Interpretation of the expected value of perfect information in economic evaluations: a review

Joanna Thorn[1], Joanna Coast[2] and Lazaros Andronis[2]

1. School of Social and Community Medicine, University of Bristol
   (joanna.thorn@bristol.ac.uk)
2. Health Economics Unit, School of Health and Population Sciences, University of Birmingham

Abstract

Background: It is becoming common for expected value of perfect information (EVPI) calculations to be performed alongside economic evaluations to guide further funding decisions. However, decision rules based on these values are not clear.

Aims: To determine whether there exists a threshold value below which research is typically not recommended, and to consider the factors affecting research recommendations.

Methods: A systematic literature review was undertaken to identify applied EVPI calculations in the healthcare field. Study characteristics were extracted, including funder, location, disease group, publication year, primary language and outcome measure. Population EVPI values and willingness-to-pay thresholds were also extracted alongside verbatim text excerpts describing the authors' research recommendations. The recommendations were classified according to whether further research was definitely recommended (a positive recommendation) or not (negative). A text analysis based on the principles of discourse analysis revealed that while some authors gave very confident recommendations either for or against further research, others were more equivocal. This analysis led to a classification schema describing the confidence with which the recommendation was presented. Factors affecting the likelihood of a positive or confident recommendation were examined.

Results and discussion: Study location, funder and disease group appear to affect both the confidence and direction of the recommendation. Language, willingness-to-pay threshold and publication year affect the nature of the recommendation, while outcome measure affects the confidence. A threshold EVPI value below which research is typically not recommended exists empirically at around £250 000. These results may help clarify the decision rules surrounding EVPI values.

1. Introduction

Decision making within health care based on cost-effectiveness considerations is rarely accomplished without some degree of uncertainty as to its accuracy (Briggs *et al*, 2006). However, the options open to a decision maker are not simply to implement immediately or not to implement; a further option to collect additional information to inform the cost-effectiveness analysis can also be invoked (Eckermann and Willan, 2007). By commissioning further studies, a decision maker can reduce, or occasionally eliminate, the uncertainty surrounding the decision. Whether gathering further information is beneficial or not depends on the net benefit the research is expected to achieve.

A formal framework, with roots in statistical decision theory (Schlaifer, 1959), has been developed to assess the value of information (VOI) to a decision maker in health care (Claxton and Posnett, 1996; Claxton, 1999). A key VOI measure is the 'expected value of perfect information' (EVPI), which represents the monetary value that can be attached to completely eliminating uncertainty in the decision-making process. The EVPI value for an individual is defined as the difference between the value associated with a decision made on the basis of current information, and the value that could be expected if perfect information were available on which a decision could be based (Briggs *et al*, 2006). However, a more appropriate comparative measure for the value of acquiring further information is the population EVPI, which takes into account the number of people who may benefit from the additional research and incorporates measures of both the time frame over which the information is expected to retain its usefulness (before, for example, newer technologies render the intervention obsolete), and the number of people with the condition.

EVPI has the potential to be used as a means of assessing research priorities in a funds-limited research environment. If the cost of obtaining further information (via a randomised controlled trial (RCT), for example) exceeds the EVPI, there is little justification for proceeding with research, and a decision maker can be confident that they could not make a better decision by waiting. Thus the EVPI exceeding the cost of running a trial is a necessary condition that must be fulfilled before research can be considered potentially worthwhile and represents a maximum amount that a rational decision maker should spend on further research (Philips *et al*, 2006). However, it should be noted that a high EVPI value is not a sufficient condition for advising further research; more information from expected value of sample information (EVSI) studies is required to determine whether a particular piece of research should be done (Ades *et al*, 2004).

Although it is now becoming relatively common for an EVPI calculation to be performed alongside a cost-effectiveness analysis, there is no overall decision framework under which guidance to a decision maker on priorities for a national or international research program can be given. One problem with implementing EVPI-based decision making is that decisions for research and implementation are carried out by different bodies in the UK (Claxton and Sculpher, 2006). The complexity of the international research-commissioning environment, with differing priorities, populations and disease incidences, further complicates the issue. A potential barrier to the practical use of EVPI values as a means of prioritising research recommendations is that there is a lack of clarity concerning decision rules based on the figures calculated *i.e.* there is no acknowledged threshold EVPI value below which research should not be recommended.

In this study, we conduct a systematic literature review of applied EVPI calculations and explore quantitative data in a descriptive manner. We employ a method based on the approach of discourse analysis, a form of qualitative study that takes a language-based approach to examining reasons for presenting information in a particular manner (Bryman, 2008). We aim to determine whether there exists an empirical magnitude of EVPI below which no recommendation for further health research is typically made (*i.e.* whether there is an empirical threshold), whether there is consistency across the literature in the recommendations for further research for a given level of EVPI, and to investigate whether there are any factors that characteristically influence recommendations.

2. Methods

2.1 Article identification

Articles reporting applied examples of EVPI calculations were systematically sought, covering 1990 to 2010 inclusive. Searches established that applied EVPI calculations were not performed or reported in health economics prior to the publication of the methodological description by Claxton and Posnett (1996); therefore, it is realistic to consider that the time frame covered will not have led to any important literature being overlooked.

Calculations of VOI are not always referred to in the abstract or keywords of an article. Therefore, in order to maximise the specificity of the searches without reducing the sensitivity, two different approaches to searching the literature were taken. In the first, standard bibliographic databases were searched with relatively broad search terms, whilst in the second, full-text searching was performed with tightly defined search terms.

The bibliographic databases Medline, EMBASE, CINAHL, Web of Science and The Cochrane Library were searched using a combination of search terms and wildcards to cover the range of different value of information phrases. Adding the qualifier "AND cost" significantly improved the specificity of the searches without reducing the sensitivity.

Full-text searching was undertaken via the websites of the journal publishers and suppliers AdisOnline, HighWire Press, IngentaConnect, Cambridge Journals Online, ScienceDirect and the UK Health Technology Assessment (HTA) site, covering the significant journals in health economics. The quality of the search engines and documentation provided was variable; therefore, search syntax was largely tested empirically for each site. To avoid the excessive noise associated with a lack of specificity, very limited and explicit search terms were employed [("expected value of perfect information" OR EVPI) AND cost] and, where it was possible to limit searches to appropriate subject areas and years, this was done. Full-text searches were also conducted using Google Scholar.

2.2 Article screening

Multiple abstracts cannot be downloaded simultaneously from Scholar; therefore, Scholar results were screened for inclusion at the abstract level via Google itself. Many could be eliminated from the Scholar text alone; where this was not possible, links to abstracts were followed, and an inclusive approach was taken to downloading them individually into a Refman 12 database *i.e.* articles were only excluded if it was very clear that they were not relevant. With all the potentially relevant abstracts from Scholar and other searches in a single Refman database, a deduplication process was undertaken electronically, with the criteria for reduction set at a high level of required concordance; this resulted in some duplicates passing to the next stage, but ensured that no abstracts were rejected erroneously. The remaining abstracts were exported into an Access 2007 database.

Abstracts were screened by one author (JT) applying a generously inclusive policy *i.e*. any abstracts that described cost-effectiveness studies of any description were included even if they did not specifically refer to VOI. Articles were excluded if they were not written in English for pragmatic reasons. Only peer-reviewed publications were included as they represented articles that could reasonably be expected to inform policy. Articles were also excluded if the EVPI calculation was carried out purely to illustrate a methodological point; articles where this was the case were defined initially as those that did not mention the disease area in the title or abstract, as these articles were unlikely to inform policy in a

particular area. A 10% sample of the abstracts was screened by a second reviewer (LA) to check for accuracy and consistency.

Electronic full-text versions in portable document format (pdf) were obtained for articles identified for inclusion via library subscriptions and free sources. For unavailable articles, health economic assessments in the NHS Economic Evaluation Database (EED) were consulted; abstracts for which the EED record did not mention VOI studies were eliminated. For the remaining articles, a request was sent to corresponding authors for whom an email address was available, asking for either confirmation that no VOI study had been performed or a copy of the article.

Multiple pdf file search functionality was used to search for the word 'perfect' in order to eliminate articles that would not contain an EVPI calculation. Non-searchable pdfs were identified, and these articles were manually scanned for EVPI calculations. For articles containing the word 'perfect', the context was examined to eliminate irrelevant material. A second screening cycle was applied to those articles that either did not contain the word 'perfect' or contained it in an irrelevant context by searching for the word 'information'.

A final screening process was undertaken by reading the full text and eliminating articles that did not describe an applied EVPI calculation. Some studies were reported twice; only the most recently published article was included to maximise the likelihood of a full report. However, the earlier report was used to supply additional details where necessary.

2.3 Data extraction

Data extracted included the funder, location and disease group (ICD-10 chapter heading) to give some background characteristics. Extracted information specific to economic evaluation included the study perspective, whether it was carried out beside a trial or purely as a model, and the cost year (where available). Both individual and population EVPI values were extracted, along with the WTP threshold and currency. The population covered, and the time frame over which the technology was expected to be useful, were also extracted. Where multiple EVPI values were cited, a pragmatic approach to choosing a single value was taken; where a single value was discussed in the text, this was the value taken. Where multiple values were described in the text, a value at a WTP of £30 000 (or other standard WTP values) was taken. Finally, brief text excerpts describing the interpretation of the values, recommendations based on the values and research prioritisation comments were extracted verbatim.

The extracted texts were classified according to whether the recommendation was for or against further research (*i.e.* positive or negative), on an ordinal scale of recommendations from 'beneficial' to do further research, through 'probably beneficial', 'possibly beneficial', 'possibly not beneficial', 'probably not beneficial' to 'not beneficial' to do further research.

Extracted data were used to classify the type of funder. Population EVPI and WTP values were converted to sterling using Bank of England exchange rates, taking the value at 31 December (or closest preceding day) of the relevant cost year of the study, or the publication year if unavailable (Bank of England, 2011). Owing to the complex nature of EVPI calculations with costs bound up in WTP thresholds, it was not possible to convert EVPI values to a common cost year. A language was assigned to the authors based on the study location. The quality of the articles was not formally assessed because this did not form one of the exclusion criteria and studies with methodological limitations were not excluded on that basis alone. For example, where an inappropriately large population had been used to derive a population EVPI resulting in a hugely inflated value, the study was included in the text analysis because a recommendation was still made and flowed logically from the value calculated.

2.4 Analyses

Texts and contextual factors were exported into Word 2010 in order to conduct a text investigation based on the principles of discourse analysis: seeking 'patterned "ways of talking"' (Dixon-Woods, 2001, p1418) is intended to illustrate how writers attempt to persuade readers (Traynor, 2006). Extracted recommendation texts were repeatedly examined analytically to look for patterns of language. Data collection and coding/analysis were carried out as distinct, unmixed phases of the study (Potter and Wetherell, 1987). Key phrases were highlighted, and grouped into similar descriptive forms, which led to the abstracts being coded according to how confident or emphatic the authors were in their recommendations for further research.

The classifications were used to consider graphically how various factors affected the recommendations. Statistical and graphical analyses were performed using Stata 11.

3. Systematic review results

3.1 Data collection

2078 potentially relevant articles were identified by searching the bibliographic databases, while a further 546 were identified via full-text searches. Following deduplication, 2497 abstracts required screening. Screening by two reviewers of a 10% sample (250 abstracts) resulted in good agreement on inclusion ($\kappa = 0.72$). Disagreements were resolved by discussion, and the remaining abstracts were screened by a single reviewer.

541 abstracts were deemed of unclear relevance or sufficiently relevant to merit acquiring the full-text version. To assist with finding full-text articles, emails were sent to 29 authors, requesting information about 30 articles; 17 responded covering 18 articles (60% return rate). 449 pdf versions were obtained, while 76 were eliminated from consideration by other methods; of these, 13 were rejected due to being written in a language other than English (eight in Spanish, and one each in German, Dutch, French, Italian and Japanese). It was not possible to locate 16 articles; however, these were all considered to have a low probability of relevance as they were published prior to the earliest EVPI calculations or covered IT applications. Of the pdf articles acquired, electronic searching revealed that only 143 contained the word 'perfect' in a relevant context and searching for 'information' revealed no further potential EVPI calculations. Only 111 appeared to be potentially relevant examples of applied EVPI calculations and were read more closely, resulting in data extraction from 94 articles. Eight quoted only individual EVPI values; although some of these still made population research recommendations, they were excluded from the final analysis, resulting in 86 articles included.

3.2 Data characteristics

The publications included were drawn from a wide range of journals. Higher numbers of EVPI calculations have been observed in recent years. Nine studies were carried out alongside trials, with the remainder being pure modelling studies. Where stated, time frames over which the technology under study was expected to remain useful varied from one to 30 years, with the majority (38) opting for 10 years. WTP thresholds for quality-adjusted life year (QALY) outcomes ranged from £500 to nearly £80 000. Other key characteristics of the included studies are listed in table 1.

4. Text analysis

The data comprise texts extracted from published (copy-edited), peer-reviewed material. As such, they are written by scientists for the scientific community, and conform to the linguistic norms of scientific communications. Of the 86 included articles, 13 suggested no further research, whilst 66 were more positive (10 implicitly through parameter research recommendations, two on the basis of factors other than VOI results and the remainder on the basis of the EVPI value). Seven made no recommendation.

The costs of carrying out research were not frequently assessed with only three of the included articles making reference to actual figures; these estimates varied from around £172 000 to £17.5 million. Categorical recommendations were rare, with only a few explicitly using the term 'recommend'; however, many were implicit. Where an absolute, rather than a comparative, value judgment was applied, the EVPI was described as 'low', 'small', 'high', 'large' or 'substantial'.

The texts were examined for evidence of linguistic patterns in which recommendations were presented and discussed. The texts had already been classified according to whether further research was recommended or not. This process revealed that, while some authors made confident positive or negative recommendations, others were more equivocal, and a theme encompassing the perceived confidence of the authors in making their recommendation emerged from the data. Categories were developed and then used to classify texts on this basis as described below. Analyses based on the classification are not intended to reach statistical significance, or to be generalisable; however, numbers are included to give background context.

4.1 Classifications on the basis of confidence

EVPI interpretations were sometimes expressed as a restatement of the meaning of EVPI, without contextual interpretation in terms of the real or estimated cost of a trial. For example, one article (Martikainen *et al*, 2005) stated that

> 'According to EVPI analysis, future research would potentially be cost effective if the costs of research were €4.1 million (maximum)'.

Without details of what research might actually cost, this does not constitute a recommendation; texts employing this device were classified as 'none', along with those that did not appear to interpret their EVPI results at all, with a total of seven articles falling into this group.

Some texts appeared confident in their recommendations and were classified as 'unequivocal'; this level of confidence could also be considered to form instructions to decision makers. A typical example (Groot Koerkamp *et al*, 2008) is

> '… a Dutch funding agency seeking to maximize future health in the Netherlands should not fund more research regarding the value of MR imaging in patients with acute knee trauma'.

The confident group also includes positive recommendations such as (Smith *et al*, 2007)

> '… the federal government should be willing to invest in further research to reduce the uncertainty associated with …'.

Ten articles fell into the unequivocal category, with only one advising against further research. Positive recommendations should not strictly follow from EVPI values alone; however, it should be noted that these were typically made with the benefit of other VOI measures, rather than on EVPI values alone.

A further group of texts described their recommendations in terms of the perceived value of possible future research, using expressions such as 'worthwhile' or 'additional value'. For example (Speight *et al*, 2006),

> '…further primary research would appear to be worthwhile, given the large cost of uncertainty in all scenarios'.

The 24 texts in this group appeared reasonably confident about their recommendations, but were not as directive as the previous group; they were classified as 'valuable'.

Many texts used words such as 'warranted' when making recommendations; this is a slightly more equivocal approach, implying that, while it may be considered beneficial to do some more research, there is no reason to dictate that it should be done. In other words, these texts leave the decision making entirely to the decision maker. This group of texts, comprising 14 articles, was classified as 'justified'. Examples include (Eddama *et al*, 2010)

> 'There is little economic justification for conducting further research…'

Another group of texts used the likelihood of research being cost-effective to describe their recommendations. For example (Henriksson *et al*, 2006),

> '… it appears unlikely that any further research regarding this decision problem would be worthwhile.'

was typical of this group. This group of 10 articles represents relatively clear recommendations that nevertheless acknowledge that there is some uncertainty over whether research is cost-effective and was classified as 'probably'.

Texts using equivocal words such as 'possible', 'may' or 'potential' were deemed to demonstrate less confident recommendations. A negative example is (Rao *et al*, 2009)

> 'Value of information analysis suggests that further research may not be cost-effective'.

This group more frequently made positive recommendations than negative. The 21 texts were labelled as 'possibly' and represented the least emphatic of the recommendations.

The descriptions above formed an ordinal confidence classification system running from the most confident 'unequivocal' recommendations, through 'valuable', 'justified' and 'probably' to the least confident 'possibly'; where texts could have fallen into more than one group, they were classified on the basis of the most emphatic wording. The 'possibly', 'probably', 'justified' and 'unequivocal' groups were relatively straightforward to identify and formed cohesive groups. However, the 'valuable' group was more diverse in its contents. The emphasis in this schema is different to the earlier positive/negative/equivocal groupings, which leads to some differences in classification; for example, studies originally designated as positive may fall into more equivocal groups.

5. Graphical analysis

Both the recommendation and confidence scales are investigated graphically. Owing to the extensive range of EVPI values observed, the graphs are plotted on a natural log scale. For aesthetic purposes, the scale has been truncated and study numbers are omitted.

5.1 Classification based on research recommendations

The initial classification describing whether further research was recommended or not is illustrated in figure 1, showing that stronger belief that no further research should be performed is clustered at lower EVPI values, while a strong belief that research should be carried out tends to be more common towards higher EVPI values. If the data are collapsed into binary categories of no further research (including all three negative categories) and any other recommendation (figure 2), there appears to be a cut-off around an EVPI value of £250 000 below which research was not typically recommended. Between £250 000 and

£2 million, recommendations were variable, while over £2 million did not typically attract recommendations against further research. The one outlying result with a recommendation against further research at an EVPI value of £10.76 million specifically noted that an RCT in this area would be particularly costly (Rogowski *et al*, 2009).

With the data grouped into firm positive recommendations, firm negative recommendations and equivocal recommendations (including all four of the probably and possibly groups from above), it can be seen that there are a few firm recommendations against further research at the lower end of the EVPI scale, but many are equivocal (figure 3).

The recommendations can be broken down by various relevant factors. Figure 4 shows an example broken down by the disease area covered; neoplasms appear to attract a higher rate of positive research recommendations than other diseases. Similar breakdowns for other factors suggested that studies from the UK favour more equivocal recommendations, whilst those from the US and the Netherlands show higher proportions of positive recommendations. When the data are broken down by funder, it appears that government-funded studies are less likely to make firm positive recommendations than industry-funded studies. A breakdown by WTP value (for QALY outcomes only) suggests that positive research recommendations are clustered at higher WTP values; however, negative recommendations were found throughout the ranges.

A breakdown by primary language suggests that authors in English-speaking regions are more likely to make equivocal recommendations than those in non-English speaking regions, while a breakdown by publication year indicates that authors have been making more positive recommendations in recent years. A breakdown by outcome measure did not reveal any patterns, although the dominance of QALY measures mitigates against observing any differences.

5.2 Classification based on confidence of recommendations

The confidence scale derived from the text analysis is shown in the following graphs with darker shading representing more confident recommendations. Confidence is not necessarily linked to actual EVPI value (figure 5), with confident and less confident recommendations found over the entire range of EVPI values observed.

As above, the effect of various factors on the confidence of the recommendation was tested graphically. Figure 6 shows an example with the EVPI values for studies broken down by country; a higher proportion of non-emphatic recommendations is observed for studies from the UK than for those carried out in the Netherlands and US, mirroring the earlier

recommendation analysis. Similarly, a breakdown by funder appears to show that industry-funded studies are slightly more confident in their recommendations than those funded by other means. Neoplasms seem to generate more confident recommendations than other disease groups. Studies that used life years gained as an outcome measure appeared to generate more confident recommendations than those using QALYs or other measures. Breakdowns by language, WTP value and publication year did not suggest any patterns arising from these factors.

6. Discussion

It appears empirically that recommendations are reasonably consistent with EVPI values and that there is a cut-off value of £250 000 below which research is typically not recommended. Although it might be anticipated that confident recommendations would be more likely at both high and low EVPI values, with less confident recommendations in between, this was not observed.

The preponderance of studies originating from the UK is likely to arise as a result of national guidelines, with NICE having formally advocated the use of VOI methods since 2004 (NICE, 2004). These UK-based studies are more likely to offer equivocal than straight positive or negative recommendations, and to show lower confidence in the recommendation which could be a result of cultural expectations of funders. Alternatively, in the context of outcome measures, there is evidence that some nations consistently produce results positively favouring the intervention (Vickers *et al*, 1998); there may be an analogous effect in which national expectations lead to confident expressions of results.

Although achieving generalisability was not the aim of this study, limiting the included articles to English language only will have resulted in some useful articles being excluded. Particularly given that country of origin appears to have a bearing on how confident the recommendation is, it seems likely that foreign-language articles would have added useful information to the research question. The higher likelihood of an equivocal recommendation from English-speaking areas could arise from native English speakers making greater use of the range of nuances in the English language. Articles from non-English speaking areas are also likely to be subject to higher levels of copy-editing changes; a published paper represents not only the work of the authors, but also includes contributions from peer reviewers, editors and copy-editors. Language can be considered to perform an active function in developing scientific ideas (Ford and Peat, 1988). That is,

language is not simply used passively to report findings, but is used to influence future directions. In this study, a relatively small number of variations in the language used to describe recommendations was observed, and possibly a different linguistic approach could present the results in a fresh fashion to decision makers.

Higher WTP values appear to lead to a greater proportion of positive research recommendations, although confidence in the recommendations is variable over the whole range. This implies that any threshold lies below the minimum of a typical EVPI–WTP curve. There may also be a relationship between willingness to pay for health benefit, and willingness to invest in research; if an entity considers health benefit more valuable, they may also consider research more valuable. The increased confidence associated with recommendations based on outcomes measured in life-years gained may be due to the unambiguous nature of the measure.

Neoplasms appear to attract a higher rate of positive research recommendations than other disease areas, with authors also exhibiting slightly stronger confidence in those recommendations. This could be a result of societal factors; cancers are overrepresented in the media compared with other diseases (Williamson *et al*, 2011) and societal interest in, and approval for, cancer research may influence authors. In the top ten therapeutic research areas focused on by pharmaceutical companies, cancer drugs outweigh other areas by a factor of at least 2.5 (*The Economist*, 2011); cancer research is well supported by multiple funding sources including charitable entities (Eckhouse *et al*, 2008), and authors may be encouraged to make positive research recommendations by the likelihood of receiving research funding.

Funding by industry sponsors is associated with the presentation of more positive cost-effectiveness results, a form of publication bias (Lexchin *et al*, 2003). This study suggests that industry sponsors are more likely to make positive recommendations and be emphatic about their VOI results which may tally well with commercial interests.

As time has progressed, there has been an increase in the rate of positive recommendations. This could be as a result of increasingly widespread use of EVPI methods, with the consequent development of methodology. Potentially, equivocal recommendations may be less likely to attract further funding. Although there is a possibility that researchers naturally have a vested interest in recommending further research, this study does not provide any significant evidence to support this idea.

6.1 Costs of research

Eckermann *et al* (2010, p702) point out that 'To establish a threshold value of EVPI requires consideration of the costs of undertaking research, which, in turn, is dependent on the proposed research type and size. There is no "one size fits all" cost across different study designs and HTAs that EVPI would need to be greater than'. The costs of further research can vary significantly. In 2005, clinical trials cost a total of $24 billion in the US, representing a mean cost of just under £3 million per trial (CMSInfo in Fee, 2007), while in the UK £950 million was spent, at an average of approximately £100 000 per trial (UKCRC, 2006). However, individual studies can vary substantially within these means. MRC grants varied between just over £100 000 and nearly £3 million in February/March 2011 (MRC, 2011). HTA reports published in 2011 had maximum funding of £1.5 million, although substantially larger trial costs are quoted; for example, over £34 million for the ProtecT trial (HTA, 2011).

The costs of running trials are not necessarily equivalent across jurisdictions, with eastern Europe noted as a less costly environment in which to conduct a trial (Babic and Kucerova, 2003). Therefore, it is unlikely that a unified 'trial cost' could be derived, and the marginal costs associated with each potential study need to be considered on a case by case basis; the level of research in which it is efficient to invest is largely an empirical observation based on circumstances for individual technologies (Drummond *et al*, 2005).

The costs of carrying out research were not frequently cited explicitly, and only one paper referred to the excessive costs of running a trial in that particular disease area. Costs that were cited covered a broad range, indicating that there is substantial variability around the estimates of trial costs. However, although not explicitly mentioned, the observed values at which research is typically recommended correspond reasonably well with average costs of running trials; it appears that authors implicitly acknowledge probable trial costs. The region of uncertainty between £250 000 and £2 million, where recommendations were not consistent, very plausibly covers typical trial costs.

6.2 Strengths and limitations

This study represents the first attempt at deriving an empirical 'threshold' value of EVPI. The search strategy was rigorous and thorough in order to identify an illustrative body of work. Owing to the variable quality of the suppliers' boolean logic implementations, some relevant material may have been overlooked; however, this is not likely to alter the broad conclusions, and the study does not include a formal evidence synthesis.

The texts examined covered a range of countries and cost years. EVPI values were converted to a common currency but not to a common cost year, which may have affected the observed threshold. Furthermore, there was a level of subjectivity in the decision of which EVPI value to choose when multiple values were given. The designated language was based on the area of the study, and may not have accurately represented the primary language of the authors.

The classification system derived from the text analysis was essentially subjective. In particular, some studies may have been classed as less confident than they really were; use of modifying adjectival language was observed but not used to adjust the classifications. The 'valuable' group is less well defined than others, and alternative interpretations might consider 'probably' more emphatic than 'justified'.

Factors that might affect the recommendations made were each analysed independently, with no allowance made for interaction. In particular, the study area and language are not independent of one another; there are likely to be other examples.

6.3 Future for EVPI calculations

Although other VOI measures may be hard to explain and justify, EVPI values have a straightforward interpretation that could aid funding decisions. From a decision-maker's perspective, EVPI calculations can provide a readily available assurance that the research being proposed may be cost-effective. EVPI values are comparable despite different circumstances; the WTP threshold and time frame for usefulness, for example, do not need to be identical for the same interpretation to be drawn. However, whether EVPI calculations become used more frequently depends on a number of factors, including the willingness of investigators to engage in performing calculations; only through greater usage will their value be truly determined.

7. Conclusions

A threshold EVPI value below which research is typically not recommended exists empirically at around £250 000, whilst research is typically not recommended against above a threshold of approximately £2 million.

Factors including study location, funder and disease group appear to affect both the confidence and nature of the recommendation, while language, willingness-to-pay threshold

and publication year affect the nature of the recommendation. The outcome measure affects the confidence with which the recommendation is made.

Although the study does not achieve a statistically significant outcome, it represents a first step towards clarifying decision rules made on the basis of EVPI calculations and may lead to increased confidence in applying knowledge gained from EVPI estimates to decision-making rules.

Funding bodies should be encouraged to require evidence that proposed research should not be rejected on cost-effectiveness grounds through a requirement for EVPI calculations to be presented at grant application stage.

8. Open questions

- Are there any improvements that could be made to the confidence classification? For example, the description ‘probably’ is used here to mean both probably and probably not – does this confuse the issue?
- Are the groupings in figures 1-3 useful or overdone?
- The breakdown graphs are useful for visualising effects; however, further work is required to place the results on a firm statistical basis – what is the best way to proceed, taking into account the variability of the cost year?
- Is there any value in exploring the reasons that people fail to publish EVPI values?

9. References

Ades A E, Lu G and Claxton K (2004) 'Expected value of sample information calculations in medical decision modeling' *Medical Decision Making* **24**(2) 207-227

Babic D and Kucerova I (2003) 'Benchmarking clinical trials practices in Central and Eastern Europe' *Applied Clinical Trials* **12**(5) 56-59

Bank of England (2011) http://www.bankofengland.co.uk/mfsd/iadb/Rates.asp?into=GBP (accessed 24/11/11)

Briggs A H, Claxton K and Sculpher M J (2006) *Decision modelling for health economic evaluation* (Oxford University Press: Oxford)

Bryman A (2008) *Social Research Methods* 3rd edn (Oxford University Press: Oxford)

Claxton K (1999) 'The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies' *Journal of Health Economics* **18**(3) 341-364

Claxton K and Posnett J (1996) 'An economic approach to clinical trial design and research priority-setting' *Health Economics* **5**(6) 513-524

Claxton K P and Sculpher M J (2006) 'Using value of information analysis to prioritise health research: some lessons from recent UK experience' *PharmacoEconomics* **24**(11) 1055-1068

Dixon-Woods M (2001) 'Writing wrongs? An analysis of published discourses about the use of patient information leaflets' *Social Science & Medicine* **52**(9) 1417-1432

Drummond M F, Sculpher M J and Torrance G W (2005) *Methods for the economic evaluation of health care programmes* (Oxford University Press: Oxford)

Eckermann S and Willan A R (2007) 'Expected value of information and decision making in HTA' *Health Economics* **16**(2) 195-209

Eckermann S, Karnon J and Willan A R (2010) 'The value of value of information: best informing research design and prioritization using current methods' *PharmacoEconomics* **28**(9) 699-709

Eckhouse S, Lewison G and Sullivan R (2008) 'Trends in the global funding and activity of cancer research' *Molecular Oncology* **2**(1) 20-32

Eddama O, Petrou S, Regier D, Norrie J, MacLennan G, Mackenzie F and Norman J E (2010) 'Study of progesterone for the prevention of preterm birth in twins (STOPPIT): findings from a trial-based cost-effectiveness analysis' *International Journal of Technology Assessment in Health Care* **26**(2) 141-148

Fee R (2007) 'The Cost of Clinical Trials' http://www.dddmag.com/the-cost-of-clinical-trials.aspx (accessed 7/1/12)

Ford A and Peat F D (1988) 'The role of language in science' *Foundations of Physics* **18**(12) 1233-1242

Groot Koerkamp B, Nikken J J, Oei E H, Stijnen T, Ginai A Z and Hunink M G M (2008) 'Value of information analysis used to determine the necessity of additional research: MR imaging in acute knee trauma as an example' *Radiology* **246**(2) 420-425

Henriksson M, Lundgren F and Carlsson P (2006) 'Informing the efficient use of health care and health care research resources - the case of screening for abdominal aortic aneurysm in Sweden' *Health Economics* **15**(12) 1311-1322

HTA (2011) http://www.hta.ac.uk/project/htapubs.asp (accessed 8/1/12)

Lexchin J, Bero L A, Djulbegovic B and Clark O (2003) 'Pharmaceutical industry sponsorship and research outcome and quality: systematic review' *BMJ* **326**(7400) 1167-1170

Martikainen J A, Kivioja A, Hallinen T and Vihinen P (2005) 'Economic evaluation of temozolomide in the treatment of recurrent glioblastoma multiforme' *PharmacoEconomics* **23**(8) 803-815

MRC (2011) 'Research board awards' http://www.mrc.ac.uk/Utilities/Documentrecord/index.htm?d=MRC002629 (accessed 8/1/12)

NICE (2004) 'Guide to the Methods of Technology Appraisal' http://www.nice.org.uk/niceMedia/pdf/TAP_Methods.pdf (accessed 8/1/12)

Philips Z, Claxton K P, Palmer S, Bojke L and Sculpher M J (2006) 'Priority setting for research in health care: An application of value of information analysis to glycoprotein IIb/IIIa antagonists in non-ST elevation acute coronary syndrome' *International Journal of Technology Assessment in Health Care* **22**(3) 379

Potter J and Wetherell M (1987) *Discourse and social psychology: Beyond attitudes and behaviour* (Sage Publications Ltd: London)

Rao C, Haycock A, Zacharakis E, Krasopoulos G, Yakoub D, Protopapas A, Darzi A, Hanna G B and Athanasiou T (2009) 'Economic analysis of esophageal stenting for management of malignant dysphagia' *Diseases of the Esophagus* **22**(4) 337-347

Rogowski W, Burch J, Palmer S, Craigs C, Golder S and Woolacott N (2009) 'The effect of different treatment durations of clopidogrel in patients with non-ST-segment elevation acute coronary syndromes: a systematic review and value of information analysis' *Health Technology Assessment* **13**(31) 1-102

Schlaifer R (1959) *Probability and statistics for business decisions: an introduction to managerial economics under uncertainty* (McGraw–Hill: New York)

Smith K J, Ness R B, Wiesenfeld H C and Roberts M S (2007) 'Cost-effectiveness of alternative outpatient pelvic inflammatory disease treatment strategies' *Sexually Transmitted Diseases* **34**(12) 960-966

Speight P M, Palmer S, Moles D R, Downer M C, Smith D H, Henriksson M and Augustovski F (2006) 'The cost-effectiveness of screening for oral cancer in primary care' *Health Technology Assessment* **10**(14)

The Economist (2011) 'The costly war on cancer' http://www.economist.com/node/18743951 (accessed 13/1/12)

Traynor M (2006) 'Discourse analysis: theoretical and historical overview and review of papers in the Journal of Advanced Nursing 1996–2004' *Journal of Advanced Nursing* **54**(1) 62-72

UKCRC (2006) 'UK Health Research Analysis' http://www.ukcrc.org/researchcoordination/healthresearchanalysis/ukanalysis/ (accessed 8/1/12)

Vickers A, Goyal N, Harland R and Rees R (1998) 'Do certain countries produce only positive results? A systematic review of controlled trials' *Controlled Clinical Trials* **19**(2) 159-166

Williamson J M L, Skinner C I and Hocken D B (2011) 'Death and illness as depicted in the media' *International Journal of Clinical Practice* **65**(5) 547-551

Acknowledgments

This work was part funded by the MRC Network of Hubs for Trials Methodology, and formed the basis of an MSc dissertation for Health Economics and Health Policy at the University of Birmingham for J Thorn.

Table 1. Characteristics of the 86 included studies.

Characteristic	Number (%)	Characteristic	Number (%)
Funder		Currency	
Government	53 (61.6)	Sterling	48 (53.5)
Industry-related	17 (19.8)	US$	16 (18.6)
Academic	14 (16.3)	Euros	13 (15.1)
Charity	2 (2.3	Can$	6 (7.0)
Disease group		Other	3 (3.5)
Circulatory system	20 (23.3)	Location	
Neoplasms	19 (22.1)	UK	46 (53.5)
Musculoskeletal	8 (9.3)	US	13 (15.1)
Genitourinary	5 (5.8)	Netherlands	12 (14.0)
Other	34 (39.5)	Canada	6 (7.0)
Outcome measure		Other	9 (10.5)
QALYs	74 (86.0)	Publication date	
Life-years gained	7 (8.1)	2000–2005	12 (14.0)
Other	5 (5.8)	2006–2010	74 (86.0)

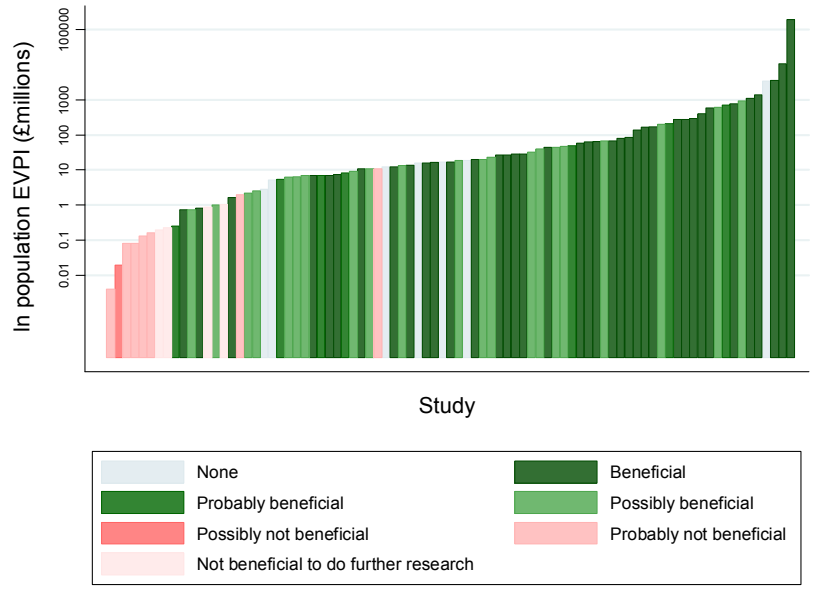

Figure 1. Recommendations for further research by EVPI value.

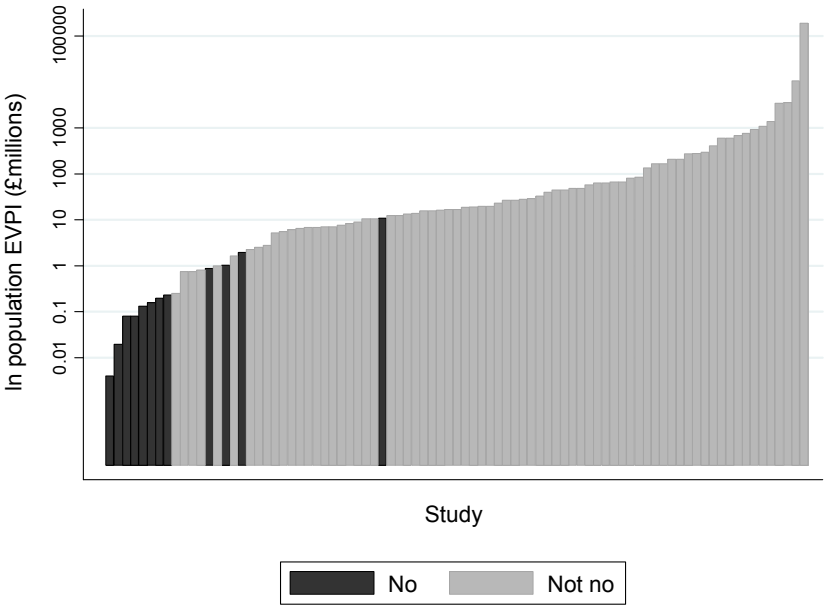

Figure 2. Recommendations for further research collapsed into binary 'no further research' versus any other recommendation.

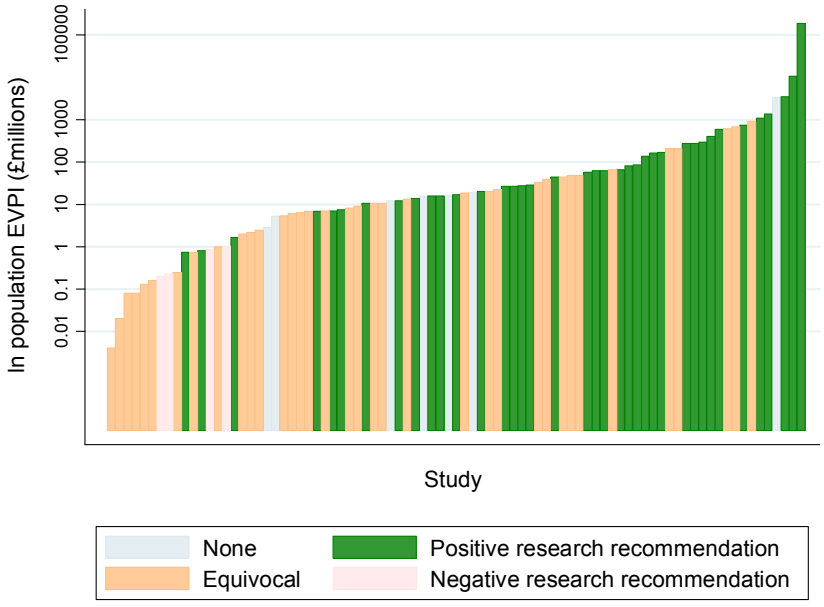

Figure 3. Three-level breakdown of research recommendations.

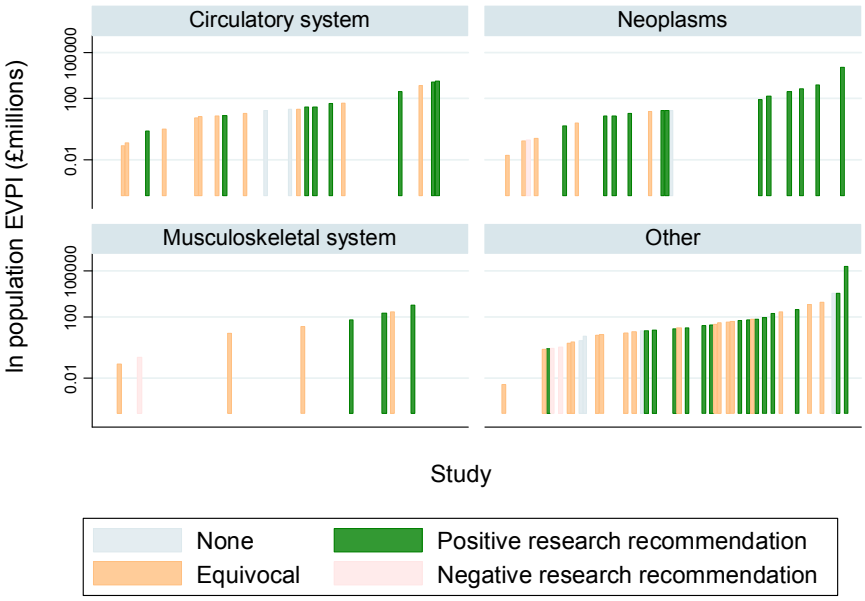

Figure 4. Breakdown of research recommendations by disease area.

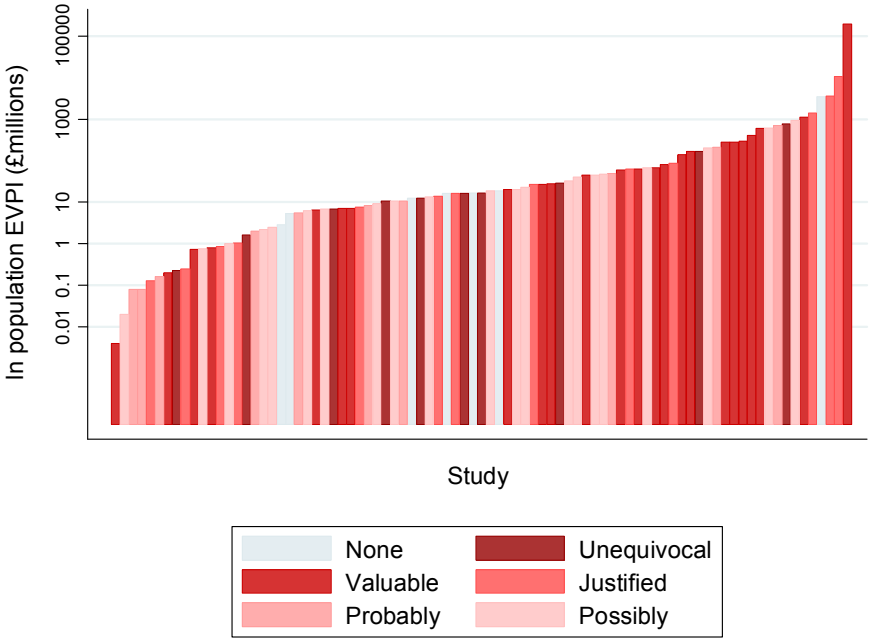

Figure 5. Confidence in recommendations.

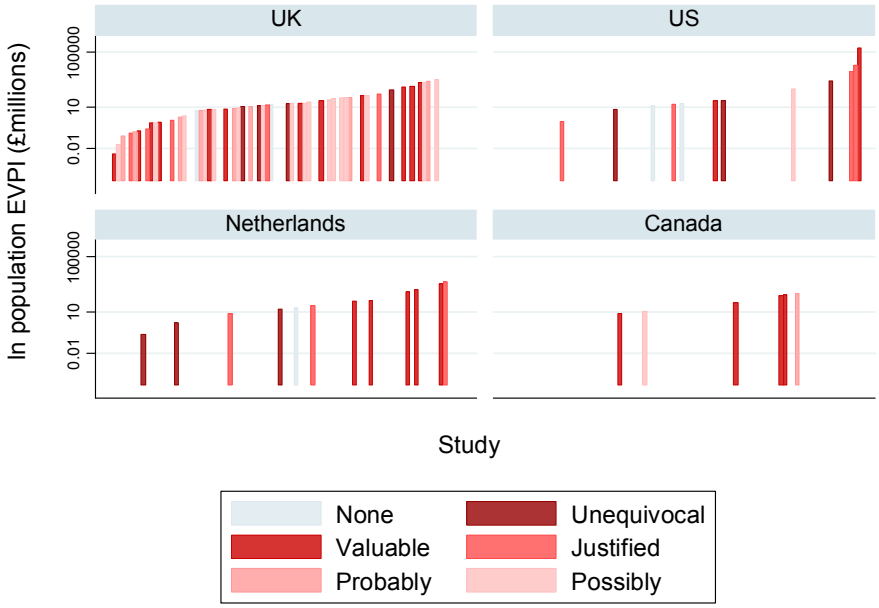

Figure 6. Breakdown of confidence in recommendation by country.