Paper Number P21

How is evidence on test performance synthesised in economic decision models of diagnostic tests? A systematic appraisal of Health Technology Assessments in the UK since 1997.

Nicola Novielli, Nicola J. Cooper, Alex J. Sutton, Keith R. Abrams

Department of Health Sciences, University of Leicester, UK.

Contact details

Nicola Novielli: Department of Health Sciences, Adrian Building, University of Leicester, University Road, Leicester LE1 7RH, UK. E-mail: nn40@le.ac.uk

WORK IN PROGRESS - DO NOT QUOTE WITHOUT PERMISSION

Abstract

Background: Evaluation of the performance of diagnostic tests in both clinical and economic terms has received less attention than for interventions. The methodology required is potentially more complex due to additional issues relating to i) threshold levels of tests; and ii) dependence between sensitivity and specificity.

Aims**:** To assess how evidence on diagnostic test accuracy is synthesised and used to inform economic-decision modelling for HTA.

Methods**:** All reports evaluating diagnostic test via an economic-decision model published by the NHS Research and Development Health Technology Assessment (HTA) programme since 1997 were identified. The methods for evidence synthesis of diagnostic test accuracy data and its use in economic-decision modelling in this sample were reviewed.

Results**:** Forty-four HTA reports out of 474 concern diagnostic accuracy, of which 11 did not do any economic evaluation. Of the remaining 33 HTAs, 14 conducted meta-analyses of diagnostic accuracy in the clinical review but only 8 used such pooled estimates to inform the decision model. A number of meta-analysis methods ranging in complexity were applied to estimate diagnostic accuracy. However, when it came to informing the economic-decision model, the majority of reviews used independent meta-analytic estimates of sensitivity and specificity.

Conclusions**:** Often, very simplistic methods to estimate diagnostic test accuracy were used for purposes of informing an economic-decision model. The assumptions made by the simplistic methods are usually invalid which may lead to sub-optimal decisions being made. It is desirable that decision modellers become aware of the rapid evolution of meta-analysis methods in this area, however further research is still required to identify how the pooled results obtained from the different meta-analysis models should best be used to inform economic-decision models.

1 Introduction

The creation of structures in the UK (i.e. National Institute for Health and Clinical Excellence) and elsewhere to facilitate evidence-based health policy decision-making has highlighted the role that systematic reviews including, where appropriate, meta-analysis, and economic evaluations have to play in the decision-making process. These methodologies provide answers to fundamental questions such as: Does the technology work, for whom, at what cost, and how does it compare with alternatives?(NICE, 2008) In the area of diagnostic test performance such evidence-based evaluations are crucial to the decision making process as early diagnosis can lead to diseases being treated more successfully than if treatment were delayed.

Evidence synthesis of diagnostic accuracy data are more complicated than for intervention studies due to additional issues relating to variable test threshold levels, dependence between sensitivity (i.e. the proportion of people with the condition correctly detected by the test) and specificity (i.e. the proportion of people without the condition correctly detected by the test) and use of multiple tests in combination. To date, at least five different approaches to the meta-analysis of diagnostic test performance data have been developed each producing results in different formats (Figure 1):

i) Separate meta-analyses of sensitivity and specificity assuming independence of the two (Deeks and Altman, 2001). Since there is usually a trade-off to be made between sensitivity and specificity which is done by varying the threshold value used to categorise diseased and non diseased, this method is only valid if all primary studies report test performance using the same threshold. This method produces a summary point on the ROC plane for the pooled sensitivity and specificity (Figure 1 panel (i)).

ii) Meta-analysis of diagnostic odds ratios (Deeks and Altman, 2001). This method relaxes the assumption that all studies use the same test threshold. In doing so, this method produces a symmetric summary receiver operating curve (sROC) (Figure 1 panel (ii)).

iii) The regression modelling approach of Littenberg and Moses (Littenberg and Moses, 1993). This approach extends the previous one parameter model to a two parameter model; that is, a further parameter is included to allow the sROC to be asymmetric (Figure 1 panel (iii)).

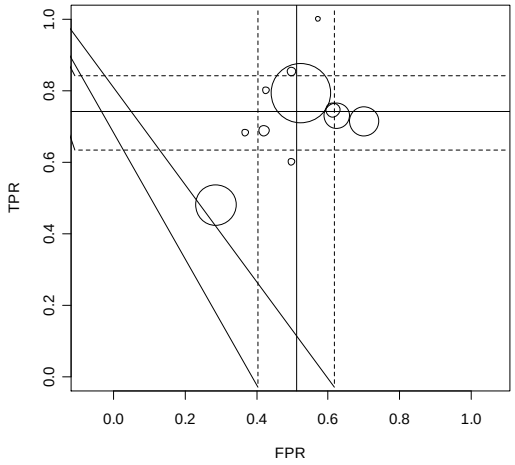
iv) Bivariate meta-analysis models. This models sensitivity and specificity and their correlation within a single model(Reitsma et al., 2005, Rutter and Gatsonis, 2001). The output from this model can be viewed as either a random effects asymmetric sROC curve(Rutter and Gatsonis, 2001) or a joint confidence region around a specific sensitivity and specificity(Harbord et al., 2007) (Figure 1 panel (iv)).

v) Synthesis of test performance at multiple thresholds from the individual studies(Dukic and Gastonis, 2003). Although rarely done in practice, it is possible to extend the Bivariate model to include multiple data points from the primary studies relating to multiple test thresholds to produce an overall sROC curve (Figure 1 panel (v)).

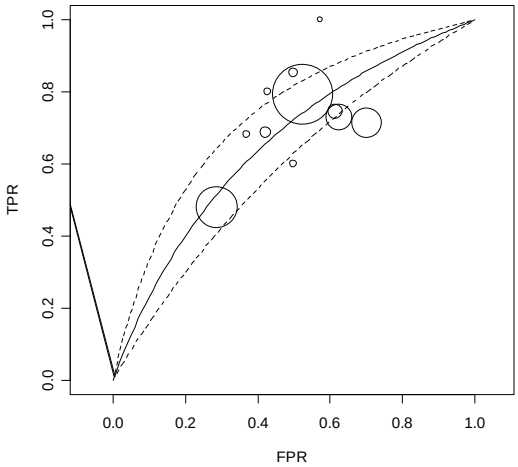
As the format of the results produced by each of the different meta-analysis models differs considerably (Figure 1), the challenge when evaluating the cost-effectiveness of diagnostic tests is how best to synthesise the available evidence and then appropriately incorporate the results of this synthesis in an economic decision model.

In this paper we investigate how evidence on test accuracy is used to inform decision models developed to evaluate the cost-effectiveness of diagnostic tests. In particular, we focus on diagnostic tests evaluated as part of the NHS Research and Development Health Technology Assessment (HTA) programme since 1997 and investigate how the evidence on diagnostic test accuracy identified as part of the systematic review is used to inform the diagnostic test accuracy parameter(s) of the economic decision model. Where evidence synthesis methods have been applied to combine test accuracy data from a number of studies, the review focuses on the specific meta-analysis models adopted and how these pooled results are used in the economic evaluation, if at all.

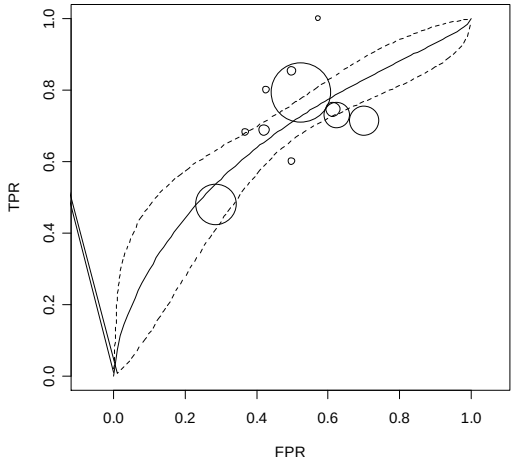
i) Meta-analyses sensitivity and specificity separately



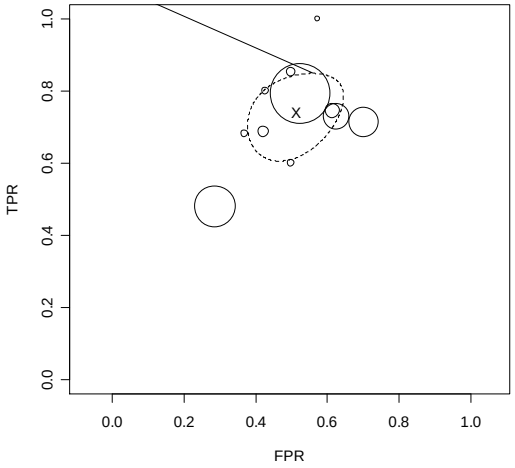
ii) Meta-analysis of diagnostic odds ratios



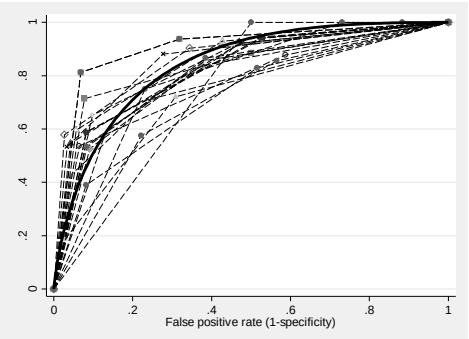
iii) Regression modelling approach



iv) Bivariate meta-analysis model



v) Meta-analysis of multiple ROC curves



TPR = True Positive Rate (Sensitivity)

FPR = False Positive Rate (1-Specificity)

Figure 1: Illustrates the different formats of the analysis output obtained when the five alternative approaches to meta-analysis of diagnostic test data are applied

2 Methods

All NHS Research & Development Health Technology Assessment (HTA) Programme reports listed on their website (http://www.ncchta.org/project/htapubs.asp) as published between 1997 and May 2009 inclusively were reviewed by one of the authors (NN) with the aim of identifying reports that evaluated the performance of diagnostic tests. First the HTA reports were categorised, based on their title, as: (i) Methodology, (ii) Treatments alone, or (iii) Testing. Where classification was unclear from the title, abstracts then executive summary and introduction were reviewed.

The second step was to sub-divide those HTA reports classified as Testing into one of the following subgroups: i) Diagnosis, ii) Screening, iii) Prognosis and iv) Monitoring. Occasionally, a report could be classified into more than one subgroup. If a report contained diagnosis and prognosis, screening or monitoring then the report was classified as diagnosis. For all other combinations the report was classified according to its main objective established by reading the main text of the report. Where the purpose(s) of the testing was unclear, categorisation was established via consensus forming discussions with two further authors (NJC & AJS).

All reports evaluating diagnostic tests were reviewed to identify whether an economic decision model had been developed as part of the HTA. Those reports where economic models had been developed were examined further to establish whether they contained meta-analyses of diagnostic accuracy data in the clinical review section of the report. Those that had defined our sample of interest, and these reports were scrutinised further. Specifically, data were extracted on:

i) All meta-analysis methods used in the clinical review.

ii) Whether any of the meta-analysis methods recorded in i) were used to derive estimates of test performance for the economic model. If yes, which method used. If no, the alternative method used to estimate diagnostic test accuracy parameters specifically for the economic model.

iii) Whether the economic model had considered pathways involving multiple test combinations, and if so, how test performance had been estimated for the combinations of tests.

3 Results

Figure 2 shows our classifications of the 474 HTA reports published between 1997 and May 2009 inclusively. 110 out of the 474 reports (23%) were classified as 'Testing' with 44 (40%) of these focusing on 'Diagnosis'. Thirty-three out of the 44 'Diagnosis' reports (75%) included an economic evaluation. Of these 33, 14 (42%) included meta-analysis of diagnostic test accuracy in the clinical review section of the report and these 14 reports define our sample of interest (A numbered reference list (S1-14) for this sample is provided in the appendix).

In Table 1 the 14 reports that performed meta-analysis as part of the clinical review are listed chronologically together with the meta-analysis method(s) used (denoted by the letter R in the table). The methods are listed broadly in order of complexity and it can be observed that most reports used more than one meta-analysis method. All of the reports except one (S9), performed meta-analyses on specificity and sensitivity separately thus assuming the two measures to be independent. One of these reports used individual participant data in their meta-analysis rather than summary data (S13)). Two reviews adopted a strategy based on heterogeneity; that is, where evidence of heterogeneity existed the Littenberg and Moses regression approach was adopted otherwise independent pooled estimates of sensitivity and specificity were obtained. (S1, S6) The most sophisticated methods of bivariate and hierarchical summary receiver operating characteristic curve were only applied by 2 of the reviews. (S9, S10) Five of the reports considered study-level covariates in their analyses (S4, S5, S9, S14, S15).

Table 1 also highlights which meta-analysis method (if any) is used to inform the test accuracy parameters in the economic decision model (denoted by the letter M in the table). Where the letters R and M appear in the same cell of the table, this indicates that one of the meta-analysis approaches used in the clinical review was also used to inform the economic decision model. Where the letter M appears in a cell on its own, this indicates that a different meta-analysis method was used specifically to inform the decision model.

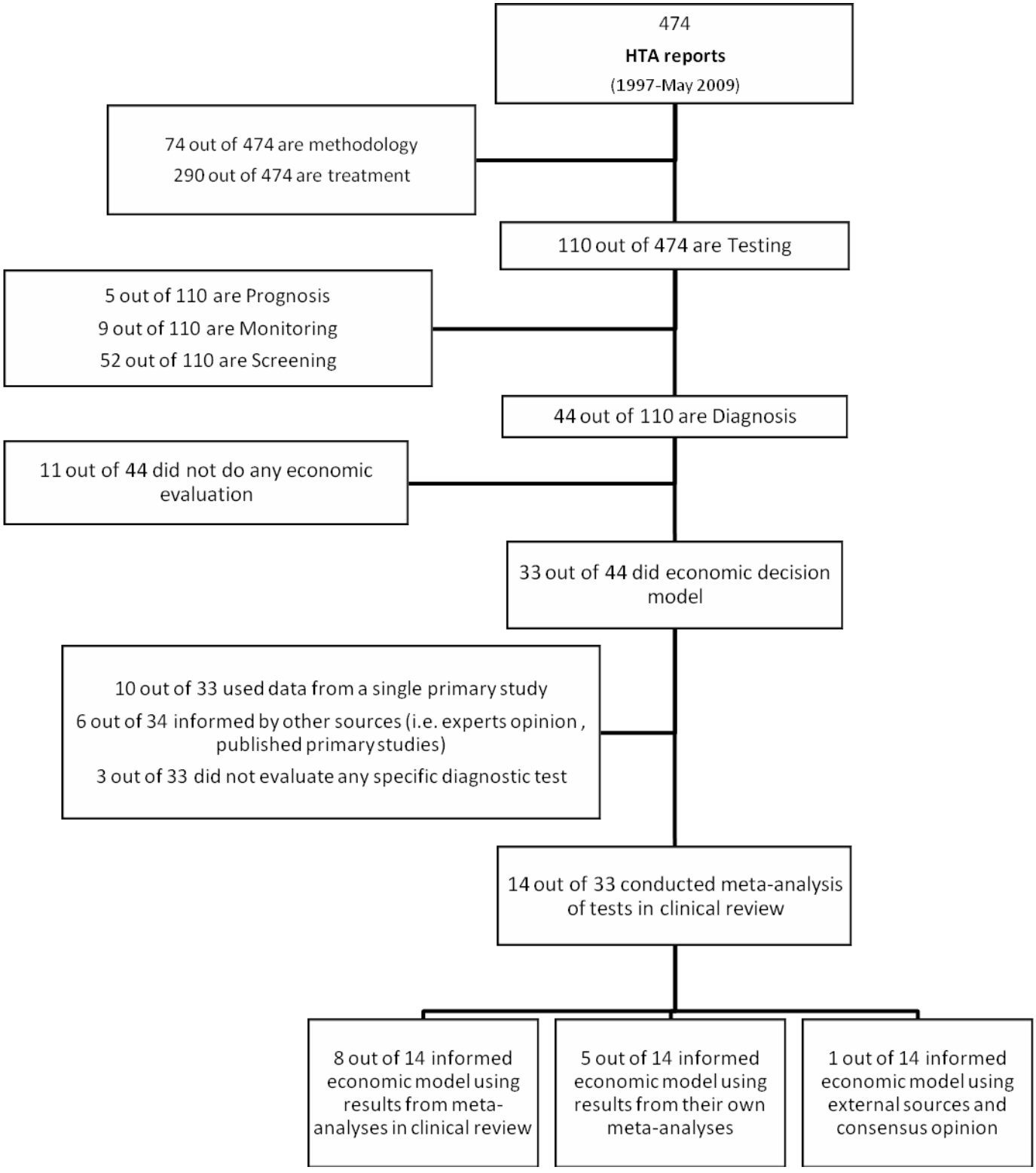


Figure 2: Flowchart of excluded and included studies

Eight out of 14 reports (57%) used pooled estimates of sensitivity and specificity obtained from a meta-analysis performed in the clinical review as inputs into the decision model, 5 (36%) used study data identified by the clinical review but performed their own meta-analyses (3 reports did meta-analyses on sensitivity and specificity separately, 1 report did a bivariate meta-analysis model, and 1 report obtained negative predictive values and ratio of test positive to test negative), and 1 report used sources external to the clinical review plus consensus opinion. Overall, the majority of reports (10 out of 14 (71%)) used pooled estimates of sensitivity and specificity obtained from the simplest meta-analysis method, that assumes the two measures are independent of one another, as inputs into the economic decision model. Only two economic decision models used estimates of sensitivity and specificity from meta-analyses that allowed for the correlation between the two quantities (i.e. a bivariate model). None of the models used a meta-analysis method that derives an sROC (i.e Diagnistic odds ratio, Littenberg and Moses regression method, HsROC curve). Ten out of the 14 models reviewed (77%) incorporated the uncertainty associated with pooled estimates to perform a probabilistic cost-effectiveness evaluation.

Evaluation of a combination of diagnostic tests

Six out of the 14 (43%) reports listed in Table 1 considered a combination of diagnostic tests in the economic decision modelling. Three of these (S5, S8, S13) assumed the tests to perform independently of one another and thus input the pooled estimates of sensitivity and specificity obtained for each test direct from the meta-analyses. Two reports (S3, S14) assumed the second test to have 100% sensitivity and 100% specificity (i.e. a perfect test). The remaining report (S12) provided no details about how the combination of tests was evaluated but we presume that independence was assumed.

	HTA Report Volume/Number													
	2002	**2004**			**2006**					**2007**		**2008**		**2009**
Meta-analytic methods used to evaluate diagnostic accuracy	S2	S7	S6	S11	S8	S5	S12	S13	S14	S3[1]	S1	S9	S10	S4[2]
Independent sensitivity and specificity	M	R, M	R, M[3]	R, M	R, M	R, M	M[4]	R, M	R	R	M		R, M	R
Likelihood ratio		R	R			R	R		R	R			R	
Diagnostic odds ratio					R			R		R	R		R	
Littenberg and Moses regression approach	R						R		R					
Littenberg and Moses if heterogeneity, if not independent			R								R			
Bivariate model									M[5]			R, M[6]		
Hierarchical Summary Receiver Operating Characteristic curve													R	

[1]Used data from systematic review to obtain negative predictive values (number of true negatives divided by total number of negatives) and ratio of test positives to test negatives
[2]Used data from external sources and consensus opinion
[3]Used median sensitivity and specificity
[4]Expert opinion used where no studies identified in the systematic review
[5]Performed a series of regression analyses to establish the relationship between sensitivity and specificity
[6]Unclear how the Bivariate data is dealt with in the probabilistic decision model

Table 1: Meta-analysis methods applied in the systematic review of diagnostic test accuracy (R) and results used as input parameter in the economic decision model (M).

4 Discussion

The focus of this review has been to assess how evidence on test accuracy is synthesised and used to inform economic decision models evaluating diagnostic pathways. The proportion of HTA reports reviewed that considered the effectiveness and cost-effectiveness of diagnostic tests is relatively small. The number of reports that developed an economic decision model and used a meta-analytic approach to summarize test performance evidence was even smaller. It would appear that, currently, there is little guidance relating to this topic in the published literature.

Many of the reports used a range of different meta-analysis methods to synthesise the test performance data. This in itself can be problematic since virtually all the methods make different assumptions, and therefore, theoretically cannot simultaneously be appropriate for a given dataset. Thus this can lead to problems of interpretation regarding the most appropriate summary of test performance. Despite the multiple methods used in many of the clinical reviews, the majority of the reports applied the simple meta-analytic approach of assuming sensitivity and specificity to be independent for informing the decision model. This is concerning because it has been established that when this model is used inappropriately (i.e. the primary studies evaluate tests at different thresholds) the resulting point estimate underestimates true test performance(Deeks, 2001) (i.e. it lies below the sROC curve that would be produced by an analysis that takes threshold into account). Additionally, if a probabilistic modelling approach is used, this approach will estimate the uncertainty incorrectly.

Although half the reports calculated pooled likelihood ratios for test performance, none went on to use these estimates to inform the decision model. This is understandable since it is not as straightforward to use likelihood ratios compared to estimates of sensitivity and specificity to estimate the number of true positives, true negatives, false positives and false negatives required by the typical parameterisation of decision models evaluating diagnostic tests. Similarly, although methods that estimate an sROC curve (i.e. diagnostic odds ratios & the regression method of Littenberg and Moses) were conducted quite frequently, the output from these analyses was never used to inform the decision model. Again, this may

well be because it is not obvious how to parameterise output in the form of an sROC curve in the decision model. Indeed, one report (S13) stated that meta-analyses were performed on sensitivity and specificity separately, rather than calculating an sROC curve, to obtain the parameters needed for the economic decision model. An sROC curve describes how test performance varies with changing test threshold, therefore it would be possible to consider the cost effectiveness of a diagnostic strategy as a function of test threshold. This could be achieved most simply by running a series of decision models using estimates of sensitivity and specificity for the test(s) at different locations on the sROC curve. In this way, it is possible to identify the optimum threshold – in terms of cost-effectiveness – to use a test at (although it should be acknowledge that, in practice, this may or may not be achievable). To our knowledge, this approach has only been attempted once in the published literature(Sutton et al., 2008).

A bivariate model, which accounts for the correlation between sensitivity and specificity, was used in two of the reports. There would appear to be growing consensus in the statistical literature that this is the most appropriate model for meta-analysing test performance data(Harbord et al., 2007) (Arends et al., 2008). Therefore, this finding could be interpreted as disappointing. However, it is important to remember that this approach to meta-analysis of diagnostic test data was only described in the literature in 2005(Reitsma et al., 2005) with custom software appearing even more recently (e.g. Stata macros). Even once the parameter estimates for the bivariate model have been obtained, for a probabilistic decision model, it will be necessary to specify a multivariate normal distribution or a re-parameterisation or approximation to it that is non-trivial (i.e. one of the papers stated using Cholesky Decomposition for this (S14)). Alternatively, it is possible to use a one-stage comprehensive approach to the decision modelling where the meta-analyses are carried out simultaneously in the same computer program that evaluates the decision model. This has been described elsewhere(Sutton et al., 2008) using the WinBUGS software(Spiegelhalter et al., 2003) which implements MCMC simulation methods, and perhaps provides the most elegant approach available to date.

Despite the above, it is the authors' belief that the bivariate approach should not be used uncritically for the following reason. The method estimates a 95% confidence region for the

average sensitivities and specificities observed in the primary studies. Therefore, it is implicit that all the studies are representative of how the test will be used in routine practice. If for example, particular studies use test thresholds which are not representative of routine practice / a particular threshold being considered, then such an analysis would seem inappropriate. In such cases, exploring cost-effectiveness as a function of an sROC curve, or at one particular point on the curve, would seem more appropriate (although, study level data relating to test threshold is not routinely included in the meta-analysis models and therefore it is not obvious which point on an sROC curve relates to a particular threshold). Given this, further research is required to establish the optimal approach in different situations and this is ongoing.

To add further confusion to this already complex area, it was recently established that the bivariate model and the hierarchical sROC approach are actually re-parameterisations of the same model(Harbord et al., 2007) although the two parameterisations lead naturally to different outputs (i.e. a confidence region in ROC space and an sROC curve respectively). Thus, owing to this re-parameterisation, it is possible to obtain an sROC curve from the bivariate analysis and therefore the discussion relating to sROC curves above is also pertinent for this model leading to even more possibilities of how diagnostic test data may be used to inform decision models.

Six of the models reviewed considered diagnostic pathways using multiple tests in combination. The use of combinations of tests is common in clinical practice, e.g. a cheap or non-invasive test may initially be used which has poor specificity and those diagnosed as diseased may go on to receive a more expensive / more invasive test with superior test performance. We are concerned that estimation of accuracy of test combinations was dealt with too simplistically in these reviews (i.e. assuming tests to be independent or the second test to be perfect). Crucially, this is perhaps a limitation of the available data as much as the modelling per se as many primary studies estimating test performance only consider a single test so results of tests conditional on the results of other tests are rarely available. We are concerned that if the strong assumption of test independence is violated, this could lead to misleading conclusions. Further work is needed to establish ways of estimating such correlations. Even if they are estimated with considerable uncertainty, including them in the

modelling allows the possibility of using value of information methods(Claxton, 1999, Ades et al., 2004) to demonstrate the importance of conducting primary studies to estimate them more accurately.

In conclusion, meta-analytic methods for diagnostic test accuracy data have developed rapidly in recent years. Decision modellers need to be aware of the recent developments in this area and appreciate the limitations of simplistic approaches used commonly in the past.

REFERENCES

ADES, A. E., LU, G. & CLAXTON, K. (2004) Expected value of sample information calculations in medical decision modelling. *Medical Decision Making,* 24**,** 207-227.

ARENDS, L. R., HAMZA, T. H., VAN HOUWELINGEN, H. C., HEIJENBROK-KAL, M. H., HUNINK, M. G. M. & STIJNEN, T. (2008) Bivaraite random effects meta-analysis of ROC curves. *Medical Decision Making,* 28**,** 621-638.

CLAXTON, K. (1999) The irrelevance of inference: A decision-making approach to the stochastic evaluation of health care technologies. *Journal of Health Economics,* 18**,** 341-364.

DEEKS, J. (2001) Systematic reviews in health care: Systematic reviews in evaluations of diagnostic and screening tests. *BMJ,* 323**,** 157-162.

DEEKS, J. J. & ALTMAN, D. G. (2001) Effect measures for meta-analysis of trials with binary outcomes. IN EGGER, M., DAVEY SMITH, G. & ALTMAN, D. G. (Eds.) *Systematic reviews in health care: Meta-analysis in context.* London, BMJ Publishing Group.

DUKIC, V. & GASTONIS, C. (2003) Meta-analysis of Diagnostic Test Accuracy Assessment Studies with Varying Number of Thresholds. *Biometrics,* 59**,** 936-46.

HARBORD, R. M., DEEKS, J. J., EGGER, M., WHITING, P. & STERNE, J. A. C. (2007) A unification of models for meta-analysis of diagnostic accuracy studies. *Biostatistics,* 8**,** 239-251.

LITTENBERG, B. & MOSES, L. E. (1993) Estimating diagnostic accuracy from multiple conflicting reports: a new meta-analytic method. *Medical Decision Making,* 13**,** 313-21.

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

REITSMA, J. B., GLAS, A. S., RUTJES, A. W. S., SCHOLTEN, R. J. P. M., BOSSUYT, P. M. M. & ZWINDERMAN, A. H. (2005) Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic review. . *Journal of Clinical Epidemiology,* 58**,** 982-90.

RUTTER, C. M. & GATSONIS, C. A. (2001) A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Statistics in Medicine* 2865-84.

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

SUTTON, A. J., COOPER, N. J., GOODACRE, S. & STEVENSON, M. (2008) Integration of meta-analysis and economic decision modeling for evaluating diagnostic tests. *Medical Decision Making,* 28**,** 650-667.

APPENDIX: HTA reports included in the review

S1 Abubakar, I., L. Irvine, C. F. Aldus, G. M. Wyatt, R. Fordham, S. Schelenz, L. Shepstone, A. Howe, M. Peck, and P. R. Hunter. 2007. A systematic review of the clinical, public health and cost-effectiveness of rapid diagnostic tests for the detection and identification of bacterial intestinal pathogens in faeces and food. *Health Technology Assessment* 11 (36):iii-110.

S2 Berry, E., S. Kelly, M. E. Westwood, L. M. Davies, M. J. Gough, J. M. Bamford, J. F. M. Meaney, C. M. Airey, J. Cullingworth, M. Barbieri, A. Jackson, and M. A. Smith. 2002. The cost-effectiveness of magnetic resonance angiography for carotid artery stenosis and peripheral vascular disease: A systematic review. *Health Technology Assessment* 6 (7).

S3 Collins, R., G. Cranny, J. Burch, R. Aguiar-Ibanez, D. Craig, K. Wright, E. Berry, M. Gough, J. Kleijnen, and M. Westwood. 2007. A systematic review of duplex ultrasound, magnetic resonance angiography and computed tomography angiography for the diagnosis and assessment of symptomatic, lower limb peripheral arterial disease. *Health Technology Assessment* 11 (20):iii-120.

S4 Fortnum, H., C. O'Neill, R. Taylor, R. Lenthall, T. Nikolopoulos, G. Lightfoot, G. O'Donoghue, S. Mason, D. Baguley, H. Jones, and C. Mulvaney. 2009. The role of magnetic resonance imaging in the identification of suspected acoustic neuroma: A systematic review of clinical and cost-effectiveness and natural history. *Health Technology Assessment* 13 (18):iii-106.

S5 Goodacre, S., F. Sampson, M. Stevenson, A. Wailoo, A. Sutton, S. Thomas, T. Locker, and A. Ryan. 2006. Measurement of the clinical and cost-effectiveness of non-invasive diagnostic testing strategies for deep vein thrombosis. *Health Technology Assessment* 10 (15):iii-99.

S6 Kaltenthaler, E., Y. B. Vergel, J. Chilcott, S. Thomas, T. Blakeborough, S. J. Walters, and H. Bouchier. 2004. A systematic review and economic evaluation of magnetic resonance cholangiopancreatography compared with diagnostic endoscopic retrograde cholangiopancreatography. *Health Technology Assessment* 8 (10):iii, 1-89.

S7 Mant, J., R. J. McManus, R. A. L. Oakes, B. C. Delaney, P. M. Barton, J. J. Deeks, L. Hammersley, R. C. Davies, M. K. Davies, and F. D. R. Hobbs. 2004. Systematic review and modelling of the investigation of acute and chronic chest pain presenting in primary care. *Health Technology Assessment* 8 (2):iii-78.

S8 Martin, J. L., K. S. Williams, K. R. Abrams, D. A. Turner, A. J. Sutton, C. Chapple, R. P. Assassa, C. Shaw, and F. Cheater. 2006. Systematic review and evaluation of methods of assessing urinary incontinence. *Health Technology Assessment* 10 (6):iii-87.

S9 Meads, C. A., J. S. Cnossen, S. Meher, A. Juarez-Garcia, G. Ter Riet, L. Duley, T. E. Roberts, B. W. Mol, J. A. Van der Post, M. M. Leeflang, P. M. Barton, C. J. Hyde, J. K. Gupta, and K. S. Khan. 2008. Methods of prediction and prevention of pre-eclampsia: Systematic reviews of accuracy and effectiveness literature with economic modelling. *Health Technology Assessment* 12 (6):1-249.

S10 Mowatt, G., E. Cummins, N. Waugh, S. Walker, J. Cook, X. Jia, G. S. Hillis, and C. Fraser. 2008. Systematic review of the clinical effectiveness and cost-effectiveness of 64-slice or higher computed tomography angiography as an alternative to invasive coronary angiography in the investigation of coronary artery disease. *Health Technology Assessment* 12 (17):iii-143.

S11 Mowatt, G., L. Vale, M. Brazzelli, R. Hernandez, A. Murray, N. Scott, C. Fraser, L. McKenzie, H. Gemmell, G. Hillis, and M. Metcalfe. 2004. Systematic review of the effectiveness and cost-effectiveness, and economic evaluation, of myocardial perfusion scintigraphy for the diagnosis and management of angina and myocardial infarction. *Health Technology Assessment* 8 (30):iii-89.

S12 Rodgers, M., J. Nixon, S. Hempel, T. Aho, J. Kelly, D. Neal, S. Duffy, G. Ritchie, J. Kleijnen, and M. Westwood. 2006. Diagnostic tests and algorithms used in the investigation of haematuria: Systematic reviews and economic evaluation. *Health Technology Assessment* 10 (18).

S13 Wardlaw, J. M., F. M. Chappell, M. Stevenson, E. De Nigris, S. Thomas, J. Gillard, E. Berry, G. Young, P. Rothwell, G. Roditi, M. Gough, A. Brennan, J. Bamford, and J. Best. 2006. Accurate, practical and cost-effective assessment of carotid stenosis in the UK. *Health Technology Assessment* 10 (30):iii-128.

S14 Whiting, P., M. Westwood, L. Bojke, S. Palmer, G. Richardson, J. Cooper, I. Watt, J. Glanville, M. Sculpher, and J. Kleijnen. 2006. Clinical effectiveness and cost-effectiveness of tests for the diagnosis and investigation of urinary tract infection in children: a systematic review and economic model. *Health Technology Assessment* 10 (36):iii-iv, xi-xiii, 1-154.