Medical Decision Making, 2002, in press

Markov Chain Monte Carlo estimation of a multi-parameter decision model: consistency of evidence, and the accurate assessment of uncertainty.

AE Ades, PhD

Medical Research Council, Health Services Research Collaboration, University of Bristol, England

S Cliffe, MPH

Communicable Disease Surveillance Centre, Public Health Laboratory Service, Colindale, England
and
Department of Epidemiology and Biostatistics, Institute of Child Health, University College London, London, England

Running title:
MCMC estimation of a multi-parameter model: consistency and uncertainty

Correspondence to: Dr A E Ades, MRC Health Services Research Collaboration, Department of Social Medicine, University of Bristol, Canynge Hall, Whiteladies Road Bristol BS8 2PR. E-mail: t-ades@bristol.ac.uk

Ades-mdm9-ref.doc
March 26, 2002

Abstract

Decision models are usually populated one parameter one at a time, with one item of information informing each parameter. Often, however, data may not be available on the parameters themselves but on several functions of parameters, and there may be more items of information than there are parameters to be estimated. We show how in these circumstances all the model parameters can be estimated simultaneously, using Bayesian Markov Chain Monte Carlo methods. Consistency of the information and/or the adequacy of the model can also be assessed within this framework. Statistical evidence synthesis using all available data should result in more precise estimates of parameters and functions of parameters, and is compatible with the emphasis currently placed on systematic use of evidence. To illustrate this, WinBUGS software is used to estimate a simple 9-parameter model of the epidemiology of HIV in women attending prenatal clinics, using information on 12 functions of parameters, and to thereby compute the expected net benefit of two alternative prenatal testing strategies, universal testing and targeted testing of high risk groups. We demonstrate improved precision of estimates, and lower estimates of the expected value of perfect information, resulting from the use of all available data.

Keywords

Evidence synthesis, decision analysis, Markov Chain Monte Carlo, Bayesian methods, incremental net benefit, Expected Value of Perfect Information, epidemiology, HIV, screening.

Standard texts and tutorial articles on decision analysis tend to assume that parameter values and their ranges are taken individually from appropriate literature [1 2]. Typically, there is uncertainty concerning the parameter values because each is based on a finite sample. If this uncertainty can be expressed in the form of statistical distributions, Monte Carlo (MC) simulation is an appropriate way of computing the expected values of particular functions of parameters, such as the incremental net benefit of one strategy over another. MC simulation, using standard spreadsheet software, also allows one to correctly propagate the uncertainties in the input parameters to generate a distribution around functions of parameters, and this 'probabilistic sensitivity analysis' [3 4 5] has become a firmly accepted tool of decision analysis [2 5 6].

In this paper we look at situations where data may not be available on the parameters themselves, but on complex functions of several parameters. In addition, information may be available on more functions of the parameters than there are parameters in the model. It is hard to say how often this situation arises, but the habit of populating decision models one parameter at a time, and unfamiliarity with methods that can take advantage of multiple evidence types, has perhaps deterred decision analysts from seeking and using additional sources of data that are available. No one seems to have noticed what a marvellous coincidence it is that the literature always seems to provide just the right number of items of data to populate a model, never more !

The proposals below for the statistical combination of evidence are in the same spirit as the Confidence Profile Method (CPM) of DM Eddy and colleagues [7 8 9 10 11] , but we attempt to take their very powerful ideas further in a number of ways. First, we show how Markov Chain Monte Carlo methods implemented in the freely available WinBUGs software [12] provide a relatively simple alternative estimation framework. Secondly, we raise the issue of evidence consistency, which was not emphasised in CPM literature on evidence synthesis. Thirdly, we show how conventional methods for populating decision trees, one parameter at a time, lead to incomplete and hence arbitrary use of evidence. In a decision context this results in a biased assessment, and usually an over-estimate, of the true uncertainty in a decision.

We illustrate our approach with a simple model of the epidemiology of HIV infection in pregnant women, set in the context of a decision about whether prenatal testing for HIV in the London metropolitan area of the United Kingdom should be universal or targeted at high risk groups [13 14 15]. This is not, at present, a particularly realistic decision problem: the decision to screen universally throughout England has now been taken, although it remains to be fully implemented. Further, targeted testing, even in areas of extremely low prevalence where universal testing may not be cost-effective, is considered politically difficult in Europe and North America, and possibly discriminatory [16 17]. Although the analysis is therefore illustrative in nature, simplifications intended to avoid unhelpful detail are made explicit, and the analysis is essentially what we would propose in a definitive application.

Methods

HIV EPIDEMIOLOGY MODEL AND SOURCES OF DATA

The decision problem is to choose between targeted and universal prenatal testing for HIV. The decision tree in Figure 1 incorporates what is, in effect, an 8- parameter model of the epidemiology of HIV infection in pregnant women: *a* is the proportion of the prenatal population born in sub-Saharan Africa (SSA), *b* is the proportion who are previous or current injecting drug users (IDU). The remaining proportion *(1-a-b*) are considered low risk: these would only be tested in a universal programme. Parameters *c,d,* and *e* are the HIV prevalence in each of these three population groups, and *f, g*, and *h* are the proportions of infected women who have already been diagnosed as HIV infected prior to prenatal attendance.

Table 1 presents the 12 data items that could be used to inform this model. Note that there is no direct data on parameters *e,f*, or *h*. Before looking at the decision problem in terms of the incremental costs and effects of universal compared to targeted testing, we begin by describing the data sources and their relation to model parameters, from the perspective of an epidemiologist interested in estimating the model.

Data item 1 provides information on parameter *a*, the proportion of women who were born in Sub-Saharan Africa. Data on mother's country of birth is collected at civil registration of every birth and compiled by the Office for National Statistics [18].

Item 2 provides information on parameter *b*, the proportion of IDUs. This is based on the 1990 National Survey of Sexual Attitudes and Lifestyles [19]. The 12/882 (0.136%) used here represents an adjustment of the original national data for the relative extent of reported female injecting drug use in London[14], and for additional variance in the estimate due to survey clustering (A.Copas, personal communication).

Anonymous newborn seroprevalence surveys with linked maternal country of birth data in North London provide data on parameter *c*, HIV prevalence in women born in Sub-Saharan Africa (item 3) [20]. (Antibody in the newborn is a highly reliable indicator of maternal infection). Other anonymised surveys (item 4) [21] give information on parameter *d*, HIV prevalence in IDUs. Although there is no direct information on HIV prevalence in the low-risk, there is data on women *not* born in Sub-Saharan Africa. Thus, data item 5 estimates a weighted average of the HIV prevalence parameters *c* and *d*, with the population sizes, *b* and *(1-a-b)* respectively, as the weights.

Item 6, also from anonymised surveys[21], provides information on overall seroprevalence in all London in 1999, a function of the parameters *a,b,c,d,* and *e* . (Note that this data source does not overlap with items 3 and 5, which relate to data collected in the years 1997-98. Setting aside the possibility of auto-correlation across time-periods, this is a somewhat artificial way of ensuring approximate independence of data items, in order to avoid unhelpful complication).

Items 7 and 8 arise from data that is collected on cases of HIV in pregnancy that have been diagnosed prior to initial prenatal attendance [22 23]. This survey can be used to provide information on the proportion of diagnosed infection that can be attributed to each of the three population groups. To maintain independence the information is expressed in the form of two proportions, diagnosed SSA women as a proportion of all diagnosed women, and diagnosed IDUs as a proportion of all diagnosed non-sub-Saharan African women. The same survey provides the total number of already diagnosed women, which is the numerator for data item 9, while the denominator, the total number of infected women, comes from the anonymised neonatal surveys [24]. Data item 10[21] provides a direct estimate of parameter *g*, the proportion of infected IDUs that are already diagnosed.

In the period 1997-8, anonymous samples from North London that were HIV-1 seropositive were serotyped [25], and 15/118 typed samples from infants born to women from sub-Saharan Africa were sub-type B. Data item 11 informs us about an ancilliary parameter, *w*, which is not part of the epidemiological model, but which allows us to incorporate serotype data on non-SSA mothers (item 12). The reasoning here is as follows. HIV infection in IDUs is invariably subtype B. A fairly well-supported assumption at this early stage in the epidemic is that the subtype profile in the low-risk group is the same as the SSA subtype profile [26]. Thus the proportion of subtype B in the non-SSA women (data item 12) is a weighted average of the proportion of subtype B in SSA women (*w*), and the proportion of subtype B in IDUs, with the weights being the proportions of infection attributable to IDU and low-risk groups. In the absence of direct data on the parameter *e*, HIV prevalence in the low risk, we introduce these two final data points, along with the ninth parameter *w*, in order to provide further information that bears on the distribution of infection in these two groups.

THE DECISION PROBLEM

Our characterisation of the decision problem is kept very simple. It is assumed that the test is 100% sensitive and specific, and that the uptake of testing in the high risk groups is the same in both strategies. Under these circumstances, it can be shown that the incremental net benefit of universal compared to targeted testing depends on the prevalence of undiagnosed infection in the low risk group, the additional group that a universal strategy tests and a targeted strategy does not [14 27].

Specifically, for a prenatal population of 105,000 per year in London, the additional number of tests carried out in a universal programme compared to targeted is: *105000 (1-a-b) (1-eh)*. The additional number of infected women diagnosed is: *105000 (1-a-b) e(1-h)*. With T as the unit cost of an HIV test and B as the net benefit [28 29] of a maternal diagnosis, the expected incremental net benefit of universal over targeted testing per year is therefore:

$$E[INB_U] = E[\ 105000\ (1-a-b)\ \{Be(1-h) - T(1-eh)\}\ ] \qquad (1)$$

INB_U is a random variable, due to the stochastic uncertainty in the parameters *a,b,e,h,* and *B.* The optimum policy is Universal if and only if $E[INB_U] > 0$. We therefore take an expectation over the expression, bearing in mind that the mean of a non-linear function of random variables is not equivalent to the function applied to their expected values.

In the illustration that follows, *T* is taken to be fixed at £3. For *B* a distribution of values is based on earlier work which took account of: the costs and effectiveness of treatment to prevent transmission to the foetus and newborn, the life-time costs of paediatric infection, the life-years gained by averting it, the life-years gained by the mother through earlier diagnosis and the additional costs incurred. Based on published models of each of these processes, and assuming a £10,000 value per additional life-year gained [30], it was concluded that the net benefit of a maternal diagnosis was approximately £50,000 with a range of £12,000 to £60,000 [13]. The extreme leftwards skew is a result of the diminishing return that is achieved by lowering the mother-to-child transmission rate still further. For present purposes we interpret these results as the mean and 95% confidence limits of a probability distribution, which can be represented by a transformed gamma distribution, truncated at an upper limit of 2:

$$B \sim 600012 - 54296\ \{gamma\ (0.56,3)\ |\ (0,2)\} \qquad (2)$$

Uncertainty in the incremental net benefit of universal versus targeted testing is expressed by the spread of the distribution of INB_U. This uncertainty results in uncertainty about which decision to adopt. If we had perfect information on INB_U the decision would be optimal and we would maximise net benefit. Uncertainty in INB_U therefore has a 'cost', arising from the probability that a decision based on its expected value may be the wrong decision. This expected loss due to uncertainty is also known as the Expected Value of Perfect Information (EVPI)[31 32], because it is the amount that a decision maker should be willing to pay in order to eliminate uncertainty in a decision.

EVPI can be calculated from the distribution of INB by integrating over the product of the loss and the probability of the loss. Assuming that universal testing has a higher expected net benefit than targeted (i.e. $NB_U - NB_T > 0$), then universal testing is the

optimal decision given current information. EVPI is the difference between the expected gain given a decision based on perfect information, $E[\max(NB_U, NB_T)]$, and the expected gain given a decision based on current information, $E[NB_U]$. Using $NB_U - NB_T = INB_U$, and assuming a 10 year horizon for the decision, and discounting at 5% per year, then:

$$EVPI_U = E[7.7127 \quad Max(-INB_U, 0)] \qquad (3)$$

ESTIMATION OF THE MODEL

Bayesian analysis

Markov Chain Monte Carlo (MCMC) simulation is a framework that is useful in both Bayesian statistical inference and decision analysis. The theoretical background and practical examples are set out in recent texts [33 34 35]. Here we employ WinBUGs software, version 1.3.[12] The BUGs website (Bayesian inference Using Gibbs Sampling), *www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml*, gives free access to the software, manuals, worked examples, tutorial papers, lectures, and further references and contacts.

The aim of Bayesian inference is to provide information on the posterior distribution of model parameters given the data. In a multi-parameter model, direct computation of the posterior distribution would require a high dimensional integration. Instead, the MCMC approach is to repeatedly draw samples from the posterior distribution, allowing the analyst to estimate its mean, standard deviation and percentiles. In Gibbs sampling [36], an MCMC algorithm implemented in WinBUGS, samples are drawn from the conditional distribution of each parameter given the current values of the other parameters and the data. It can be shown that, given some initial values, the resulting distributions converge to a stationary distribution which correctly reflects the required posterior distributions [37]. The first few thousands of 'burn in' simulations, while the distributions are not yet stationary, are discarded. In the examples presented, tests for convergence [38] incorporated in WinBUGS suggested that adequate convergence had been achieved within 3000 iterations. In results reported below the first 10,000 were discarded, and results are based on the subsequent 100,000 samples. This took about 115 seconds on a PC with a Pentium III 700MHz processor.

The BUGS system computes the conditional distributions, manages the sampling process, accumulates the samples from the posterior distributions of the parameters and of any function of the parameters, and produces statistical summaries of them including the mean, standard deviation and centiles. All the user need do is to specify the relationships between the data and the model parameters, exactly as expressed in Table 1, provide prior probability distributions for each of the parameters, and specify the likelihood for the data, given values of the parameters. All the 9 basic parameters are assigned beta distributions, which is a natural choice for probability parameters. An additional constraint is imposed to ensure that $(1-a-b) > 0$. All parameters are given minimally informative, uniform, beta(1,1) priors indicating a prior belief that all values are equally likely. Vague priors mean that inference will be dominated by the data rather than prior assumptions. Beta(0.5,0.5) priors would be an alternative giving virtually identical results. Each of the twelve data items are generated by binomial processes, so that the likelihood of the data given values of the parameters is a product of 12 binomial likelihood contributions. The WinBUGS code, given in full in the appendix, therefore specifies that the data numerators are drawn from binomial distributions, whose probability parameters are given by the model parameter or function of model parameters shown in Table 1, and with denominators given by the data.

Model checking and diagnostics

Besides looking at the posterior distributions of the estimates and various functions of them, we also compute certain diagnostic statistics both to assess the overall fit of the model, and to identify data points that are inconsistent with other data given the model. Bayesian global goodness of fit statistics used to compare models in this paper are: the sum of the squared standardised residuals, the posterior mean deviance, the estimated deviance at the maximum likelihood solution,[39] and the Deviance Information Criterion (DIC).[39] The latter is a measure that captures goodness of fit but penalises additional parameters. To assess the consistency of individual data points we use: the standardised residual, the deviance contribution, and the probability of getting a more extreme observation. The latter is the posterior predictive p-value, obtained by comparing the observed data with a simulated 'replicate' observation, whose distribution captures both the uncertainty in the fitted probability and the

uncertainty due to binomial sampling.[40] These are not necessarily the most sophisticated model diagnostics, but they are simple to compute and serve to illustrate the feasibility and importance of model checking in decision analysis. The Appendix describes how each of the model diagnostics, and all the results presented in tables, can be derived from the WinBUGS output.

Results

GOODNESS OF FIT AND CONSISTENCY OF EVIDENCE

Table 2 compares the mean and the 2.5 and 97.5 percentiles of the posterior distribution of the predicted probabilities, along with the point-by-point model checking information. The posterior mean probabilities for each data point are reasonably consistent with the observed data, except for 3 data points: item 2, the proportion of the prenatal population who are or were IDUs (parameter *b*); item 4, the prevalence of HIV in this group (parameter *d*); and item 12, the proportion of HIV infected women not from sub-Saharan Africa who have serotype B. In these three cases, the observed data lie towards the edge of the posterior distribution.

The three indicators of goodness of fit at the individual data-point level confirm this picture. The posterior mean of the deviance contributions, which would be expected to be unity if the data were consistent with the model, is higher at these three points. The standardised residuals for these three items are also the largest. The residuals for items 2 and 4 are both positive, suggesting that either the proportion of IDUs or the HIV prevalence in IDUs in the data, or both, is too great to be consistent with the remaining information. The probability of obtaining a more extreme observation is also at its lowest at these three points.

To obtain further insight into which data source or sources are inconsistent with the others, the model was re-estimated with each of the three deviant data points left out in turn, and results compared with the original 12-item dataset in terms of global goodness of fit statistics (Table 3). For all criteria a lower number indicates a better fit. Both the full dataset and the dataset without item 12 produce a poor fit, compared to the datasets without item 2 or without item 4. With either of these two data points removed, the remaining data items appear to be consistent and the model appears to fit

well, in the sense that the estimated deviance at the maximum likelihood solution is close the number of degrees of freedom, the posterior mean deviance is close to the number of data points, and the DIC is equal to the number of data points plus the number of parameters. However, removal of the serotyping data item 12 still leaves evidence of poor model fit as judged by these global diagnostics. All this suggests that it is the conflict between items 2 and 4 rather than the serotyping data item 12, that is the cause of the inconsistency. There is little to choose between the two datasets that appear most consistent. This is because in each case information on the missing parameter, *b* or *d*, comes from evidence on the other together with the same evidence on the product *db*.

As the choice of dataset cannot be made on the basis of global goodness-of-fit statistics, one must consider the epidemiological interpretation of the data. The information on the proportion of IDUs in the prenatal population is based on a structured representative survey, and should be unbiased. However, information on HIV prevalence in IDUs is based on current, and probably longer-term, drug users attending drug user clinics. HIV prevalence in this group is likely to be higher than among the 'ever-IDUs' among the prenatal population. For this reason we would prefer to adopt a model based on dropping data item 4 in a definitive analysis.

Full results for this 11-item dataset (Table 4) indicate predicted values are more concordant with the observed data, and the point-by-point diagnostics all suggest the remaining data are consistent with each other. Note that the modeling still gives us an estimate of parameter *d*, HIV prevalence in IDUs, although there is no longer evidence that informs this directly. The estimate is 0.5% (95% confidence interval: 0.15% - 1.2%), lower than the full 12-item dataset estimate of 1.2%, and far less than the 2.1% in the data itself.

PRECISION OF ESTIMATES

Bringing more data to bear on a decision problem should improve the precision of estimates. This can be illustrated most readily in parameters and functions on which there is direct data. In Table 5 we compare the coefficient of variation (CV) of the posterior distributions of the fitted probabilities at each data item with the CVs of the

original data represented as a beta distribution. The improvements in precision are particularly noticeable in items relating to IDUs, but estimates of HIV prevalence in women from Sub-Saharan Africa and of overall prevalence have also benefited.

The uneven nature of the changes in precision illustrates how the influence of parameters and the functions of parameters on each other depends on the model structure and on the amount of information supporting different aspects of the model. Although MCMC estimation is superficially similar to forward Monte Carlo simulation as practiced in probabilistic sensitivity analysis, the critical difference is that both the structural constraints imposed by distributional assumptions, and the excess of data items over parameters to be estimated, induce correlations between parameters. These are illustrated in Figure 2. Notice for example the strong negative correlation between items 2 and 4. Within a framework where we have information – direct or indirect – on the proportion of all HIV infection that is attributable to IDU, a scenario of high HIV prevalence implies a low number of IDUs in the population, and vice versa. Without the constraint imposed by information on the overall amount of IDU-related infection via the product *db*, this correlation cannot occur.

ESTIMATES OF LATENT PARAMETERS, INCREMENTAL NET BENFIT AND EXPECTED VALUE OF PERFECT INFORMATION

As we have seen, multi-parameter synthesis via MCMC readily generates estimates of model parameters and functions of model parameters which have not been directly observed, and further functions of them. Posterior means and confidence intervals for parameters *e,h*, along with the Incremental Net Benefit of universal over targeted tests, and the EVPI, are presented in Table 6, based on the full 12-item dataset and on the 11-item dataset with item 4 removed. The table shows that the latent parameters, INB and EVPI are not particularly sensitive to the inclusion of this data point.

The results from the 12-item and 11-item datasets contrast with those that might be obtained if one were to choose subsets of the data. Table 6 gives results for 6 of over one hundred 'minimum' subsets of the data from which INB and EVPI could be computed. (Minimum in the sense that the number of data points equals the number of parameters that need to be estimated, so that they provide just enough data to uniquely identify the parameters *a,b,e,h* that are required for INB). Among these subsets some,

like {1,2,3,4,5,7,8,9} and {1,2,3,4,6,7,8,9}, incorporate the strongest items of data and would form a reasonable basis for estimation. Others, such as {2,3,4,5,8,9,10,11,12}, leave these items out, and would not be considered an intelligent use of available data. Although the expected INB_U is positive in all these datasets, so that policy decision in favour of Universal would not be affected, one observes that not only are the estimates of the latent parameters and of INB strongly dependant on the dataset chosen, but the precision of these estimates is compromised by incomplete use of available data. Further, even the more reasonable datasets result in seriously biased over-estimates of EVPI.

Discussion

This example illustrates how Bayesian MCMC methods can be used to statistically estimate a decision model in situations where some of the data informs complex functions of parameters rather than the parameters themselves, and where there are more items of data than there are parameters. In this section we compare statistical evidence synthesis to the traditional approach where model parameters are informed individually, and then review earlier related work on evidence synthesis, before noting some of the wider implications and directions for future work.

Compared to populating decision models one parameter at a time, statistical synthesis via Bayesian MCMC, or by other methods, adds to the conceptual and technical complexity of decision modelling, but it has important advantages. The first is that by using all the available information we should obtain more precise estimates than is possible if only a subset of the data is used. If the number of data points is limited to equal the number of parameters, the choice of which data points to use becomes arbitrary. For example, the parameters *a,b,c,d,e* could be estimated from data points (1,2,3,4,5), or (1,2,3,4,6). In fact there are over 100 different 'minimum' subsets of the 12 data items that could be used to estimate INB. The results of fitting MCMC models to just a few of these subsets (Table 6) illustrated how this can lead to extreme over-estimation of uncertainty, and of the expected cost of making a decision under uncertainty as measured by EVPI. EVPI is now increasingly recognised as a method for sensitivity and analysis and for research prioritisation [28 41 42 43 44 45 46] in medical

decision making, and environmental health risk assessment [47 48] . Within a decision theory framework, over-estimation of EVPI would lead to incorrect assessment of research priorities, and inappropriate allocate of resources. But even if a less formal view is taken of the research process, it is clear that an accurate identification and assessment of uncertainty must play a critical role.

In the present example, even if the number of available data items really did equal the number of parameters, it would not in any case be possible to carry out a probabilistic sensitivity analysis or estimate EVPI using conventional MC methods[41 42 46]. This is because INB is a function of unobserved parameters. It might appear that this can be circumvented. For example, in order to obtain an estimate of uncertainty in the parameter *e* from data items 1 through 5, one could consider carrying out MC simulation at each data point, and then solve the simultaneous equations for *e* on each MC cycle [49]. However, this would not necessarily lead to correct propagation of uncertainties, as there would be no way of preventing MC samples of latent parameters being outside their possible range. In the first four 'minimum' datasets in Table 6, at least one probability parameters is assigned a value outside the {0,1} range in 81%, 38%, 58%, and 2% of MC simulations.

A further advantage of using all available data is that if there is data on more functions of parameters than there are parameters, it becomes possible to check the consistency of the data sources, and the overall fit of the model to the data. We have illustrated this with a number of diagnostic and model checking tools. Bayesian model diagnostics are a very active area of research at present, and the interested reader is referred to standard texts [33 34] and recent work.[39] The use of formal methods for model checking has not been particularly common in medical decision making. It is occasionally recommended that additional data items, relating to functions of several parameters, are used to 'validate' the chosen values and ranges for the basic model parameters [50]. Alternatively, an investigator might use this additional information to 'calibrate' the basic parameter ranges in an attempt to harmonise the model with additional data sources not incorporated in the model itself. This is certainly feasible when there is just one additional source of data, but increasingly complex if there are several. Although the emphasis on calibration and validation reflects an appreciation of the need to make parameter ranges as credible as possible, we see our proposal as

having two advantages: first, the tests of consistency are formal and based on all available evidence, including other indirect evidence on other functions of parameters; second, the validating data are – if consistent – incorporated into the model, increasing precision, and leading to a more realistic assessment of uncertainty.

These proposals for statistical combination of evidence are an attempt to carry forward the pioneering work of DM Eddy and colleagues on the Confidence Profile Method [7 8 9 10], but they go beyond the CPM literature in some respects. Bayesian MCMC via WinBUGS has several technical advantages over the FASTPRO software made available for CPM [51], in which maximum likelihood and non-MCMC Bayesian approaches were offered. These include the flexibility to programme ad hoc ancillary analyses like those required for model checking, in some cases increased accuracy in estimated posterior distributions [52], and the facility to generate programme code directly from conditional independence graphs[52]. It should be noted that an alternative, non-Bayesian, estimation could also be achieved using the non-linear regression modules within a number of statistics packages. However, the partial derivatives of each of the 12 functions with respect to each of the 9 parameters would have to be provided to ensure stability: a substantial additional programming effort in itself. Further programming would then be required to obtain confidence intervals for arbitrarily complex functions of parameter such as INB, and to compute EVPI.

Although CPM in both its maximum likelihood and Bayesian forms is capable of evidence synthesis in cases where the number of functions of parameters on which there are data exceeds the number of parameters, there are relatively few examples of this in the CPM literature. Further, in these examples, the issue of consistency of evidence was not addressed. The approach to statistical evidence synthesis proposed here both enables one to use all the available evidence that bears on a problem, and to check that this evidence is consistent before proceeding. Although the best response to conflicting evidence remains a matter of judgement, the ability to check and evaluate complex evidence synthesis models has wide implications in epidemiology as well as decision making. From an epidemiologist's point of view, it becomes possible to estimate unobserved variables from indirect evidence, and to discover that some evidence, for example HIV prevalence data from women attending drug abuse clinics, is probably biased, at least as an estimate of HIV prevalence in pregnant former

and/or current IDUs. Interestingly, in HIV epidemiology, two separate methods have been proposed for estimating HIV prevalence in sections of the United Kingdom population, the 'direct' method relying on the kind of data represented by items 2 and 4 [53 54], and the 'indirect' method relying on items such as 4 and 10 [55]. Both methods over-estimate the uncertainty in the estimated prevalence of infection because they have no way of capturing the negative correlation between the estimates of group population size and group-specific prevalence.

The idea that parameter estimates should be based on 'all available evidence' echos the emphasis now rightly placed on systematic review in meta-analysis [56 57], and carries it over into the multi-parameter setting. The claim that an entire decision model is based on all available evidence, which has been checked for consistency, is a powerful one, that could enhance the model's credibility and the acceptability of decisions based on it. Further if research prioritisation is to be driven by the need to reduce uncertainty, with EVPI as one possible way of quantifying both the extent of uncertainty and its consequences, there appears to be a need to move towards a systematic use of available evidence and towards statistical methods for evidence synthesis that measure uncertainty accurately.

MCMC methods for multi-parameter evidence synthesis can be applied to many other evidence structures, including indirect treatment comparisons and multiple intermediate outcomes. There are several areas for further research. First, an extension is required to situations where data on some functions of parameters itself arises from multiple sources, as in a standard meta-analysis, which may exhibit a within-study heterogeneity. Second, current MC methods for assessing the EVPI for subsets of the data[41 46], assume both a linear relation between parameters and net benefit and a lack of correlation between parameters. New methods will be needed for EVPI on parameters in correlated structures, and for EVPI on functions of parameters. Finally, we began with a decision problem that required four parameters to estimate INB, *a*, *b*, *e* and *h*, but ended by estimating a 9-parameter model. If the concept of using all available information is interpreted as including information on ancillary parameters, this raises the question of how systematic review can identify which information is relevant.

Table 1. Data sources available, and the parameters or functions of parameters estimated. SSA Sub-Saharan Africa, IDU injecting drug user.

	Description of data items	Parameter or Function of parameters estimated	Data	Ref
1	Proportion born Sub-Saharan Africa, 1999	a	11044 /104577	[18]
2	Proportion IDU last 5 years	b	12 / 882	[19]
3	HIV prevalence, women born in SSA, 1997-8	c	252 / 15428	[20]
4	HIV prevalence in female IDUs, 1997-9	d	10/ 473	[21]
5	HIV prevalence, women not born in SSA, 1997-8	[db + e(1-a-b)]/(1-a)	74 / 136139	[20]
6	Overall HIV seroprevalence in pregnant women, 1999	ca + db + e(1-a-b)	254 / 102287	[21]
7	Diagnosed HIV in SSA women as a proportion of all diagnosed HIV, 1999	fca / [fca + gdb + he(1-a-b)]	43 / 60	[23]
8	Diagnosed HIV in IDUs as a proportion of non –SSA diagnosed HIV, 1999	gdb / [gdb + he(1-a-b)]	4 / 17	[23]
9	Overall proportion HIV diagnosed	[fca+gdb+he(1-a-b)] / [ca + db +e(1-a-b)]	87 /254	[21 24]
10	Proportion of infected IDUs diagnosed, 1999	g	12/15	[21]
11	Prop of serotype B in infected women from SSA, 1997-8	w	14/118	[26]
12	Prop of serotype B in infected women not from SSA, 1997-8	[db /[db + e(1-a-b)]] + we(1-a-b)/ [db + e(1-a-b)]	5 / 31	[26]

Table 2. Estimates of parameters and functions of parameters and model diagnostics for full dataset. SSA sub-Saharan Africa, IDU injecting drug user.

Parameters/functions	**Data**	**Estimate**	**95% interval**	**Dev**	**St. Res**	**P(extr)**
Proportion SSA	0.106	0.106	0.104, 0.018	1.01	-0.10	0.47
Proportion IDU	0.0137	0.00894	0.0047, 0.015	2.90	1.77	0.13
HIV prevalence in SSA	0.0163	0.0172	0.016, 0.019	1.33	-0.98	0.26
HIV prevalence in IDU	0.0211	0.0124	0.0063, 0.021	3.41	2.26	0.09
HIV prevalence in non-SSA	0.000544	0.000597	0.00048, 0.00073	1.51	-0.85	0.29
Overall HIV prevalence	0.00248	0.00235	0.0022, 0.0025	1.19	1.36	0.23
Diagnosed SSA as a proportion of all diagnoses	0.717	0.688	0.58, 0.79	1.02	0.53	0.37
Diagnosed IDU as prop of non-SSA diagnoses	0.235	0.304	0.16, 0.47	0.85	-0.86	0.32
Proportion of all HIV that is diagnosed	0.343	0.351	0.30, 0.41	1.03	-0.30	0.42
Probability already diagnosed, IDU	0.800	0.739	0.51, 0.91	1.11	0.59	0.37
Proportion subtype B in SSA	0.119	0.113	0.066, 0.17	0.95	0.23	0.41
Proportion subtype B non-SSA	0.161	0.288	0.20, 0.39	2.98	-2.58	0.09

Table 3. Comparison of models using global measures of goodness of fit. SSA sub-Saharan Africa, IDU injecting drug user.

	All 12 data points	Exclude item 2, Proportion IDU in population	Exclude item 4. HIV prevalence in IDUs	Exclude item 12, Serotype B in non-SSA
Data points	12	11	11	11
Parameters	9	9	9	9
Degrees of freedom	3	2	2	2
Posterior mean deviance	19.3	11.5	11.5	16.0
Deviance Information Criterion	28.4	20.1	20.1	24.9
Estimated deviance at ML solution	10.37	2.53	2.60	7.39
Sum of squared standardised residuals	20.0	5.8	5.8	10.7

Table 4. Estimates of parameters and functions of parameters with model diagnostics from reduced 11-item dataset, without data on HIV prevalence in IDUs. SSA sub-Saharan Africa, IDU injecting drug user.

Parameters/functions	**Data**	**Estimate**	**95% interval**	**Dev**	**St. Res**	**P(extr)**
Proportion SSA	0.106	0.106	0.104, 0.108	1.00	-0.10	0.47
Proportion IDU	0.0137	0.0136	0.0070, 0.022	1.02	0.00	0.27
HIV prevalence in SSA	0.0163	0.0172	0.016, 0.019	1.32	-0.96	0.35
HIV prevalence in IDU	0.0211	0.00514	0.0015, 0.012	-	-	0.20
HIV prevalence in non-SSA	0.000544	0.000581	0.00046,0.00071	1.21	-0.59	0.49
Overall HIV prevalence	0.00248	0.00234	0.0021, 0.0025	1.39	1.53	0.40
Diagnosed SSA as a proportion of all diagnoses	0.717	0.713	0.60, 0.81	0.91	0.07	0.47
Diagnosed IDU as prop of non-SSA diagnoses	0.235	0.214	0.082, 0.39	0.77	0.27	0.45
Proportion of all HIV that is diagnosed	0.343	0.346	0.29, 0.40	0.98	-0.10	0.46
Probability already diagnosed, IDU	0.800	0.770	0.56, 0.93	0.88	0.31	0.23
Proportion subtype B in SSA	0.119	0.117	0.068, 0.17	0.88	0.06	0.47
Proportion subtype B non-SSA	0.161	0.227	0.15, 0.33	1.13	-1.40	0.27

Table 5. Effect of evidence synthesis on coefficient of variation (posterior standard deviation divided by posterior mean). SSA sub-Saharan Africa, IDU injecting drug user.

Parameters/functions	**Data**	**Model based on 12 data points**	**Model excluding item 4: HIV prevalence in IDUs**
Proportion SSA	0.0090	0.0090	0.0090
Proportion IDU	0.29	0.29	0.29
HIV prevalence in SSA	0.062	0.050	0.051
HIV prevalence in IDU	0.31	0.31	0.52
HIV prevalence in non-SSA	0.12	0.11	0.11
Overall HIV prevalence	0.063	0.041	0.042
Diagnosed SSA as a proportion of all diagnoses	0.080	0.079	0.078
Diagnosed IDU as prop of non-SSA diagnoses	0.42	0.26	0.37
Proportion of all HIV that is diagnosed	0.087	0.083	0.085
Probability already diagnosed, IDU	0.13	0.14	0.13
Proportion subtype B in SSA	0.25	0.24	0.23
Proportion subtype B non-SSA	0.40	0.17	0.21

Table 6. Comparison of full evidence synthesis based on all available evidence, with synthesis based on 'minimum' datasets. Posterior means and standard deviations of: latent parameters *e*, HIV prevalence in low-risk women; *h*, the proportion of infected low-risk women diagnosed before prenatal attendance; Incremental Net Benefit, and EVPI..

Dataset	*e (sd)*	*h (sd)*	$E[INB_U]$, £ (*sd)*	$EVPI_U$, £
MCMC full evidence synthesis				
All 12 data points	0.000485 (0.000065)	0.427 (0.12)	1,028,000 (516,000)	71,240
Excluding item 4	0.000517 (0.000066)	0.408 (0.11)	1,149,000 (543,700)	73,290
MCMC with data items:				
1,2,3,4,5,8,10	0.000381 (0.000074)	0.759 (0.18)	163,900 (395,600)	532,400
1,2,3,4,5,7,8,9	0.000298 (0.000086)	0.652 (0.19)	250,900 (417400)	404,800
1,2,3,4,6,8,10	0.000692 (0.00018)	0.713 (0.20)	687,600 (851,400)	292,500
1,2,3,4,6,7,8,9	0.000536 (0.00021)	0.464 (0.20)	1,202,000 (1,066,000)	178,800
2,3,4,7,8,9,11,12	0.899 (0.07)	0.086 (0.077)	157,000,000 (65,350,000)	3,661,000
1,2,3,4,8,10,11,12	0.158 (0.23)	0.129 (0.19)	714,000,000 (1120,000,000)	18,790,000

APPENDIX

The BUGS code for the full model based on all 12 data items is:

```
model;
{
#  SET PRIORS
   a ~ dbeta( 1,2)
   z ~ dbeta (1,1)
   b <- z * (1-a)            #  sets constraint (1-a-b > 0)
   c ~ dbeta (1,1)
   d ~ dbeta (1,1)
   e ~ dbeta (1,1)
   f ~ dbeta (1,1)
   g ~ dbeta (1,1)
   h ~ dbeta(1,1)
   w ~ dbeta(1,1)

# VECTOR p[1:12] HOLDS THE EXPECTED PROBABILITIES FOR EACH DATA POINT
   p[1] <- a
   p[2] <- b
   p[3] <- c
   p[4] <- d
   p[5] <- (d*b + e*(1-a-b))/(1- a)
   p[6] <- c*a + d*b + e*(1-a-b)
   p[7] <- f*c*a / (f*c*a + g*d*b  + h*e*(1-a-b))
   p[8] <- g*d*b / (g*d*b + h*e*(1-a-b))
   p[9] <- (f*c*a + g*d*b  + h*e*(1-a-b)) / p[6]
   p[10] <- g
   p[11] <- w
   p[12] <- d*b/(d*b+e*(1-a-b))  +  w*e*(1-a-b)/(d*b + e*(1-a-b))

# NET BENEFIT OF MATERNAL DIAGNOSIS, INCREMENTAL NET BENEFIT, EVPI
   y ~ dgamma( 0.56,3)I(, 2)                                  #  distribution for net benefit of maternal diagnosis
   nbmd <-  60012 - 54296*y                                                                  #  see equation 2
   inb.u <-  105000*(1-a-b) * (nbmd * e * (1-h) - 3.0*(1-e*h))                          # annual INB, see equation 1
   evpi.u <- 7.7217 * max(-inb.u,0)                        # expected value of perfect information, see equation 3

#  LIKELIHOOD AND DIAGNOSTICS
   for(i in 1: 12) {
      r[i] ~ dbin(p[i],n[i])                                                  #  data numerators and denominators
      rhat[i] <- p[i] * n[i]                                                        # predicted value of numerators
      dev[i] <- 2 * (r[i] * (log(r[i])-log(rhat[i]))  +  (n[i]-r[i]) * (log(n[i]-r[i]) - log(n[i]-rhat[i])))          #  deviance contribution
      p.rep[i] ~ dnorm(p[i],t[i])         # replicate observation, normal approximation to binomial, mean p[i], variance 1/t[i]
      t[i] <- n[i]/( p[i]*(1-p[i]))                                                  #  1/variance of a proportion
      p.smaller[i] <- step( (r[i]/n[i]) - p.rep[i])                    # counts number of times replicate data < observed
    }
   sumdev <- sum(dev[])                                                          # sum of deviance contributions
}
```

with the following data:

```
list(
r=c(11044,12,252,10,74,254,43,4,87,12,14,5),
n=c(104577,882,15428,473,136139,102287,60,17,254,15,118,31)
)
```

The model checking diagnostics and measures of goodness of fit are computed as follows:

Deviance (Tables 2,4): mean of the posterior distribution of dev[i].

Standardised residuals (Tables 2,4): (Observed – expected probabilities)/standard deviation of expected. For example for data item 1: observed= 12/882= 0.0136, mean

of posterior distribution of p[2] is 0.0089, sd of posterior distribution of p[2] is 0.00466.

Probability of obtaining a more extreme observation (P(Extr)) (Tables 2,4): The node p.smaller[i] monitors the number of MCMC cycles in which the observed data exceeds a replicate data point p.rep[i] based on the p[i] on that cycle and the denominator. The tabulated P(Extr) is Min(Prob($p.rep_i<r_i/n_i$),1- Prob($p.rep_i<r_i/n_i$)). (Note that the normal approximation to the binomial gives adequate accuracy and speeds processing with large denominators. The function step(x) returns 1 if x = 0, else 0).

Posterior mean deviance, E[D] (Table 3): posterior mean of sumdev, a node that monitors the sum of the deviance contributions of each item.

Deviance Information Criterion (Table 3): a measure of global fit that combines model fit as expressed by E[D], the posterior mean deviance, together with a penalty for the effective number of parameters p_D.[39] DIC = E[D] + p_D, where p_D = E[D] - D(E[θ]), and D(E[θ]) is the deviance calculated by plugging in the posterior mean values of the parameters into the formula for the deviance contribution. In the present case, p_D should approximate the degrees of freedom.

Estimated deviance at the Maximum Likelihood solution (Table 3): The posterior mean deviance minus half its variance is an estimate of the deviance at the maximum likelihood solution.[39]

Sum of squared standardized residuals (Table 3): the sum of the squared standardized residuals presented in Table 2.

Coefficient of variation (Table 5). Posterior standard deviation of the fitted probabilities divided by the posterior mean.

FIGURE LEGENDS

Figure 1.

Model of the epidemiological prenatal screening. The values of the eight parameters, *a,b,c,d,e,f,g,h* are to be estimated. The decision problem relates to a choice between universal testing of all groups, or targeted testing of Black African and IDU women.

Figure 2.

Posterior correlations between parameters based on the full 12-item dataset. The correlations are calculated from the successive estimates of each parameter from the Monte Carlo Markov Chain. Forward slash indicates positive correlation, backward negative.

References

1. Naglie G, Krahn MD, Naimark D, Redelmeier DA, Detsky AS. Primer on medical decsion analysis: part 3 - estimating probabilities and utilities. *Medical Decision Making* 1997;**17**:136-141.

2. Petitti DB. *Meta-analysis, decision analysis, and cost-effectiveness analysis: methods for quantitative synthesis in medicine.* New York: Oxford University Press, 2000.

3. Doubilet P, Begg CB, Weinstein MC, Braun P, McNeill BJ. Probabilistic sensitivity analysis using Monte Carlo simulation: a practical approach. *Medical Decision Making* 1985;**5**:157-177.

4. Critchfield GC, Willard KE. Probabilistic analysis of decision trees using Monte Carlo simulation. *Medical Decision Making* 1986;**6**:85-92.

5. Thompson KM, Burmaster DE, Crouch EAC. Monte Carlo techniques for quantitative uncertainty analysis in public health risk assessments. *Risk Analysis* 1992;**12**:53-63.

6. Manning WG, Fryback DG, Weinstein MC. Reflecting uncertainty in decision analysis. In: Gold MR, Siegel JE, Russell LB, Weinstein MC, eds. *Cost-effectiveness in health and medicine*. New York: Oxford University Press, 1996;247-275.

7. Eddy DM, Hasselblad V, Shachter R. *Meta-analysis by the Confidence Profile Method*. London: Academic Press, 1992.

8. Eddy DM, Hasselblad V, Shachter R. An introduction to a Bayesian method for meta-analysis: the confidence profile method. *Medical Decision Making* 1990;**10**:15-23.

9. Eddy DM, Hasselblad V, Shachter R. A Bayesian method for synthesising evidence: the confidence profile method. *International Journal of Technology Assessment in Health Care* 1990;**6**:31-35.

10. Eddy DM. The confidence profile method: a Bayesian method for assessing health technologies. *Operations Research* 1989;**37**:210-228.

11. Shachter R, Eddy DM, Hasselblad V. An influence diagram approach to medical technology assessment. In: Oliver RM, Smith JQ, eds. *Influence diagams, belief nets and decision analysis.* New York: Wiley, 1990;321-350.

12. Spiegelhalter DJ, Thomas A, Best N. WinBUGS User Manual. (1.3). 2000. Cambridge.

13. Ades AE, Sculpher MJ, Gibb DM, Gupta R, Ratcliffe J. Cost effectiveness analysis of antenatal HIV screening in United Kingdom. *British Medical Journal* 1999;**319**:1230-1234.

14. Ades AE, Gupta R, Gibb DM, et al. Selective versus universal antenatal HIV testing: epidemiological and implementational factors in policy choice. *Aids* 1999;**13**:271-278.

15. Ades AE, Sculpher MJ, Gibb D, Ratcliffe J, and Gupta R *A cost-effectiveness analysis of antenatal HIV testing in the UK*. 1999.

16. Ades AE, Ratcliffe J, Gibb DM, Sculpher MJ. Economic issues in the prevention of vertical transmission of HIV. *Pharmacoeconomics* 2000;**18**:9-22.

17. Anonymous *Reducing the odds: preventing perinatal transmission of HIV in the United States*. Washington DC: National Academy Press, 1999.

18. Office for National Statistics *Birth Statistics, 1999*. London: The Stationary Office, 2000.

19. Johnson AM, Wadsworth J, Wellings K, Field J. *Sexual attitudes and lifestyles*. Oxford: Blackwell Scientific Publications, 1994.

20. Ades AE, Walker J, Botting B, Parker S, Cubitt D, Jones R. Effect of the worldwide epidemic on HIV prevalence in the United Kingdom: record linkage in anonymous neonatal seroprevalence surveys. *Aids* 1999;**13**:2437-2443.

21. Public Health Laboratory Service, Institute of Child Health, and Scottish Centre for Infection and Environmental Health *Unlinked anonymous prevalence monitoring programme: annual report Supplementary Data set: data to end of 1999*. 2000.

22. Ades AE, Davison CF, Holland FJ, et al. Vertically transmitted HIV infection in the British Isles. *British Medical Journal* 1993;**306**:1296-1299.

23. Nicoll A, McGarrigle C, Brady AR, et al. Epidemiology and detection of HIV-1 among pregnant women in the United Kingdom: results from national surveillance 1988-96. *British Medical Journal* 1998;**316**:253-258.

24. Holland FJ, Ades AE, Davison C, et al. Use of anonymous newborn serosurveys to evaluate antenatal HIV screening programmes. *Journal Of Medical Screening* 1994;**1**:176-179.

25. Hu DJ, Dondero TJ, Rayfield MA, et al. The emerging genetic diversity of HIV: the importance of global surveillance for diagnostics, research, and prevention. *JAMA* 1996;**275**:210-216.

26. Ades AE, Parker S, Walker J, Weber JN. Serotype, parental country of birth, and ethnic status in unlinked anonymous neonatal HIV seroprevalence surveys. *Journal Of Acquired Immune Deficiency Syndromes* 2000;**24**:292-294.

27. Zeuner D, Ades AE, Karnon J, Brown J, Dezateux C, and Anionwu AN. Antenatal and neonatal haemoglobinopathy screening in the UK: review and economic analyss. *Health Technology Assessment* 1999;**3**:1-183.

28. Claxton K, Posnett J. An economic approach to clinical trial design and research priority-setting. *Health Economics* 1996;**5**:513-524.

29. Stinnett A, Mullahy J. Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analyses. *Medical Decision Making* 1998;**18**:S68-S80.

30. Department of Health *Register of cost-effectiveness studies*. London: Department of Health, 1995.

31. Raiffa H. *Decision analysis: introductory lectures on choices under uncertainty.* Reading, Mass: Addison-Wesley, 1961.

32. Schlaiffer R. *Probability and statistics for business decisions*. New York: McGraw-Hill, 1958.

33. Gilks WR, Richardson S, Spiegelhalter DJ. *Markov Chain Monte Carlo in Practice.* London: Chapman & Hall/CRC, 1996.

34. Carlin JG, Louis TA. *Bayes and empirical Bayes methods for data analysis.* Boca Raton: Chapman & Hall/CRC, 2000.

35. Congdon P. *Bayesian Statistical Modelling*. Chichester: Wiley, 2001.

36. Gelfand AE, Smith AFM. Sampling-based approaches to calculating marginal densities. *Journal Of The American Statistical Association* 1990;**85**:398-409.

37. Besag J. Spatial interaction and the statistical analysis of lattice systems (wth discussion). *Journal of the Royal Statistical Society (B)* 1974;**36**:192-236.

38. Geweke J. Evaluating the accuracy of sampling based approaches to the calculation of posterior moments. In: Bernardo JM, Berger J, Dawid AP, Smith AFM, eds. *Bayesian Statistics 4*. Oxford: Oxford University Press, 1992;169-193.

39. Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society (B)* 2002;**64**:in press.

40. Gelman A, Carlin JG, Stern HS, Rubin DB. *Bayesian data analysis*. London: Chapman and Hall, 1995.

41. Felli JC, Hazen GB. Sensitivity analysis and the expected value of perfect information. *Medical Decision Making* 1998;**18**:95-109.

42. Thompson KM, Graham JD. Going beyond the single number;: using probabilistic risk assessment to improve risk management. *Human and Ecological Risk Assessment* 1996;**2**:1008-1034.

43. Thompson KM, Evans JS. The value of improved national exposure information for perchloroethylene (Perc): a case study for dry cleaners. *Risk Analysis* 1997;**17**:253-271.

44. Claxton K. The irrelevance of inference: a decision-making approach to the stochastic evaluation of heath care technologies. *Journal of Health Economics* 2000;**18**:341-364.

45. Claxton K, Lacey LF, Wlaker SG. Selecting treatments: a decision theoretic approach. *Journal of the Royal Statistical Society (A)* 2000;**163**:211-226.

46. Fenwick E, Claxton K, Sculpher M, and Briggs A *Improving the efficiency and relevance of heath technology assessment: the role of iterative decision analytic modelling*. York: Centre for Health Economics, University of York, 2000.

47. Brand KP, Small MJ. Updating uncertainty in an integrated risk asessment: conceptual framework and methods. *Risk Analysis* 1995;**15**:719-731.

48. Dakins ME, Toll JE, Small MJ, Brand KP. Risk-based environmental remediation: Bayesian Monte Carlo analysis and the expected value of sample information. *Risk Analysis* 1996;**16**:67-79.

49. Ratcliffe J, Ades AE, Gibb D, Sculpher MJ, Briggs AH. Prevention of mother-to-child transmission of HIV-1 infection: alternative strategies and their cost-effectiveness. *Aids* 1998;**12**:1381-1388.

50. Mandelblatt JS, Fryback DG, Weinstein MC, Russell LB, Gold MR, Hadorn DC. Assessing the effectiveness of health interventions. In: Gold MR, Siegel JE, Russell LB, Weinstein MC, eds. *Cost-effectiveness in Health and Medicine*. New York: Oxford University Press, 1996;135-175.

51. Eddy DM, Hasselblad V. *Fast*Pro: software for meta-analysis by the confidence profile method*. San Diego: Academic Press, 1992.

52. Spiegelhalter DJ, Myles JP, Jones DR, and Abrams KR. Bayesian methods in Health Technology Assessment. *Health Technology Assessment* 2000;**4**:

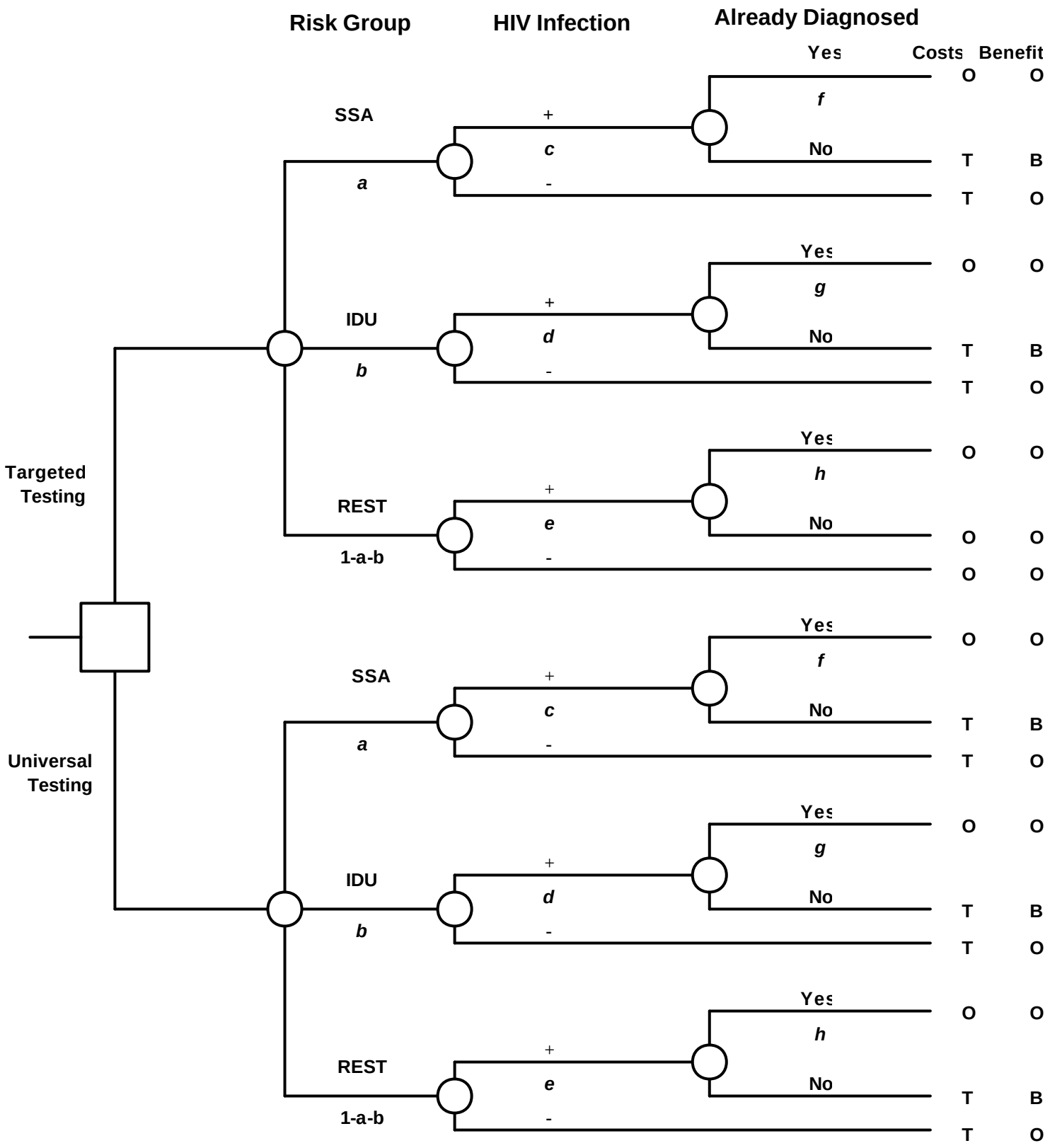
53. Giesecke J, Johnson A, Hawkins A, et al. An estimate of the prevalence of human immunodeficiency virus in England and Wales by using a direct method. *Journal of the Royal Statistical Society (A)* 1994;**157**:89-103.

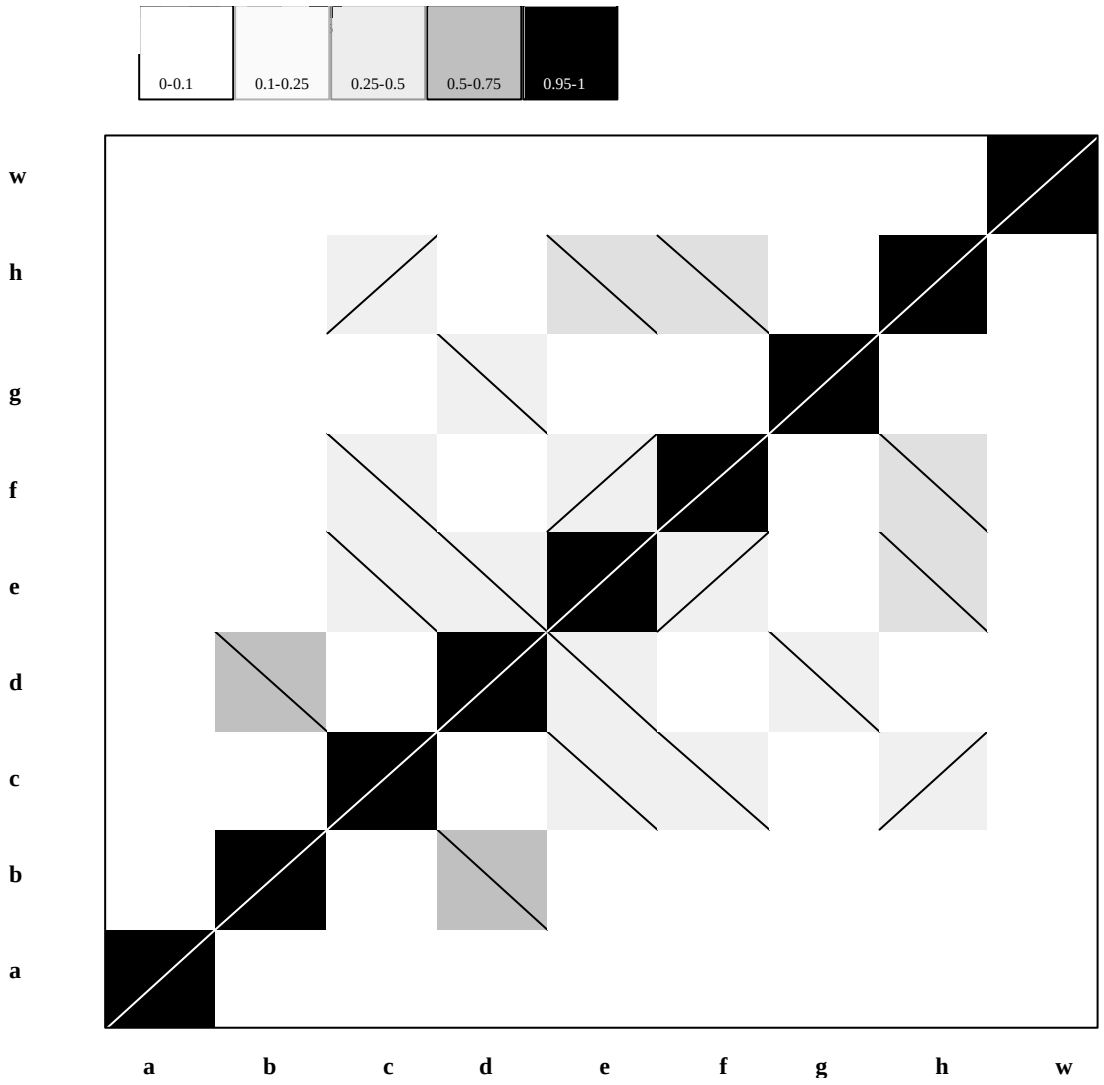
54. Petruckevitch A, Nicoll A, Jonhson AM, Bennet D. Direct estimates of prevalent HIV infection in adults in England and Wales for 1991 and 1993. *Genitourinary Medicine* 1997;**73**:348-354.

55. Hughes G, Porter K, Gill N. Indirect methods for estimating prevant HIV infections: adults in England and Wales at the end of 1993. *Epidemiology And Infection* 1998;**121** :165-172.

56. Chalmers I, Altman DG. *Systematic Reviews*. London: BMJ Publishing Gorup, 1995.

57. Song F, Eastwood AJ, Gilbody S, and Sutton AJ. Publication and related biases. *Health Technology Assessment* 2000;**4**:


Risk Group
HIV Infection
Already Diagnosed
Yes
Costs
Benefit
Targeted Testing
Universal Testing
SSA
IDU
REST
1-a-b
a
b
1-a-b
+
-
c
d
e
f
g
h
Yes
No
O
T
B
O


0-0.1
0.1-0.25
0.25-0.5
0.5-0.75
0.95-1
w
h
g
f
e
d
c
b
a
a
b
c
d
e
f
g
h
w