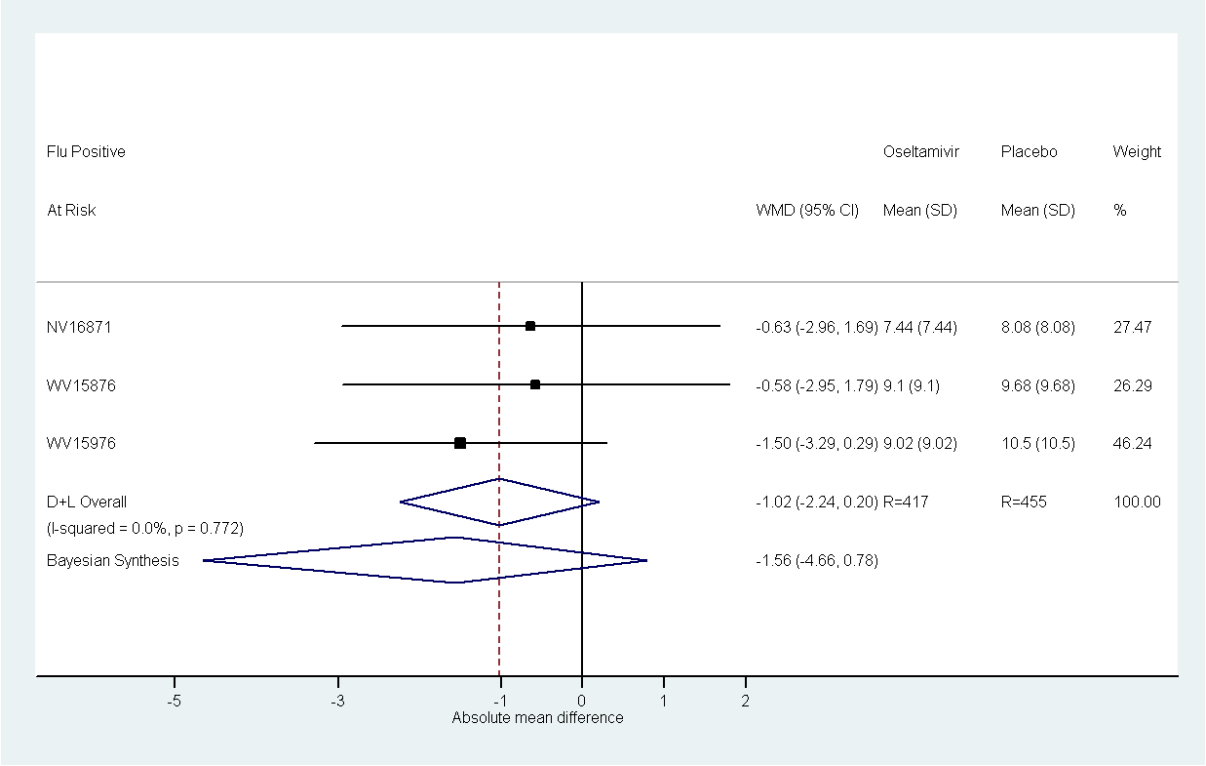


Figure 7b: Time to symptoms alleviation. Flu positive 'At-risk' in the Oseltamivir treatment trials

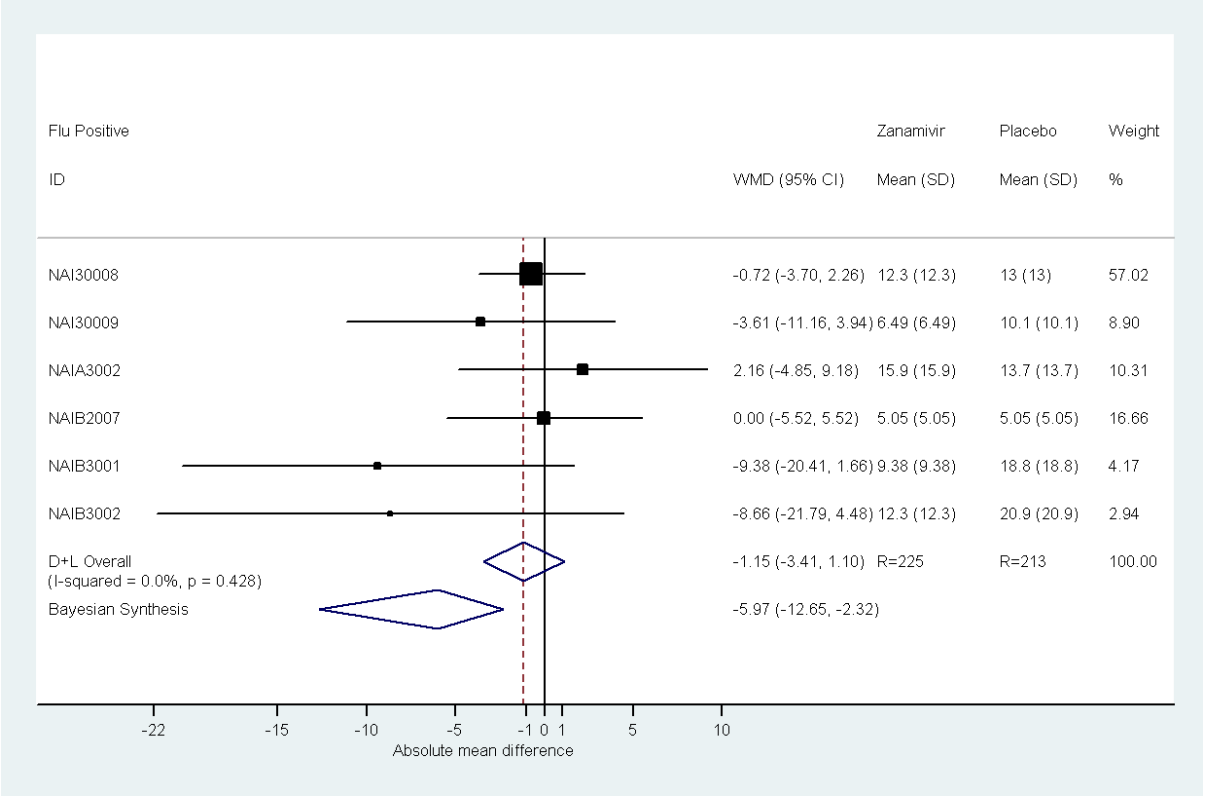


Figure 7c: Time to return to normal activities. Flu positive 'At-risk' in the Zanamivir treatment trials

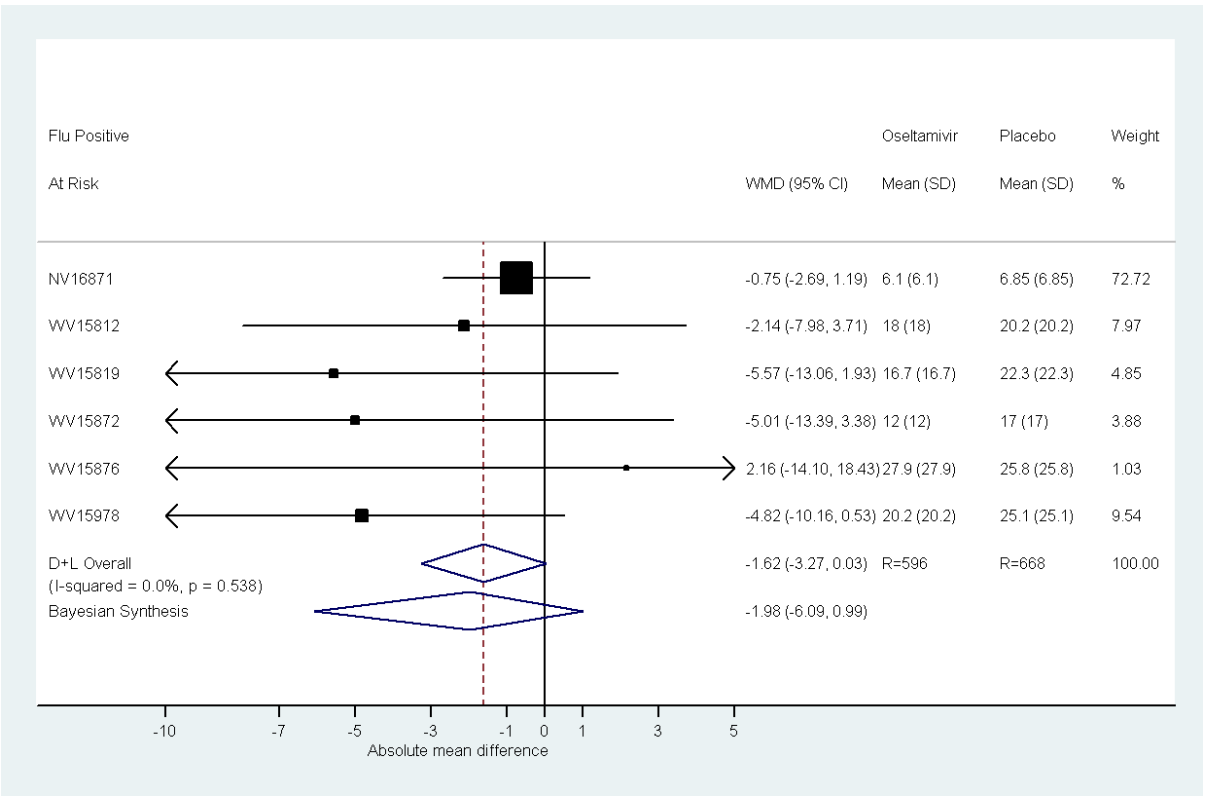


Figure 7d: Time to return to normal activities. Flu positive 'At-risk' in the Oseltamivir treatment trials

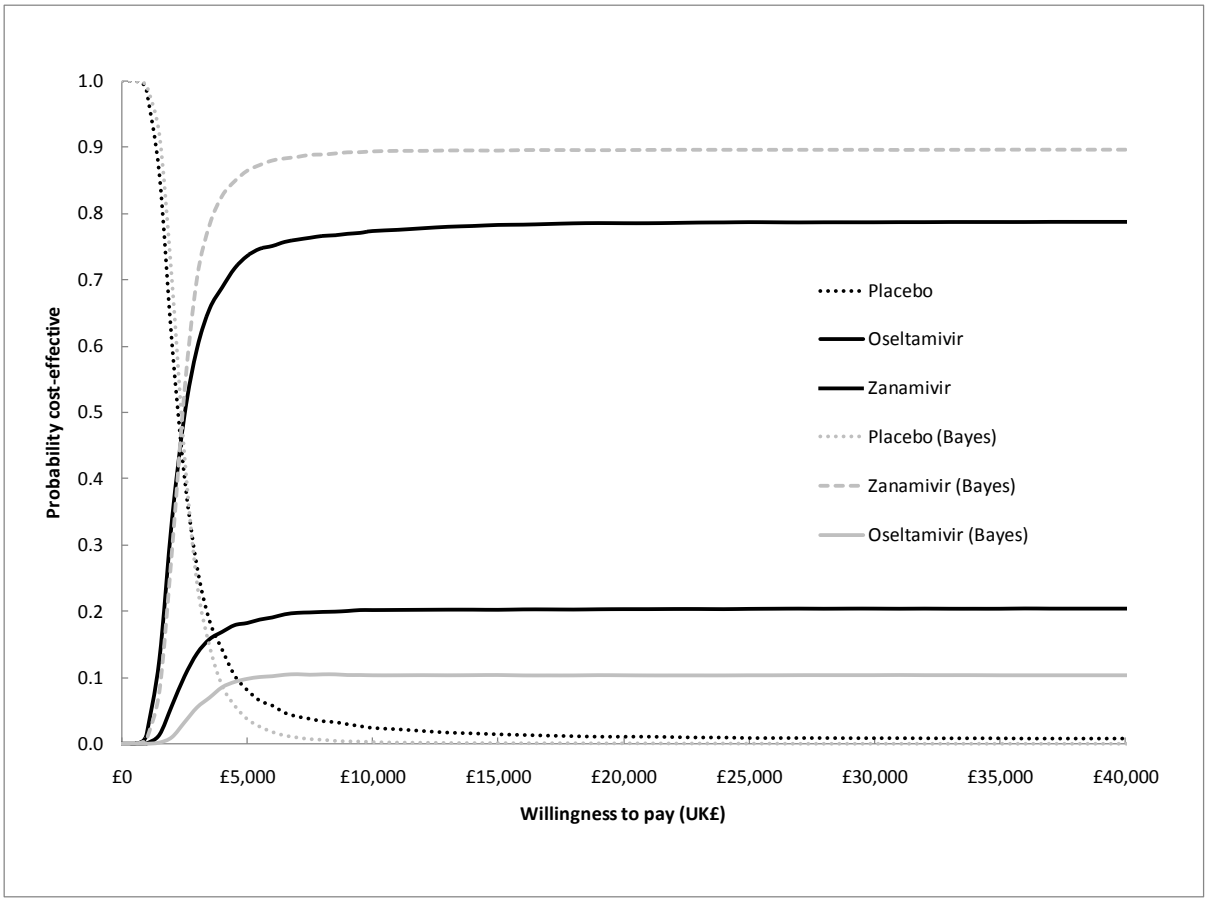


Figure 8: Cost-effectiveness acceptability curves for at-risk adults in the influenza example.

DISCUSSION:

In this paper we have attempted to compare three state-of-the-art Bayesian synthesis models with more standard (pair-wise) meta-analytic approaches for estimation of clinical effectiveness. Since the latter two case studies consider MTC models with further modelling refinements, these comparisons are not "pure" MTC versus pair-wise synthesis but 'complex' Bayesian evidence synthesis versus pair-wise meta-analysis.

Important findings that these case studies highlight include: i) MTC methods allow the inclusion of evidence that is ignored in pair-wise modelling, and this inclusion of further evidence generally reduces the uncertainty in effectiveness parameters; ii) Imposing hierarchical structures on data (e.g. to allow for multiple time points etc) can decrease uncertainty through the inclusion of extra evidence. However, other changes in the evidence synthesis modelling considered here meant this was not always observed (e.g. in example 3, in order to include a second time-point to estimate a survival curve, the curve was assumed to follow a Weibull rather than Exponential shape); and iii) Both MTC and hierarchical aspects of the synthesis modelling can change point estimates, and it is

difficult to anticipate by how much and in which direction prior to carrying out the evidence synthesis and thus the impact on the results of the cost-effectiveness models.

A common, but un-workably vague, phrase in guidance for decision models is that "all relevant evidence" should be used to inform (effectiveness) model parameters(NICE, 2008). We believe this paper highlights just how difficult it is to come up with a workable definition of relevant evidence, but one would clearly need to address issues relating to trial networks (i.e. as shown in case study 1), and time points (i.e. case studies 2 and 3). Both of these issues relate to evidence that may impact on the effectiveness parameters of interest, although it may not be immediately obvious that such evidence is "relevant". Further, this paper has only considered randomised evidence, although it is acknowledged that observational evidence, or even expert opinion, may sometimes be considered "relevant", for example in situations where there is no or limited trial evidence, or where the trial evidence may not relate to the patient populations being considered in the decision modelling. We are currently exploring further case studies where we compare the results of using trial data only with results obtained by using trial data augmented with observational data and expert opinion to further explore these issues.

The three case studies considered here compare different approaches to estimating effectiveness parameters for use in decision models. Many of the authors of this paper have research interests in MTC methods and Bayesian synthesis modelling more generally, and thus an enthusiasm for the more complex approaches considered here. However, it is important to note that the case studies cannot demonstrate one method is superior to the other with respect to bias and precision of parameter estimates since no 'gold standard' approach exists and thus there is no way of knowing what the "truth" is. What we can say is that the more complex approaches consider the evidence as a coherent whole, include more data, and sometimes relax the assumptions made in the pair-wise approaches. Against this is the acknowledgement that the more complex models can be very non-intuitive to understand and time-consuming to undertake.

ACKNOWLEDGEMENTS:

We would like to thank Professor Adrian Bagust for providing the necessary information which allowed us to replicate the original decision model for Case study 2.

REFERENCES:

BUCHER, H. C., GUYATT, G. H., HUTCHINSON, A. & ECCLES, M. (1997) The results of direct and indirect treatment comparisons in meta-analysis of randomised controlled trials. *Journal of Clinical Epidemiology,* 50**,** 683-691.

BURCH, J., PAULDEN, M., CONTI, S., STOCK, C., CORBETT, M., WELTON, N., ADES, A. E., SUTTON, A., COOPER, N., ELLIOT, A., NICHOLSON, K., DUFFY, S., MCKENNA, C., WESTWOOD, M. & PALMER, S. (2009) Antiviral drugs for the treatment of influenza: A Systematic Review and Economic Evaluation. . *NICE Appraisal,* 168.

CALDWELL, D. M., ADES, A. E. & HIGGINS, J. P. T. (2005) Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *BMJ,* 331**,** 897-900.

COLLETT, D. (2003) *Modelling survival data in medical research,* New York, Chapman and Hall/CRC.

COOPER, N. J., SUTTON, A. J., ABRAMS, K. R., TURNER, D. & WAILOO, A. (2004) Comprehensive decision analytical modelling in economic evaluation: a Bayesian approach. *Health Economics,* 13**,** 203-226.

COOPER, N. J., SUTTON, A. J., LU, G. & KHUNTI, K. (2006) Mixed comparison of stroke prevention treatments in individuals with non-rheumatic atrial fibrillation. *Archives of Internal Medicine,* 166**,** 1269-1275.

GREEN, C. J., HADORN, D. C., BASSETT, K. & KAZANJIAN, A. (1997) Anti-coagulation in chronic nonvalvular atrial fibrillation: A critical appraisal and meta-analysis. *Canadian Journal of Cardiology,* 13**,** 811-815.

HART, R. G., BENAVENTE, O., MCBRIDE, R. & PEARCE, L. A. (1999) Antithrombotic therapy to prevent stroke in patients with atrial fibrillation: A meta-analysis. *Annals of Internal medicine,* 131**,** 492-501.

HART, R. G., PEARCE, L. A. & AGUILAR, M. I. (2007) Meta-analysis: Antithrombotic therapy to prevent stroke in patients who have nonvalvular atrial fibrillation. *Annals of Internal Medicine,* 146**,** 857-867.

HILL, R. A., BOLAND, A., DICKSON, R., DÜNDAR, Y., HAYCOX, A., MCLEOD, C., MUJICA MOTA, R., WALLEY, T. & BAGUST, A. (2007) Drug-eluting stents: a systematic review

and economic evaluation. *Health Technology Assessment,* 11.

LU, G. & ADES, A. E. (2004) Combination of direct and indirect evidence in mixed treatment comparisons. *Statistics in Medicine,* 23**,** 3105-3124.

LU, G., ADES, A. E., SUTTON, A. J., COOPER, N. J., BRIGGS, A. H. & CALDWELL, D. M. (2007) Meta-analysis of mixed treatment comparisons at multiple follow-up times. *Statistics in Medicine,* 26**,** 3681–99.

LUMLEY, T. (2002) Network meta-analysis for indirect treatment comparisons. *Statistics in Medicine,* 21**,** 2313-2324.

NICE (2008) Guide to the methods of technology appraisal. *National Institute for Health and Clinical Excellence*.

PHARMACEUTICAL BENEFITS ADVISORY COMMITTEE. (2009) Report of the Indirect Comparisons Working Group to the Pharmaceutical Benefits Advisory Committee:Assessing indirect comparisons. Australia.

SEGAL, J. B., MCNAMARA, R. L., MILLER, M. R., KIM, N., GOODMAN, S. N., POWE, N. R., ROBINSON, K. A. & BASS, E. B. (2000) Prevention of thromboembolism in atrial fibrillation: A meta-analysis of trials of anti-coagulants and antiplatelet drugs. *Journal of General Internal Medicine,* 15**,** 56-67.

SINGER, D. E. (1993) Overview of the randomised controlled trials to prevent stroke in atrial fibrillation. *Annals of Epidemiology,* 3**,** 563-567.

SPIEGELHALTER, D., THOMAS, A., BEST, N. & LUNN, D. (2003) *WinBUGS user manual: Version 1.4,* Cambridge, MRC Biostatistics Unit.

STETTLER, C., WANDEL, S., ALLEMANN, S., KASTRATI, A., MORICE, M. C., SCHÖMIG, A., E PFI STERER, M., STONE, G. W., LEON, M. B., SUAREZ DE LEZO, J., GOY, J. J., PARK, S.-J., SABATÉ, M., SUTTORP, M. J., KELBAEK, H., SPAULDING, C., MENICHELLI, M., VERMEERSCH, P., DIRKSEN, M. T., CERVINKA, P., PETRONIO, A. S., NORDMANN, A. J., DIEM, P., MEIER, B., ZWAHLEN, M., REICHENBACH, S., TRELLE, S., WINDECKER, S. & JÜNI, P. (2007) Outcomes associated with drug-eluting and bare-metal stents: a collaborative network meta-analysis. *Lancet,* 370**,** 937-48.

SUTTON, A. J., ABRAMS, K. R., JONES, D. R., SHELDON, T. A. & SONG, F. (2000) *Methods for meta-analysis in medical research,* Chichester, England, John Wiley & Sons, Ltd.

SUTTON, A. J., ADES, A. E., COOPER, N. J. & ABRAMS, K. R. (2008) Use of indirect and mixed treatment comparisons for technology assessment. . *Pharmacoeconomics,* 26**,** 753-767.

TAYLOR, F. C. & EBRAHIM, S. (2001) Systematic review of long term anticoagulation or antiplatelet treatment in patients with non-rheumatic atrial fibrillation. *BMJ,* 322**,** 321-326.

TURNER, D., WAILOO, A., NICHOLSON, K., COOPER, N. J., SUTTON, A. J. & ABRAMS, K. R. (2003) Systematic review and economic decision modelling for the prevention and treatment of influenza A and B. *Health Technology Assessment Report,* 7.

VAN WALRAVEN, C., HART, R. G., SINGER, D. E., LAUPACIS, A., CONNOLLY, S., PETERSEN, P., KOUDSTAAL, P. J., CHANG, Y. & HELLEMONS, B. (2002) Oral anticoagulants vs aspirin in nonvalvular atrial fibrillation: An individual patient meta-analysis. *JAMA,* 288**,** 2441-2448.

WELLS, G. A., SULTAN, S. A., CHEN, L., KHAN, M. & COYLE, D. (2009) Indirect evidence: Indirect treatment comparisons in meta-analysis. *Ottawa: Canadian Agency for Drugs and Technologies in Health*.

WELTON, N. J., COOPER, N. J., ADES, A. E., LU, G. & 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**,** 5620-5639.