Evidence Working Group Report

Robert Heggie¹, Neil Hawkins², Olivia Wu¹

1. Health Economics and Health Technology Assessment
   Institute of Health and Wellbeing, University of Glasgow

2. Department of Health Services Research and Policy
   London School of Hygiene and Tropical Medicine

November 2015
EXECUTIVE SUMMARY

Introduction
The iDSI Reference Case, developed through the Methods for Economic Evaluation Project (MEEP), has set out the core principles to which economic evaluations should adhere to inform policy decisions in low and middle income countries (LMICs). While these core principles should generally remain universal, the methodological approaches may justifiably differ across decision context. In particular, analysts must consider the processes used to identify all relevant evidence, and critically assess the quality and relevance of these evidence used to inform their decision problem, as this will ultimately determine the level of confidence that policymakers can have in their policy decisions.

Aim
The Evidence Working Group seeks to develop an analytical framework which can be used to assess the quality and relevance of any evidence – clinical and cost evidence, that are used in economic evaluations within a health technology assessment (HTA) in LMICs.

Methods
Our approach consisted of three main stages: (1) a scoping literature review of tools and guidance used to assess evidence used in economic evaluations; (2) a survey of analysts and policymakers involved in HTA in LMICs; and (3) the development of a framework for assessing evidence.

The scoping review involved searching relevant databases to identify and summarise the available literature on relevant assessment tools. The results of our review formed the basis of a survey. A questionnaire accompanied by a case study was sent to analysts and policymakers currently working on HTA in LMICs. The case study presented a typical economic evaluation in a LMIC setting, where data were sourced from a variety of sources. This survey was used to gain insight into how practising analysts and policymakers decide when, and when not, to include certain forms of evidence in their analysis. Subsequently, the results of both the scoping review and survey were used to form the basis of an analytical framework which can be used to assess the relevance and validity of any evidence to be used in an economic evaluation.

Results
The literature broadly described two types of tools available for the evaluation of the quality of evidence – checklists and rating systems. Checklists simply provide a list of inclusion criteria by
which you can ensure all relevant issues have been considered. Rating systems go beyond checklists by grading individual aspects of the evidence and providing the user with an overall rating as to the quality of their evidence. However, one limitation which both these methods share is that they both simply provide the user with an abstract measure of the “quality” of their evidence; they do not tend to inform the user of the usefulness of the evidence. In practice, users will always be faced with imperfect information and be required to make a decision as to the relevance and appropriateness of this to their own decision problem. Hence, a gap exists in the literature for a tool which provides users with, not simply some abstract quality rating, but also clear assessment of the relevance of the evidence to the decision problem at hand.

They survey response rate was poor. Nonetheless, of the five responses received, there was good agreement among respondents on all the issues addressed. They found the case study to be an interesting and a valid contribution to the existing literature base. The respondents generally considered the clinical data from the broader international literature to be reliable, but were concerned with the lack of local cost data. While the respondents recognised that demanding local data for all components of the economic evaluation would not be feasible, none felt they could make a policy recommendation based on the case study alone. Overall, the survey has highlighted the importance of external validity to analysts and decision-makers, when considering evidence for decision-making.

Our literature review and survey formed the basis of our framework for assessing evidence. Our framework deviates from the standard practice in that we do not take a prescriptive approach to assessing evidence. It can be tempting to take the view that as long as the evidence informing economic evaluations is of high quality, then the final estimates of cost-effectiveness estimates will be reliable. It is further assumed that if one or more items of evidence used are of low quality, then you cannot rely on your final cost-effectiveness estimate as the basis for your decision problem. We take the view that it is the overall decision problem that is important and, given this, the real issue is to determine whether there is significant decision uncertainty, which items of evidence contribute to this uncertainty, and whether further research is feasible. It is plausible that there can be uncertainty in model inputs but little uncertainty in the final decision. Hence, we propose a framework in which sensitivity analysis is used to “work backwards” through the model to determine where the uncertainty exists and to what extent this is a problem. Given this, decision-makers can decide how to proceed – whether further research is required, or whether a decision can be made based on available evidence and policymakers’ judgement.
Conclusion

This report explored the problem of how to assess the quality and relevance of evidence used in economic evaluations. In assessing evidence, the majority of the existing tools have been designed to assess quality – i.e. internal validity of the studies. While high-quality economic evaluations require top-quality and unbiased evidence, the relevance – external validity of the included evidence are also highly important. Information will always be imperfect, but decisions need to be made upon the best evidence available. Therefore, we propose a framework which takes a less prescriptive and more decision-centered approach to evidence assessment. We stress the importance of using sensitivity analysis to understand the sources of uncertainly and then using judgement to decide whether or not this level of uncertainty is acceptable and, given this, what the subsequent steps should be adopted.
GLOSSARY

Bias
Systematic deviation of the results of a study from the 'true' results, which is caused by the way the study is designed or conducted.

Evidence
Information on which a decision or guidance is based. Evidence can be obtained from a range of sources, including randomised controlled trials, observational studies and expert opinion (of healthcare and other professionals and/or patients).

Generalisability
Evidence is considered generalisable if it can apply, without adjustment, to other populations or settings.

Transferability
Evidence is considered transferable if it can be adapted to apply in other settings.

Validity
Whether a test or study actually measures what it aims to measure. Internal validity shows whether study or test is appropriate for the question posed. External validity shows whether findings can be applied to the relevant populations of interest.
TABLE OF CONTENTS

1. INTRODUCTION ........................................................................................................7
2. AIM ..........................................................................................................................8
3. METHODS ................................................................................................................8
   3.1. Scoping Literature Review ..................................................................................8
   3.2. Survey of HTA Researchers and Decision-makers .............................................9
   3.3. Framework Development .................................................................................10
4. RESULTS ................................................................................................................10
   4.1. Scoping Literature Review – Clinical Evidence .................................................10
   4.2. Scoping Literature Review – Cost Evidence ......................................................15
   4.3. Survey of HTA Researchers and Decision-makers .............................................19
   4.4. Framework for Assessing Evidence ..................................................................22
5. Discussions ............................................................................................................26
6. References ..............................................................................................................28

Table 1 Comparison of Tools for Evaluating Study Quality ...........................................32
Table 2 Survey Responses ..........................................................................................34
Figure 1 Process for Sensitivity Analysis ....................................................................36
Appendix I ..................................................................................................................37
1. INTRODUCTION

This report was produced as part of the Research arm of the International Decisions Support Initiative (iDSI), in a collaboration between the University of Glasgow, the London School of Hygiene and Tropical Medicine, Erasmus University, Rotterdam, and the Centre for Health Economics, University of York.

In the first phase of iDSI (2014-2016), Methods Working Groups were established to investigate three areas of economic evaluation for which additional methods guidance for analysts and policymakers in low- and middle-income countries (LMICs) was deemed particularly valuable. These linked directly to specific Principles for the practice of economic evaluation detailed in the iDSI Reference Case (1). The three research areas were identified by policymakers in LMICs, as being particularly challenging and requiring additional methods guidance in order to support the realisation of their corresponding Reference Case Principles.

This report details the findings from the investigation into assessing relevant evidence for economic evaluation and decision-making. This is linked to Principle 3: “An economic evaluation should consider all available evidence relevant to the decision problem”. Additional information about the findings from the Methods Working Group can be viewed at: www.idsihealth.org/knowledge_base.

In the assessment of evidence that is used to inform the economic evaluations, taking into account the local context is critical. The issue of determining relevance is of utmost importance, as it will determine which items of evidence are, and are not, considered; the resources and technical skills required for an evaluation; and whether individual evaluations themselves are deemed “sound”. Overall, health technology assessment (HTA) practitioners must critically assess the quality and the relevance of all evidence used, and the processes used to identify that evidence, to inform their decision problem, as this will ultimately determine the level of confidence that policymakers can have in their policy decisions.
2. AIM

The Evidence Working Group seeks to develop an analytical framework which can be used to assess the quality and relevance of any evidence – clinical and cost evidence, that are used in economic evaluations in LMICs. Specifically, the Working Group seeks to develop an analytical framework which can be used to assess:

- the quality of the evidence – internal validity and
- the relevance of the evidence to the decision problem – external validity,

of evidence on clinical effectiveness and costs that are used in economic evaluations for HTA in LMICs.

3. METHODS

Through scoping literature reviews and survey of researchers and decision-makers, the Working Group reviewed existing approaches to evaluating evidence, explored the types of data sources that are appropriate for such economic evaluations, and developed a framework to assess such evidence.

3.1. Scoping Literature Review

Two scoping literature reviews were undertaken: (i) to identify existing tools for assessing evidence, with particular focus on those used to generate clinical and health policy guidance; and (ii) to identify sources of cost data relevant for economic evaluations in LMICS. In addition to the peer-reviewed and electronically indexed literature base, we also explored grey literature such as documents and protocols from evidence-based agencies (e.g. the World Health Organisation (WHO), Guidelines International Network (GIN)) that have a role in assessing evidence to generate guidance in LMICS.

Consequently, two separate searches were undertaken based on the pearl growing approach (2):

1. Tools for assessing clinical evidence – the basic search strategy consisted of Google Scholar search for “tools and methods for assessing evidence” and “assessing evidence”; Medline and Embase for “assessing evidence”. The main organisations for developing
and accessing guidelines were identified and, from this, citations were used to find other relevant papers.

ii. Identifying sources of cost data – the basic search strategy consisted of Google Scholar search for a combination of the following terms: costs, cost source, effectiveness, low-income/developing country. Medline and Embase were searched for "cost evidence".

3.2. Survey of HTA Researchers and Decision-makers

Typically, multiple sources of evidence are required to inform an economic evaluation. In order to explore how these individual sources of evidence are currently assessed, a two-page questionnaire with closed and open questions was developed (Appendix I). Based on the findings of the scoping review, the questionnaire explored the current practice on evaluating evidence for economic evaluations. In particular, the criteria for internal and external validity, and the acceptability of uncertainty were explored amongst the responders.

A survey of researchers and decision-makers in LMICS was conducted by email. Efforts were made to identify a representative sample of local decision-makers through our network of colleagues in HTA (e.g. members of iDSI Reference Case development group), evidence-based medicine (e.g. guideline developers) and public health and health policy (e.g. colleagues in Glasgow, the Scottish Government, WHO and the London School of Hygiene and Tropical Medicine (LSHTM)), who are active collaborators with LMICs.

In order to encourage consistency in interpreting the survey questions, all participants were also sent a case study alongside the survey. A published cost-effectiveness analysis of preventative interventions for cardiovascular disease in Tanzania (3). The study represents a typical methodological approach in the evaluation of cost-effectiveness – clinical evidence was sourced from international literature and cost data were adjusted to reflect local setting. For instance, treatment effects were based on meta-analysis of international literature and the Framingham risk score was used for risk stratification. In terms of cost data, drugs costs were obtained from international price guide and calibrated to the Tanzanian setting; facility costs were obtained from a bespoke costing study of the Tanzanian healthcare system; and capital costs and recurrent costs (such as health check-ups) were taken from published literature and local health centers in Tanzania. All the data were synthesized within a decision analytical framework to estimate potential cost-effectiveness of the individual interventions.
3.3. Framework Development

The results of the scoping literature and the survey formed the basis for the development of a framework for assessing evidence in context. The first draft of the framework was presented at an internal meeting and an international HTA conference. The final framework took into account the feedback and was revised accordingly.

4. RESULTS

4.1. Scoping Literature Review – Clinical Evidence

The scoping literature identified a number of rating systems, checklists and reporting guidance intended to help users evaluate the quality and generalisability of the evidence used to evaluate healthcare interventions (Table 1). In general, these tools consist of a list of criteria for assessment that are specific to individual study designs. They share a common goal of assessing internal validity – to determine whether the study under assessment has answered the research questions being posed in an unbiased manner. As a result, the quality or the trustworthiness of the study findings may be qualitatively and/or quantitatively described.

However, a small proportion of the tools identified in the review go beyond the assessment of internal validity. These tools also include criteria for assessing transferability and generalisability of the study findings. These criteria seek to assess external validity – to determine whether the study population under assessment is generalisable to the wider real world population (population validity); and whether the conduct of the study is generalisable to the real world setting (ecological validity). This is particularly relevant when the findings of the individual studies are being used to inform a clinical or policy decision in practice.

Randomised Controlled Trials

Randomised controlled trials (RCTs) are the most common source of information for informing clinical effectiveness within an economic evaluation. The CONsolidated Standards of Reporting Trials (CONSORT) Statement that was introduced to 1996, is currently the gold standard on the minimum requirements for reporting RCTs (4). The statement consists of a 25-item checklist (and a flow diagram) relating to trial design (appropriate methodological approach to minimize
bias) and analysis (appropriate sample size and statistical analysis) – internal validity; and interpretation of findings (limitations, generalizability and interpretation of results) – internal and external validity. More recently, the CONSORT collaboration has developed extensions to the CONSORT statement for more specific RCT designs such as cluster trials, non-inferiority trials, pragmatic trials, and non-pharmacological interventions.

One other RCT assessment tool was also found to address both internal and external validity when considering RCTs. The Critical Appraisal Skills Programme (CASP) provides support for individuals who are using research evidence in their everyday practice – professional and personal decision-making and policy and guidelines development (5). A series of CASP tools have been developed for specific study types. These tools consist of a series of questions, guiding the user through an assessment of interval and external validity. For instance, the CASP RCT tool addresses three broad issues: (i) internal validity of the results; (ii) the results of the RCT; and (iii) whether the results “help locally”.

Other RCT assessment tools include the JADAD score and the Cochrane risk of bias tool, which are designed to assess methodological quality of studies (internal validity) (6). The JADAD score is a simple-to-use tool, specifically addressing randomization, blinding, and withdrawals and drop-outs; quality are expressed as a score between 0 and 5. However, it has been criticised for being flawed due to the failure of accounting for other essential elements of good RCTs such as adequate allocation concealment and adopting the intention-to-treat principal. The Cochrane risk of bias tool is more comprehensive in capturing relevant methodological criteria. The tool specifically addresses selection bias, performance bias, detection bias, attrition bias, reporting bias and other important biases. Individual RCTs are assessed to these six criteria and results are expressed qualitatively as low, medium or high risk of bias.

**Observational Studies**

Although RCTs are typically the preferred source of clinical evidence, with regard to treatment effects, the incorporation of observational data within an economic evaluation is commonplace. For instance, RCT data are rarely available for long-term outcomes which are relevant to economic evaluations; or there may be health interventions that cannot be or unlikely to be evaluated in an RCT setting. Similar to CONSORT, the STrengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement is a recognised good practice guide for reporting observational studies (7). This 22-item checklist consists of criteria relevant to both internal and external study validity. In particular, item number 22 relates to generalisability – “discuss the generalisability (external validity) of the study results”. Similarly to
RCTs, CASP appraisal tools are also available for cohort and case-control studies. We found one set of guidance that had been specifically developed to assess “real world” data in the context of health policy decision-making. This was produced by the Task Force on “Real World” Data for the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) (8). Although not in the format of a specific checklist or scoring system, this report highlighted eight key issues for considering observational data: importance; limitations; context; guidance on collecting and reporting data; guidance on using such data in decision-making; costs and benefits of collecting such data; need for modelling; and need for continued stakeholder engagement.

Assessment tools that focus solely on methodological quality (internal validity) include: the Newcastle-Ottawa scales for cohort and case-control studies (9); the NICE assessment tools for cohort and case-control studies (10); the Cochrane risk of bias tool for non-randomised studies of interventions (ACROBAT-NRSI) (11); and the Good ReseArch for Comparative Effectiveness (GRACE) checklist (12). These tools are similar, in terms of assessment criteria, e.g. selection bias, exposure bias, reporting bias, and adjustment for confounding. The overall methodological quality can be described quantitatively (score) or qualitatively (level of risk).

**Multiple and Mixed Studies**

A single source of information is rarely sufficient to inform clinical or policy decisions. The synthesis multiple sources of studies, often of varying study design is often required. Approaches to evaluating such efforts are needed.

Historically, when considering multiple sources of evidence (e.g. during clinical guideline development), researchers have implemented hierarchies of evidence – that is, the rating of evidence according to the study design. Typically, systematic reviews and meta-analyses of RCTs are placed at the top, followed by RCTs, then observational studies. However, the supremacy of RCTs in the literature may lead to other forms of evidence being unfairly discarded (13). The design of the study is only one aspect of the quality of the evidence provided. Randomised control trials (RCTs) may not be the most appropriate when considering how an intervention may be transferred to another setting. Furthermore, RCTs are also known to suffer from high internal validity but low(er) external validity – which is of particular concern when considered how this evidence may generalised to other jurisdictions.
The approaches to evaluating of multiple sources of evidence have evolved over the years. Currently, the most commonly used rating system to using multiple sources of evidence to inform clinical practice is the Grade of Recommendations Assessment, Development and Evaluation (GRADE) approach (14). This approach provides a system for rating quality of evidence and strength of recommendations that is explicit, comprehensive, transparent, and pragmatic and is widely adopted by organisations worldwide, in particular, in guideline development. Prior to the development of the GRADE approach, the GRADE Working Group conducted a review of existing grading systems (15). The authors note the need for consistency in grading evidence and how the variety of then available, disparate systems, undermines the purpose of having explicit systems for evidence assessments. The study critically appraised six prominent systems for grading evidence against a set of 12 criteria to determine “the sensibility” of the different approaches. All assessors had experience with at least one of the systems and most had helped to develop at least one of them. The findings showed little agreement on the sensibility of the six systems. Only one of the systems was found to be suitable for all of the features under consideration – effectiveness, harm, diagnosis and prognosis. None were appropriate for all of the stakeholders involved – professionals, patients and policy makers. All of the systems suffered from low reproducibility of results. The authors concluded that those grading systems all suffered from their own specific weaknesses.

The GRADE approach was proposed as a means to capture the best features of the current systems, while correcting for any shortcomings. The goal was to create a system which was widely used and disseminated in an attempt to create a “standard approach”. The GRADE process works by posing an explicit question to be answered, along with a specification of all the important outcomes. Once the evidence has been collected and summarized, GRADE assesses the quality of the evidence against an explicit set of criteria, which include: risk of bias, imprecision, inconsistency, indirectness, publication bias, effect size, dose response and control for confounding. Recommendations are based on the quality of evidence for the intervention and the balance between the risk of desirable and undesirable consequences of alternative interventions (none or current best practice). Under the GRADE approach, evidence obtained from a randomised control trial (RCT) would initially be deemed to be of “high quality”, before being evaluated against the aforementioned set of criteria. If these criteria raised no concerns over the quality of the evidence, it would remain “high quality”, alternatively, it may be rated down. All other forms of evidence (most commonly, that obtained from observational studies) are initially deemed to be of “low quality”. This is evaluated against an explicit set of criteria and may be rated up or down accordingly. As such, it is possible that a piece of observational evidence may be regarded as higher quality than evidence obtained from an RCT.
Further, the Method for Evaluating Research and Guideline Evidence (MERGE) sets out an explicit approach to reviewing and incorporating scientific evidence into guidelines (16). MERGE can be used for two main purposes: to assess the quality of evidence in individual studies and to assess the validity of intervention guidelines. In terms of the basic principles, MERGE is similar to the GRADE approach. Evaluation criteria are coded according to the extent to which the criteria are fulfilled. The process can either be applied to a review of studies (systematic reviews, meta-analyses) or individual studies (RCTs, cohort studies, etc.). These codes can then be aggregated to give an overall quality rating for the evidence assessed. One unique feature of MERGE is that it takes into account the relevance of the evidence to the decision question by asking the question – “What is the level of evidence for outcomes central to decision-making?” This encourages the reviewer to focus on the evidence that ‘drives’ the decision, rather than whether it meets the criteria for “ideal” evidence.

In the context of economic evaluations, typically systematic reviews and meta-analyses are carried out to formally synthesise multiple sources of data to be incorporated in an economic model. This review has identified several tools for the assessment of these studies. The Preferred Reporting Items for Systematic Reviews and Meat-Analysis (PRISMA) has set the current gold standard for reporting systematic reviews and meta-analyses (17). This is a 27-item checklist that follows the basic principles set out in CONSORT and STROBE. The checklist consists of criteria that are relevant to the reporting of methodological approach to identifying, summarizing and synthesizing multiple data sources, and addressing within individual study and across multiple studies. However, the PRISMA checklist does not take into account the issue of external validity. Similar to the CONSORT collaboration, PRISMA has also developed extensions to the original checklist. Current extensions include checklists for reporting systematic review of abstracts, equity, individual patient data, network meta-analysis and protocols. The original PRISMA checklist was developed with a focus on synthesizing RCT evidence, some have argued that PRISMA is not appropriate for synthesizing observational data. However, recent extensions to the PRISMA, such as the explanation and elaboration document have provided clear guidance on how PRISMA can be used for other study designs. Nonetheless, this review has identified one guideline for the synthesizing observational evidence produced by the Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group (18). This guideline address broadly similar issues to PRISMA, but does include specific items relevant to observational studies – e.g. assessment of confounding; regression on possible predictors of study results.
The Critical Appraisal Skills Programme (CASP) has also developed a simple tool for evaluating systematic reviews. Similar to the CASP tool for RCT and observational studies, this tool addresses both internal and external validity of systematic reviews. One tool identified in the review, has been developed to focus on external validity of systematic reviews and meta-analyses of intervention studies (19). The External Validity Assessment Tool (EVAT) consists of three core domains – (i) identification of source population in the original studies; (ii) generalizability of the source population; and (iii) setting relevant to the real world setting. Each domain can be graded as ‘well covered’ (++); ‘adequately covered’ (+); ‘poorly addressed’ (-) or not applicable (0). We also found one assessment tool that provides specific guidance for network meta-analysis. The ISPOR Academy of Managed Care Pharmacy (AMCP) and the National Pharmaceutical Council (NPC) Good Practice Task Force produced a set of guidance for the statistical pooling of data when comparing multiple treatment interventions (20). This guidance assesses both internal and external validity of network meta-analyses. The 26 questions consist of criteria relating to: relevance (internal and external validity), credibility (internal validity), analysis (internal validity), reporting quality and transparency (internal validity), interpretation (internal and external validity) and conflicts of interest (internal validity).

4.2. Scoping Literature Review – Cost Evidence

In addition to clinical evidence, economic evaluations also require high-quality and relevant cost evidence to determine cost-effectiveness of interventions. The approaches to collecting cost data and the accessibility of such data vary substantially across different countries. While costs related to healthcare resources that are typically reimbursed by the government or through insurance systems are more likely to be routinely recorded; this will not be the case for costs related to healthcare resources that are paid by the patients themselves. Therefore, acquiring cost evidence in the LMIC-setting, where much of the healthcare resources are paid by the patients themselves, is particularly challenging (21).

The WHO suggests that recommendations regarding cost valuations in economic evaluations, particularly those used in LMICs, need to be applicable in many different settings, not have stringent data requirements and be understandable to the non-specialist (22). The WHO also recommends the use of the “ingredients approach” to valuing resources. This involves reporting on the individual resources, quantities, and the cost applied to each. Contrary to simply reporting total expenditure, this allows analysts to judge whether the resources used and the way in which they have been valued is relevant to their own context.
However, complete and reliable cost data for healthcare resource use are scarce in the LMICs setting. As a result, researchers often use multiple, disparate sources to estimate the costs of healthcare interventions. These generally include: hospital records, academic studies, national health surveys, international development institutions, national reference costs and expert opinion. The scoping literature on cost evidence in the context of LMICs has identified few commonly used sources of cost data.

The WHO-CHOICE Project
The WHO-CHOICE (CHOosing Interventions that are Cost-Effective) project is an initiative developed with the aim of providing policy makers with evidence for deciding on interventions which maximize the use of health resources (23). The WHO-CHOICE project essentially uses a regression model to predict “average” country-specific unit costs for health care services. Cost and effectiveness estimates are available for a wide range of health interventions at the WHO sub-regional level. The WHO-CHOICE also contains unit costs for primary and secondary health care services for each of their 191 member states. Estimates of unit costs are available to download alongside an Excel spreadsheet containing the model. Users are able to vary the parameters of the model (e.g. length of stay) to obtain updated estimates based on any additional information they may have. Costs are available for services such as cost per hospital stay, cost per outpatient visit, etc.

Condition-specific Databases
Historically, economic evaluations of healthcare interventions in LMICs have a tendency to focus on interventions for infectious diseases. Conditions such as HIV/AIDS are the focus on many global institutions (for example, the WHO and the World Bank) and, as such, are the subject of much research in both the developed and developing world. As a result, more efforts have been spent on collecting high quality clinical, resource use and cost data on these diseases. One of these examples is a repository of country-specific vaccine data – Online International Vaccine Economics and Statistics (OLIVES).

OLIVES (24) is a country-specific database which collects data on demography, disease burden, health service utilisation and costs, vaccine coverage, effectiveness and programme costs for 130 LMICs. These data are obtained from national databases and literature reviews (carried-out by the Institute for Clinical Effectiveness and Health Policy (ICES) and the LSHTM). All studies with relevant data retrieved from the literature reviews are graded according to their internal and external validity according to a standardized format. Currently, OLIVES include data on the
human papilloma virus (HPV) vaccine, pneumococcal conjugate vaccine (PCV) and rotavirus vaccine (RV).

Local Data
Developed nations generally have published national guides for medical service fees. For example, in England, unit costs for primary health and social care are published annually by the Personal Social Services Research Unit (PSSRU) at the University of Kent at Canterbury and the London School of Economics and Political Science (25). Costs of secondary healthcare are published annually by: the UK Government in England (NHS Reference costs) (26), and by the Information Services Division (IDS) in Scotland (27).

However, such reliable databases are rarely available in LMICs. Health care costs in these countries are generally obtained in a more “ad hoc” basis. For example, hospital costs are often obtained directly from local or national hospitals (finance departments, annual reports, etc). In the case of in-patient care, some basic services are provided by unpaid “helpers” and food provided only by the family members of the patient. Some cost data for which formal records were unavailable, particularly for those services which are less likely to be recorded (such as out-patient care), were obtained through interviews with local physicians.

One guideline provides a clear step-by-step approach for conducting disease-specific costing studies in LMICs. The guideline made recommendations on methods and data requirements for:

i. Definition of the study perspective – the relevant cost items to be included is dependent on the study perspective. For instance, from a policymaker’s perspective, the inclusion of productivity losses is necessary to assess societal impact. Whereas, adopting the perspective of a healthcare provider at a local setting, costs associated with productivity loss is are less relevant. One thing to note, the patient perspective is important when undertaking economic evaluations in LMIC, where patients often incur significant personal costs.

ii. Characterisation of the unit of analysis – the unit of analysis is dependent on the decision problem. For instance, if the purpose of the study were to estimate the cost incurred within a hospital, or hospital department, then this would be the unit of analysis. Alternatively, if the purpose of the study was to estimate the potential pay-out of an insurer for a specific disease, then the unit of analysis would be the disease (e.g. no. of cases).

iii. Identification of cost items – each health service activity will consist of resource use and individual cost items (e.g. staff, equipment, drugs). Depending on data availability, different methods may be chosen to estimate activity costs – “gross-costing” and “micro-costing”.
Gross-costing is less data-intensive but provides a limited level of detail; micro-costing is more data intensive but provides the user with a higher level of detail and accuracy.

iv. Measurement of cost items – this step involves the measurement of the volume of utilisation of items to be costed. Similar to identification of costs, the measurement of costs involves a trade-off between accurate, patient-level cost estimates and data availability. This requires a trade-off between what giving consistent guidelines and ensure that your these guidelines are transferable to other settings. The recommendation is to make best use of the data available locally and to adjust their methodology as their environment requires. The availability of utilisation data will differ between health service items and data will often need be to obtained from multiple sources (patient registers and records, expert opinion, laboratory and pharmacy registers, insurance databases, databases of ongoing studies, etc.). In the absence of utilisation data, a data collection mechanism will need to be setup to obtain this.

v. Valuation of cost items – two approaches are available for valuing cost items: “top-down costing” and “bottom-up costing”. Top-down costing may, for example, involve obtaining the total cost of a hospital department per year alongside the number of patients treated and using this to estimate the cost per patient. This less data-intensive but provides a limited level of detail. Alternatively, bottom-up costing would involve identifying each service item and applying unit costs to these. This approach is more data intensive but provides the user with a higher level of detail and accuracy.

vi. Uncertainty analysis – characterizing uncertainty is important in estimating costs. Cost parameters should be subject to thorough sensitivity analysis. This may involve univariate or multivariate sensitivity analysis or probabilistic sensitivity analysis.

Our scoping review did not find any assessment tools that were designed specifically to evaluate the quality of cost evidence. However, the GRADE approach discussed in the previous section (section 4.1) can be extended to assess the quality of evidence used in economic evaluations. Specifically, this can also be used to assess evidence on resource use within an economic evaluation (28). The assessment relates to: (i) identification of relevant resource use items that is expected to vary between alternative interventions; (ii) measure of resource use in each intervention and estimates for the difference between alternative interventions; (iii) rating of the confidence these estimates; and (iv) valuation of the resource use for the specific setting in which the evaluation is expected to take place.
4.3. Survey of HTA Researchers and Decision-makers

The survey was sent to 302 currently practicing HTA analysts and/or policymakers involved in LMICs by email. The response rate was low; of those who were surveyed, we received five completed questionnaires. All respondents were academic health economists who are currently working in the UK, in Canada and Switzerland, and in Thailand.

The responses to the closed questions are shown in Table 2. Overall, there was good agreement among respondents on all the issues addressed. They found the case study to be an interesting and a valid contribution to the existing literature base. The respondents generally considered the clinical data from the broader international literature to be reliable, but were concerned with the lack of local cost data. While the respondents recognize that demanding local data for all components of the economic evaluation would not be feasible, none of them felt they could make a policy recommendation based on the case study alone.

The responses to the open questions are summarized qualitatively according to the key themes below:

• Importance of quality, reliability, and relevance of the evidence used to the decision problem

It is not always feasible, or even desirable, to wait until perfect evidence is available to carry out an economic evaluation. However, it is not simply a question of obtaining reliable evidence and cost-effectiveness estimates, the analyst must ensure that the problem their analysis seeks to answer is relevant to the decision maker. For the Tanzanian case study, the respondents felt more evidence specific to the region is necessary. This includes sound local cost data, information about the risk structure of the population and a budget impact assessment. One of the respondents voiced that although the case study was interesting, it was designed in such a way that it is not likely to be of direct relevance to the decision problem faced by policy makers.

“I find this paper interesting in terms of raising the question that this area is important and may be potentially cost-effective (and has not been the focus of global attention in LMICs). I would not find it sufficient if I were advising the MoH in Tanzania to go ahead and spend the large amount of resources required to strengthen these services.”

• Importance of generalizability of findings
Although study design will always be a key factor in determining the level of confidence you can place in an RCT study, of particular concern in this case is the extent to which this RCT data is generalizable to another setting. Issues such as the representativeness of the population, and a consideration of the constraints which would be faced in the real world should be considered. This was reflected in comments from respondents.

“*My confidence is less to do with the design and evidence quality – but more to do with generalizability.*”

“*Increase (my confidence): population or sub-population of RCT is comparable to study population; intervention and comparator of RCT are comparable to study setting; outcome of RCT comparable to study; results of RCT are statistically significant and clinically relevant.*”

- **A pragmatic approach to considering cost data.**

All respondents were in agreement that there is a need to be pragmatic when it comes to cost data. The respondents acknowledged that there is no simple hierarchy to observe, rather the analyst must think what would be appropriate to their context and be transparent about any limitations of their evidence base and any trade-offs which will have to be made. Possible sources of local cost evidence would include hospital databases and vaccine/drug database. This data should then be validated with local health care practitioners to ensure they meet face validity.

“(Cost data can be obtained)... by collecting the data (assuming not available through routine systems), using local researchers and strong links to those in policy to get face validity.”

- **Valid contribution to literature, but not sufficient for policy decision-making.**

One of the respondents felt the study might have been overly ambitious in scope. Cardiovascular disease is a complicated area including multiple conditions; the definition used in the study may have been too broad. Furthermore, one of the respondents felt there was a lack of information on the profiling of the high-risk population. A common concern was that there was insufficient practical information on implementation and policy-orientated information to enable a policy maker to decide on whether to fund an intervention or not. The lack of information on how the interventions could be incorporated into the Tanzanian healthcare system was also highlighted. The respondents showed preference for a narrower scope with clearer objectives. One respondent suggested the results to be presented in conjunction with a
more practical consideration reflecting the realities of the intervention and the decision problem faced by policy makers. Finally, all respondents agreed that, due to the uncertainty associated with some of the evidence sources, thorough sensitivity analysis (and expected value of information analysis) would be required.

“The practical conclusions are not there at all. There is nothing about health care financing in Tanzania. Is the recommendation that the Government health sector provide the drugs free of charge? How would that work in practice? How would patients get access to the drugs? I'm aware that there issues are not part of the study question in the paper, but these are issues that need to be addressed for such a study to be helpful for decision making.”

“It is useful but needs to always be contextualized, and used in conjunction with relevant qualitative data that caters for the realities and values of decision making.”
4.4. Framework for Assessing Evidence

In general, there is a broad agreement that ‘relevant’ evidence to the decision problem should be incorporated in an economic evaluation. This was echoed by the views of the respondents of the survey (Section 4.3). Having identified a candidate set of evidence we need to determine which items of evidence should be included in our analysis. In general, evidence will be selected that minimises the bias in the final estimates. A number of tools have been developed to assess study ‘quality’, effectively the risk of bias (Section 4.2). The majority of these tools focus on the internal validity. There are few exceptions and focuses on the assessment of the external validity of a study. This perhaps because it is relatively straightforward to define objective criteria related to study design and conduct to assess internal validity. Whereas external validity is context dependent and judgement is required to evaluate the potential impact differences between study and clinical populations. We should be wary of over emphasizing internal validity and ignoring issues of external validity and precision when selecting evidence. As John Tukey put it: “Far better an approximate answer to the right question, which is often vague, than an exact answer to the wrong question, which can always be made precise”.

The objectives of maximising precision and minimising bias may conflict. In general, studies are selected that minimise bias in the final estimates. For example, we may want to minimise bias by limiting our consideration to local studies, or to randomised trials. But this may reduce the number of studies we can use reducing precision. In some circumstances, particularly when the likely direction is bias is known, a precise bias estimate may be preferred over an imprecise unbiased estimate. For example, when estimating cost-effectiveness for infectious diseases a biased, but more precise, static estimate may be preferred over the less biased dynamic estimate when the intervention is cost-effective according to the static analysis and any effects on transmission are believed to improve cost-effectiveness [WHO Guidelines].

It is tempting to assess the validity of a cost-effectiveness analysis based on an assessment of the quality of the individual items of evidence used in the analysis. If all of the items of evidence are of high quality, and judged to be unbiased, the final estimates of cost-effectiveness should be unbiased, and decisions based on the analysis reliable. However, we should be wary of “denying the antecedent”; if one of the items of evidence is of low quality, and possibly biased, it is not axiomatic that the final decision is unreliable. For example, it is possible that the decision is simply not sensitive to the uncertainty in a particular parameter. At the extreme, we may be
able to make reliable decisions in the absence of any empirical data regarding certain parameter, if the decision does not change when we vary those parameters across their credible ranges.

It is therefore useful to work backwards, using methods of sensitivity analysis, to evaluate the contribution of the uncertainty in individual parameters to the overall decision uncertainty. There are a number of steps in such a sensitivity analysis:

- First, we determine whether there is uncertainty in the decision given the uncertainty in the parameter estimates. The uncertainty of the decision can be evaluated using either deterministic or probabilistic sensitivity analysis (29) (30) (31). In deterministic sensitivity analysis (DSA) individual parameters are varied. The effect of this variation on the ICER and the decision is observed. The results of multiple sensitivity analyses may be summarized as a “tornado” diagram (31). In probabilistic analysis, the effect of a set of parameters varied simultaneously according to random distributions describing the uncertainty in individual parameters (32). Following this analysis, the impact of this variation on the incremental costs and effects can be represented on a cost-effectiveness plane (33), confidence intervals can be estimated for the ICER (34) (35), and the impact on the treatment decision can be illustrated using a cost-effectiveness acceptability curve (CEAC) (36), or a “javelin” diagram (37). Sometimes it is useful to combine deterministic and probabilistic sensitivity analysis to observe the impact on decision uncertainty of varying some parameters deterministically, for example treatment price, while other parameters vary randomly.

Ideally the assessment of uncertainty should consider both bias and precision, rather than focusing solely on precision as represented by estimates of sampling error. It should also include consideration of the uncertainty in the model structure (“structural uncertainty”) (38) (39). Sometimes the uncertainty in a parameter is not well characterized. In this case it may be useful to conduct a deterministic sensitivity analysis varying the parameter across a range representing the highest and lowest values we think the parameter could take. This has been referred to as an “analysis of extremes” (29). Our analysis will then provide an upper bound of the effects of uncertainty in that parameter. If the sensitivity analysis indicates that there is little overall decision uncertainty, decision-makers can make a recommendation. Probabilistic sensitivity analysis is particularly useful in this regard as it captures the joint contribution of uncertainty in multiple parameters study parameters.
• If there is uncertainty in the decision, we first need to determine whether that uncertainty is material. We may be highly uncertain about a decision, but the consequences, in terms of opportunity cost, of making the wrong decision may be inconsequential.

The potential losses associated with decision uncertainty can be evaluated by using deterministic sensitivity analysis to assess the impact on incremental costs and effects on the cost-effectiveness plane (33). Alternatively, estimates of expected value of perfect information (EVPI) can be derived from a probabilistic sensitivity analysis (40). Expected value of perfect information is an estimate of the value of obtaining perfect information on all parameters, essentially removing all uncertainty. The value comes from the reduction in the probability of making “incorrect” decisions and selecting the incorrect treatment. Again if the impact of decision uncertainty is not material, decision-makers can make recommendations based on the current evidence.

• If the impact of parameter uncertainty on decision uncertainty is material, we need to identify which parameters make the greatest contribution to decision uncertainty. Deterministic sensitivity analysis can be used to assess the impact of varying individual parameter estimates on incremental costs and effects. Alternatively, estimates of expected value of partial perfect information (EVPPI) can be derived from a probabilistic sensitivity analysis. The EVPPI is an estimate of the value of obtaining perfect information on individual parameters (41). It provides a simple measure of the contribution of individual parameters to decision uncertainty.

Having identified which parameters make the greatest contribution to decision uncertainty, the next step depends on whether decision-makers are able to make decisions that are contingent on the collection of further information.

a. If decision-makers are able to make contingent decisions, they need to decide whether it is feasible to collect further evidence that would reduce uncertainty. If is not feasible, decision makers should substitute their best judgements for these key parameters and make a recommendation. They may be able to use the results to existing deterministic sensitivity analyses to substitute their judgements or they may need to request re-analysis. Alternatively, they may have access to interactive models that let them conduct the sensitivity analysis themselves (42). If it is feasible to collect further evidence, they need to decide whether to make their recommendations contingent on the collection of this evidence.
b. If decision-makers are not able to make contingent decisions, on the collection of further information or if it is not feasible to collect further information, decision makers should substitute their best judgements for these key parameters and make a recommendation.

This process is illustrated in Figure 1.
5. Discussions

Although developed countries have been active in undertaking economic evaluations for HTA, the transferability of the findings is limited. An ISPOR Good Research Practices Task Force produced a paper on “The Transferability of Economic Evaluations across Jurisdictions” (43). The authors suggested many reasons for HTAs to differ between settings, such as: incidence and severity of disease; the availability of health care resources, clinical practice patterns; and relative prices. They cited findings from a study (Barbieri et al, 2005) which found that in 17 out of the 27 reviewed economic evaluations of pharmaceuticals (conducted in two or more European countries), the variation in the incremental cost-effectiveness ratios estimated could be considered to be substantial (44). Further, in reviewing the advice given in current guidelines, the authors found that 16 out of 21 guidelines recognised the potential for differences in clinical parameters between settings. In contrast, Boehler et al suggested that the variation in published cost-effectiveness estimates is related more to differences in study methods than to differences in national context (45). The authors suggested that Barbieri et al failed to assign variance to differences in study questions or methods, and hence, any observed differences were attributed to differences between countries. The authors noted, however, that they do not mean to suggest that national context is not important, but simply that differences between countries may not be fully reflected in published cost-effectiveness estimates.

There is a need for economic evaluations to be conducted in the LMIC-setting, to support local stakeholders in clinical and policy decision-making. This report explored the problem of how to assess the quality and relevance of evidence used in economic evaluations. In assessing evidence, there are many existing tools available for individual study designs. The majority of these tools have been designed to assess quality – i.e. internal validity of the studies, based on a list of criteria. The main exceptions are GRADE, CASP and EVAT. Given the typical lack of local studies and differences in settings, it may be that external validity is a more important issue than interval validity in LMIC evaluations. The majority of the tools identified in the scoping review are therefore not directly helpful in determining an appropriate course of action. This relates to both clinical and cost evidence, and was highlighted by the respondents of our survey.

In undertaking our review of the current literature, our focus has been to highlight the range of tools available for assessing evidence and to comment on their approach to evidence assessment. We do not explicitly recommend the use of any one tool. Indeed, the usefulness of any particular tool will always depend on the decision problem at hand. Rather, we seek to
highlight any gaps we believe currently exist in the literature and to propose a new, and complementary, approach to evidence assessment.

Information will always be imperfect, but decisions need to be made upon the best evidence available. Therefore, we proposed a framework which takes a less prescriptive and more decision-centered approach to evidence assessment. We stress the importance of using sensitivity analysis to understand the sources of uncertainty and then using judgement to decide whether or not this level of uncertainty is acceptable and, given this, what the subsequent steps should be adopted.

Conclusions

It is insufficient to simply assess the quality of the available evidence. Although "high quality" (unbiased) evidence is more likely to lead to a reliable decision. It may be possible to reach a reliable decision given lower quality evidence, and/or it may be impractical to collect high quality evidence. We have developed a tool that focuses on the contribution of individual items of evidence to decision uncertainty; that considers both internal and external evidence; and considers the feasibility of future research. These assessments may be qualitative or quantitative given the available analytic resources. Some of these assessments are specific to the decision and the decision-maker.

What’s next?

- In order for the proposed framework to be used in practice by stakeholders in LMICs, further development is required to convert the framework into a simple toolkit.
- Further engagement of stakeholders in LMICs is required to ensure the successful development of this toolkit. We have learned from the experience of our survey, that face-to-face interaction such as focus group type set up would be required to ensure a good level of feedback.
6. References


<table>
<thead>
<tr>
<th>Study Type</th>
<th>Tool</th>
<th>Assesses</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Internal Validity</td>
</tr>
<tr>
<td>All</td>
<td>Hierarchy of evidence</td>
<td>✔</td>
</tr>
<tr>
<td></td>
<td>GRADE (Grading of Recommendations Assessment, Development and Evaluation)</td>
<td>✔</td>
</tr>
<tr>
<td>RCTs</td>
<td>ISPOR AMCP Task force on ITC Network Meta-Analysis</td>
<td>✔</td>
</tr>
<tr>
<td></td>
<td>Cochrane Risk of Bias tool for RCTs</td>
<td>✔</td>
</tr>
<tr>
<td></td>
<td>JADAD</td>
<td>✔</td>
</tr>
<tr>
<td></td>
<td>CASP (RCTs)</td>
<td>✔</td>
</tr>
<tr>
<td></td>
<td>CONSORT</td>
<td>✔</td>
</tr>
<tr>
<td>Observational studies</td>
<td>ISPOR AMCP Task force on Real World Evidence</td>
<td>✔</td>
</tr>
<tr>
<td></td>
<td>ACROBAT-NRSI (A Cochrane Risk Of Bias Assessment Tool: for Non-Randomized Studies of Interventions)</td>
<td>✔</td>
</tr>
<tr>
<td></td>
<td>CASP (cohort)</td>
<td>✔</td>
</tr>
<tr>
<td></td>
<td>CASP (case control)</td>
<td>✔</td>
</tr>
<tr>
<td></td>
<td>MOOSE</td>
<td>✔</td>
</tr>
<tr>
<td></td>
<td>STROBE</td>
<td>✔</td>
</tr>
<tr>
<td></td>
<td>GRACE</td>
<td>✔</td>
</tr>
<tr>
<td></td>
<td>Newcastle-Ottawa (case control)</td>
<td>✔</td>
</tr>
<tr>
<td></td>
<td>Newcastle-Ottawa (cohort)</td>
<td>✔</td>
</tr>
<tr>
<td></td>
<td>NICE (cohort)</td>
<td>✔</td>
</tr>
<tr>
<td></td>
<td>NICE (case control)</td>
<td>✔</td>
</tr>
<tr>
<td></td>
<td>MERGE (Method for Reviewing Research Guideline Evidence)</td>
<td>✔</td>
</tr>
<tr>
<td>Systematic review/meta-analysis</td>
<td>CASP (systematic review)</td>
<td>✔</td>
</tr>
<tr>
<td></td>
<td>NICE (systematic review/meta-analysis)</td>
<td>✔</td>
</tr>
<tr>
<td></td>
<td>EVAT (External validity assessment tool)</td>
<td>✔</td>
</tr>
<tr>
<td></td>
<td>PRISMA</td>
<td>✔</td>
</tr>
<tr>
<td>Economic evaluations</td>
<td>NICE: Guide to the methods of technology appraisal 2013</td>
<td>✔</td>
</tr>
<tr>
<td></td>
<td>Canadian Agency for Drugs and</td>
<td>✔</td>
</tr>
<tr>
<td>Study Type</td>
<td>Tool</td>
<td>Assesses</td>
</tr>
<tr>
<td>---------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Internal Validity</td>
</tr>
<tr>
<td></td>
<td>Technologies in Health: Guidelines for the Economic Evaluation of Health Technologies 2006</td>
<td>✔</td>
</tr>
<tr>
<td></td>
<td>Scottish Medicines Consortium: Guidance to Manufacturers for Completion of New Product Assessment Form 2014</td>
<td>✔</td>
</tr>
</tbody>
</table>
The authors explained that the burden of cardiovascular disease in sub-Saharan Africa was substantial. Despite not having comprehensive local data, they felt they had to undertake this study, using the best evidence available they had.

<table>
<thead>
<tr>
<th>Do you agree with this decision?</th>
<th>Yes</th>
<th>No</th>
<th>No Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Respondent 1</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Respondent 2</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Respondent 3</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Respondent 4</td>
<td></td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>- Respondent 5</td>
<td></td>
<td>✔</td>
<td></td>
</tr>
</tbody>
</table>

Do you think that resources should always be directed towards studies for which there are adequate local data available?

| - Respondent 1                   |    | ✔  |             |
| - Respondent 2                   |    | ✔  |             |
| - Respondent 3                   |    | ✔  |             |
| - Respondent 4                   |    | ✔  |             |
| - Respondent 5                   |    | ✔  |             |

In this study, European and US data on treatment effect were used due to a lack of RCT data on CVD in an African setting.

<table>
<thead>
<tr>
<th>Do you think this is reasonable?</th>
<th>Yes</th>
<th>No</th>
<th>No Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Respondent 1</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Respondent 2</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Respondent 3</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Respondent 4</td>
<td></td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>- Respondent 5</td>
<td></td>
<td>✔</td>
<td></td>
</tr>
</tbody>
</table>

Do you think observational data can be used as a data source for treatment effect?

| - Respondent 1                   | ✔   |    |             |
| - Respondent 2                   | ✔   |    |             |
| - Respondent 3                   | ✔   |    |             |
| - Respondent 4                   |    | ✔  |             |
| - Respondent 5                   |    | ✔  |             |

Both costs and DALYs were discounted at 3% in the study.

<table>
<thead>
<tr>
<th>Do you agree with the choice of discount rate?</th>
<th>Yes</th>
<th>No</th>
<th>No Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Respondent 1</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Respondent 2</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Respondent 3</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Respondent 4</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Respondent 5</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Would you require a range of discount rates to be tested?

| - Respondent 1                                | ✔   |    |             |
| - Respondent 2                                | ✔   |    |             |
| - Respondent 3                                | ✔   |    |             |
| - Respondent 4                                | ✔   |    |             |
| - Respondent 5                                | ✔   |    |             |
The study reported incremental cost-effectiveness ratios ranging from USD 85 to USD 4,589.

<table>
<thead>
<tr>
<th>Question</th>
<th>Yes</th>
<th>No</th>
<th>No Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Would you be satisfied with a decision to adopt treatment based on the findings of this study?</td>
<td></td>
<td></td>
<td>✔</td>
</tr>
<tr>
<td>- Respondent 1</td>
<td>✔</td>
<td></td>
<td>✔</td>
</tr>
<tr>
<td>- Respondent 2</td>
<td>✔</td>
<td></td>
<td>✔</td>
</tr>
<tr>
<td>- Respondent 3</td>
<td>✔</td>
<td></td>
<td>✔</td>
</tr>
<tr>
<td>- Respondent 4</td>
<td>✔</td>
<td></td>
<td>✔</td>
</tr>
<tr>
<td>- Respondent 5</td>
<td>✔</td>
<td></td>
<td>✔</td>
</tr>
<tr>
<td>If a similar evaluation were conducted in your setting, would you consider the results based on similar evidence to be useful?</td>
<td></td>
<td></td>
<td>✔</td>
</tr>
<tr>
<td>- Respondent 1</td>
<td>✔</td>
<td></td>
<td>✔</td>
</tr>
<tr>
<td>- Respondent 2</td>
<td>✔</td>
<td></td>
<td>✔</td>
</tr>
<tr>
<td>- Respondent 3</td>
<td>✔</td>
<td></td>
<td>✔</td>
</tr>
<tr>
<td>- Respondent 4</td>
<td>✔</td>
<td></td>
<td>✔</td>
</tr>
<tr>
<td>- Respondent 5</td>
<td>✔</td>
<td></td>
<td>✔</td>
</tr>
<tr>
<td>Did the study adequately answer the question it had set out to address?</td>
<td></td>
<td></td>
<td>✔</td>
</tr>
<tr>
<td>- Respondent 1</td>
<td>✔</td>
<td></td>
<td>✔</td>
</tr>
<tr>
<td>- Respondent 2</td>
<td>✔</td>
<td></td>
<td>✔</td>
</tr>
<tr>
<td>- Respondent 3</td>
<td>✔</td>
<td></td>
<td>✔</td>
</tr>
<tr>
<td>- Respondent 4</td>
<td>✔</td>
<td></td>
<td>✔</td>
</tr>
<tr>
<td>- Respondent 5</td>
<td>✔</td>
<td></td>
<td>✔</td>
</tr>
<tr>
<td>Would you consider taking into account expert opinion?</td>
<td></td>
<td></td>
<td>✔</td>
</tr>
<tr>
<td>- Respondent 1</td>
<td>✔</td>
<td></td>
<td>✔</td>
</tr>
<tr>
<td>- Respondent 2</td>
<td>✔</td>
<td></td>
<td>✔</td>
</tr>
<tr>
<td>- Respondent 3</td>
<td>✔</td>
<td></td>
<td>✔</td>
</tr>
<tr>
<td>- Respondent 4</td>
<td>✔</td>
<td></td>
<td>✔</td>
</tr>
<tr>
<td>- Respondent 5</td>
<td>✔</td>
<td></td>
<td>✔</td>
</tr>
</tbody>
</table>
Figure 1  Process for Sensitivity Analysis

Is there decision uncertainty? (CEAC, DSA)

Yes

Does the uncertainty matter? (CE Plane, EVPI)

Yes

No

Identity Which Parameters contribute to uncertainty (Partial EVPI, DSA)

Can more information be collected / requested?

Yes

No

Substitute Decision-makers best judgement (DSA re-analysis, Interactive models)

Make interim decision

Obtain further information

Re-analyse

Make Decision
Appendix I

Cost-effectiveness of medical interventions to prevent cardiovascular disease in a sub-Saharan African country – the case of Tanzania (Robberstad et al, 2007)

BRIEF SUMMARY

Aim
To calculate the costs, effects, and cost-effectiveness of 14 medical interventions related to the primary prevention of cardiovascular disease (CVD) in Tanzania. The drugs are off patent, and included Acetylsalicylic acid, a diuretic drug (Hydrocholothiazide), a beta-blocker (Atenolol) a calcium channel blocker (Nifedepine), a statin (Lovastatin) and various combinations of these.

Methods
An economic model (Markov model) was used to estimate clinical outcomes and costs under the different treatment scenarios. Several modelling assumptions were made, relating to model structure and parameter estimates. All patients start out as non-symptomatic, and after each year, they may: remain well; experience a stroke or CHD; or die from other causes not related to preventive cardiology. The probability for each of these outcomes is dependent on the individual risk profiles, age and the drugs being offered. Some specific features are highlighted below:

• Individual’s risk profile was based on the Framingham equation. Tazanian diabetes prevalence data and observed mean values from Framingham were used.
• Treatment effects were obtained from US and European systematic reviews and meta-analyses (due to a lack of local African data).
• Drugs costs were based on the International Drug Price Indicator Guide. The median buyer prices represented the ‘c.i.f’ (cost free on board + insurance + freight), and adjusted with a domestic margin to calibrate to the local (Tanzanian) setting.
• Facility costs (e.g. those associated with health check up, outpatient visit, etc.) were based on a comprehensive costing study of Tanzanian health facilities.
• Disability adjusted life years (DALYs) were the main measure of clinical outcome, and were calculated using standard assumptions on life expectancy based on a Tanzanian life table.

Results
The incremental cost-effectiveness ratios for the fourteen interventions ranged from USD 85 to USD 4589 per DALY saved. Preventive cardiology was not found to be cost-effective for any patient group in this setting until the willingness to pay exceeds USD 85 per DALY.
We would like to seek your views on the conduct of this economic evaluation.

The authors explained that the burden of cardiovascular disease in sub-Saharan Africa was substantial. Despite not having comprehensive local data, they felt they had to undertake this study, using the best evidence available they had.

1. Do you agree with this decision?

2. Do you think that resources should always be directed towards studies for which there are adequate local data available?

3. If we should wait until more evidence was available, what specifically would you require before you felt you had enough (valid) evidence to make a decision?

Please complete your response here.

In this study, European and US data on treatment effect were used due to a lack of RCT data on CVD in an African setting.

4. Do you think this is reasonable?

5. Do you think observational data can be used as a data source for treatment effect?

6. What factors would increase or reduce your confidence in an RCT as a source of evidence?

Please complete your response here.
Drug costs were calibrated to a Tanzanian setting. Facility costs were based on a costing study of Tanzanian health facilities. How do you ensure that you cost data (e.g. facilities, drugs, labour, etc.) reflects local opportunity cost?

7. When conducting economic evaluation, how would you ensure that your cost data reflect the costs in your local setting?

Please complete your response here.

Both costs and DALYs were discounted at 3% in the study.

8. Do you agree with the choice of discount rate?  

9. Would you require a range of discount rates to be tested?  

The study reported incremental cost-effectiveness ratios ranging from USD85 to USD 4589.

10. Would you be satisfied with a decision to adopt treatment based on the findings of this study?  

11. If a similar evaluation were conducted in your setting, would you consider the results based on similar evidence to be useful?  

12. Did the study adequately answer the question it had set out to address?  

13. What would be your biggest concern over using this study for decision-making?  

Please complete your response here.
In general, in the context of economic evaluations for health technology assessment:

Yes  No

14. Would you consider taking into account expert opinion?

15. What factors would lead you to seek expert opinion despite the existence of trial data? (E.g. concerns over bias, accuracy, relevance, etc.)

Please complete your response here.

16. What factors would lead you to seek expert opinion despite the existence of trial data? (E.g. concerns over bias, accuracy, relevance, etc.)

Please complete your response here.

17. Do you have any additional comments about the data sources that were used in this economic evaluation, and how it would impact decision-making?

Please complete your response here.