Oxford Policy Management

Integration of iDSI's Reference Case principles for economic evaluation and DFID's approach to value for money analysis.

Opportunities and challenges

Alex Jones

October 2015

Acknowledgements

This report has been produced by Oxford Policy Management, with funding and technical assistance from NICE International.

In particular, the advice and input from members of IDSI's network of health economists was very useful and much appreciated – including from Paul Revill, Tony Culyer, Pete Smith and Mark Sculpher.

Significant input was also provided by Thomas Wilkinson and Kalipso Chalkidou, who, in addition to reading through and commenting on many drafts of the report, were central in setting the scope and objectives of the paper and coordinating the expert review and feedback process.

This assessment is being carried out by Oxford Policy Management. The project manager is Alex Jones. For further information contact Alex Jones at alex.jones@opml.co.uk

The contact point for the client is Thomas Wilkinson, who can be emailed at thomas.wilkinson@nice.org.uk.

Oxford Policy Management Limited

Registered in England: 3122495

6 St Aldates Courtyard
38 St Aldates
Oxford OX1 1BN
United Kingdom

Tel +44 (0) 1865 207 300
Fax +44 (0) 1865 207 301
Email admin@opml.co.uk
Website www.opml.co.uk

Executive summary

This report explores to what extent the principles recommended in the International Decision Support Initiative's (iDSI) Reference Case for economic evaluation could contribute to the Value for Money analysis done by the UK's Department for International Development (DFID) to help maximise the impact of its expenditure.

In the financial year up to April 2014 DFID expenditure exceeded £10 billion. Since April 2011 it has spent over £26 billion (HM Treasury, 2014). As part of the department's efforts to make sure this money was spent wisely, it uses (and continues to use) Value for Money (VfM) analysis to assess programme costs.

In 2013 the Bill and Melinda Gates Foundation (BMGF) commissioned a "Reference Case" (RC) for health economic evaluation. This is a principle-based guideline for planning, conducting and reporting economic evaluations of health technologies, services and programmes; designed to help improve the quality, comparability and usefulness of health technology economic evaluations in low and middle income countries.

While the VfM analysis done by DFID is not health economic evaluation, and is used across a wider range of sectors (not just health), both VfM analysis and economic evaluation are used to inform the same policy question – 'how can a limited budget constraint be best used to maximise a benefit?' It may be the case, then, that DFID could improve the usefulness of its VfM analysis through consideration of the RC.

It is argued that each of the RC's principles could contribute to DFID's VfM analysis. They each address separate components of maximising value subject to a limited budget, and thus could better enable DFID to maximise the impact of its expenditure. However, DFID's approach to VFM does not currently offer guidance on how to adhere fully to them, or demand that they are adhered to. It may be the case that some VfM frameworks for some DFID funded programmes adhere to some or all of these principles, but this will be a result of individuals or programme-specific analytical frameworks rather than any universal DFID VfM requirements.

A central component of the RC that may not be present in DFID's approach to VfM is a consideration of allocative efficiency. DFID's approach appears to focus on technical efficiency and measuring practical elements of good financial governance of funded programmes. While the useful nature of understanding a programme's technical efficiency is not disputed, particularly in terms of identifying potential improvements in efficiency, understanding a programme's allocative efficiency is necessary for maximising the impact of the department's overall expenditure.

DFID's approach to VfM is already a great step in the right direction. It is accessible online, and logical frameworks, including sections specifically focussing on VfM, are prepared in advance of programme implementation and then monitored throughout. How much was spent on a programme? What were the goals and objectives? What was achieved? These are all questions that, to a large extent, can be answered about DFID funded activities. They are crucial first steps in making resource allocation a systematic and accountable process, increasing efficiency and, hopefully, raising the quality of outcomes. However, the RC for economic evaluation may be a document that DFID can draw on in improving the usefulness of its analysis.

The recommendations made in this report are based on a comparison between DFID's approach to VfM and the RC. They are considered feasible given the data and methodological techniques available. They represent a significant opportunity for DFID to further improve its efficiency and extend the benefits of its expenditure.

Table of contents

Acknowledgements i

Executive summary ii

List of tables and figures iv

List of abbreviations v

Introduction 1

The question 1

Structure of the report 2

Discussion 2

Recommendations 3

1 Value for Money analysis and Economic Evaluation 4

1.1 Value for Money analysis and DFID's approach 4

1.2 Health economic evaluation and the Reference Case 5

2 Technical and Allocative efficiency 7

3 Can the RC principles contribute to DFID's VfM analysis? 9

3.1 Principle one: Transparency 11

3.2 Principle two: Comparators 12

3.3 Principle three: Evidence 14

3.4 Principle four: Measure of outcome 16

3.5 Principle five: Measurement of costs 17

3.6 Principle six: Time horizon for costs and effects 19

3.7 Principle seven: Costs and effects outside health 20

3.8 Principle eight: Heterogeneity 21

3.9 Principle nine: Uncertainty 22

3.10 Principle ten: Impact on other constraints 23

3.11 Principle eleven: Equity implications 25

4 Discussion 27

5 Recommendations 30

5.1 Recommendations for DFID 30

5.2 Recommendations for iDSI and the BMGF 31

5.3 Recommendations for future research 32

References 34

List of tables and figures

Figure 1: 3Es in the input to impact pathway 5
Figure 2: Allocative vs Technical efficiency 7

Table 1: Summary of the key findings 9

List of abbreviations

ART	Anti-Retroviral Therapy
BMGF	Bill and Melinda Gates Foundation
DFID	Department for International Development
HIV	Human Immunodeficiency Virus
NICE	National Institute for Health and Care Excellence
OECD	Organisation for Economic Cooperation and Development
OPM	Oxford Policy Management
RC	Reference Case
VfM	Value for Money
WHO	World Health Organisation

Introduction

The question

To what extent can the principles outlined in the International Decision Support Initiative's (iDSI)[1] Reference Case for economic evaluation contribute to the Value for Money analysis conducted by the UK's Department for International Development (DFID) and help it further maximise the impact of its expenditure?

In the financial year up to April 2014 DFID spent more than £10 billion. Since April 2011 it has spent over £26 billion (HM Treasury, 2014). As part of the department's efforts to make sure this money was spent wisely, it uses (and continues to use) Value for Money (VfM) analysis to assess programme costs. DFID's methodological approach to VfM is public and outlined in the document *DFID's Approach to Value for Money (VfM)* (DFID, 2011). The core objective is "maximising the impact of each pound spent to improve poor people's lives". Every year each programme funded by DFID must report on VfM as part of the programme evaluation framework – this is generally done first as part of an internal self-assessment, then corroborated through an independent external assessment.

This report focuses on these routine VfM assessments. It is acknowledged that DFID may also commission VfM analyses as part of larger scale one off impact evaluations. While relevant for these analyses as well, the focus here is on the annual assessments of VfM conducted as part of the routine monitoring and evaluation throughout programme implementation – as this analysis is most likely to follow DFID's *Approach to VfM*.

Separately, in 2013, the Bill and Melinda Gates Foundation (BMGF) commissioned the International Decision Support Initiative (iDSI) to lead a collaboration of researchers, methodologists and policy makers with an interest in economic evaluation to develop a "Reference Case" (RC) for health economic evaluation. This is a principle-based guideline for planning, conducting and reporting economic evaluations of health technologies, services and programmes. Published in June 2014, it is intended to:

- clarify meanings that are becoming confusing and inconsistent in health economic evaluation,
- emphasise the advantage of and encourage the use of common methodologies with comparable outcome measures and standard interpretations,
- serve as a standard guidance for all stakeholder institutions helping countries progress towards Universal Health Coverage.

While the VfM analysis done by DFID is not health economic evaluation, and is used across a wider range of sectors (not just health), both VfM analysis and economic evaluation are used to inform the same policy question – 'how can a limited budget constraint be best used to maximise a benefit?' While there are many different interpretations of what constitutes value to individuals, institutions and society, a key tenant of economics is that maximising any type of benefit within a

[1] The International Decision Support Initiative (iDSI) is a mechanism to provide policymakers (at sub-national, national, regional and international levels) with coordinated support in priority-setting as a means to Universal Health Coverage (UHC) with a particular focus on Low and middle income countries (LMIC). IDSI is coordinated by NICE International and consists of a network of academic units, priority setting institutions and think tanks. [Ref www.idsihealth.org].

limited budget demands consideration of the opportunity cost of spending decisions – what value could have been gained had existing resources been spent in other ways.

This report explores whether the principles and methodological techniques of the RC could contribute to DFID's approach to VfM analysis throughout all of the sectors in which it works. Better analysis here could mean better decision making by one of the world's biggest donors, and these better decisions could lead to improved outcomes for millions of people all over the world.

Structure of the report

Section 1 introduces the concepts of value for money analysis and economic evaluation. In this, DFID's approach to VfM and the RC for economic evaluation are outlined in more detail.

Section 2 outlines the difference between allocative and technical efficiency, and argues that DFID's approach to VfM is currently focuses more towards technical efficiency.

In section 3 each principle of the RC is addressed in turn (the RC is composed of 11 principles), asking whether it could contribute DFID's VfM analysis. This is done in four recurrent sections:

- A detailed outline of the principle.
- Is this relevant for DFID's approach to VfM? Does adhering to the principle mean analysis will better inform DFID on how to maximise the impact of its expenditure?
- Is the principle currently adhered to in DFID's approach to VfM?
- What are the feasible improvements on current practice that could be made?

Section 4 is a discussion of the findings, and in section 5 recommendations are made separately for DFID, iDSI and the BMGF, and for future research. Feasible improvements are recommended for each.

Discussion

It is argued that each of the principles could contribute to DFID's VfM analysis. They each address separate components of maximising value subject to a limited budget, and thus could better enable DFID to maximise the impact of its expenditure. However, it is argued that only one of the principles is currently adhered to (partially), but that '*DFID's approach to VFM*' does not demand adherence or offer guidance on how to adhere to the remaining ten. It may be the case that some VfM frameworks for some DFID funded programmes adhere to some or all of these principles, but this will be a result of individuals or programme-specific analytical frameworks rather than any universal DFID VfM requirements.

A central component of the RC that is missing from DFID's approach to VfM is the generation of evidence to allow an evaluation of allocative efficiency. DFID's approach focuses instead on something closer to technical efficiency. While the useful nature of understanding a programme's technical efficiency is not disputed – particularly in terms of identifying potential improvements in efficiency (Smith, 2009), understanding a programme's allocative efficiency is necessary for maximising the impact of overall expenditure. Addressing allocative efficiency would require that opportunity costs be properly integrated into analysis through adequate reflection of comparator programmes using a generalised measure of benefit.

This represents a significant opportunity for DFID to further improve the usefulness of its analysis to help extend the benefits of its expenditure.

Recommendations

It is proposed that at least partial adherence to eight of the 11 principles could feasibly be achieved through technical improvements to *DFID's approach to VfM* guidelines. These include:

- Defining the scope of relevant evidence on both costs and benefits.
- Being explicit that infeasibility of data collection does not mean irrelevance of data, and that missing data should be labelled as missing.
- Offering guidance on the incorporation of future costs and benefits, and how to inflate and deflate costs and benefits to reflect their present value.
- Requiring that the implications of total programme costs on all relevant budgets be expressed.
- Requiring that the implications of programmes on non-financial constraints (such as the stock of skilled labour) be presented.
- Requiring that benefits from multiple sectors be presented disaggregated by sector.
- Requiring that the heterogeneity of target populations be described.
- Requiring that the uncertainty of the conclusions due to the low quality and quantity of data used to inform them be characterised.

However, two of the most important principles will be much harder to adequately integrate and successfully implement – appropriate selection of comparators and quantification of benefits in a single measure. Not adhering to these principles is central to the observation that DFID's VfM analysis focuses on technical rather than allocative efficiency.

Drawing on the VfM focus on technical efficiency, it is also recommended that iDSI further develop the RC such that it can be applicable to programme evaluation as well as other investment decision making tools used by policy makers (beyond health economic evaluation). This could involve the development of new principles specific to economic evaluation of programmes or further methods specification of existing principles. To do this they will need to work closely with donors and governments in order to better understand how what allocation decisions are currently made, and what the constraints are.

Finally, three areas are highlighted for critical further research. The first is our ability to measure the benefits of the work institutions like DFID do in a comparable manner across all areas of their expenditure. Properly understanding the VfM of programmes demands a better understanding of the benefits programmes are having. There is arguably a long history of attempts at this, but an adequate solution is still elusive. The second is to better understand how to characterise uncertainty when the quality and quantity of data is very low, and the vulnerability to relatively likely shocks is high. If evidence based policy making is to work in low and middle income countries, we need to understand what the evidence is actually telling us. Ignoring what little data we do have because it is not enough, or because it is not of a good enough quality may lead to wasting relevant information. However, putting too much stock in it may lead to incorrect conclusions and inappropriate recommendations; a middle ground is needed so that VfM analysis can make the most of the available data. The third area recommended for further research is in measuring equality in low income settings. Only when it is possible to affordably estimate the equity implications of an intervention will it be possible to integrate equity concerns into an evidence based decision making process. Equity is too often ignored in the more rigorous components of VfM analysis and left for a qualitative discussion point at the end.

1 Value for Money analysis and Economic Evaluation

1.1 Value for Money analysis and DFID's approach

Value for Money (VfM) analysis is intended to inform decision making, often on resource allocation, with the objective of obtaining the best possible outcomes given limited resources. It grew out of the UK's New Public Management agenda of the 1980s and became an important part of the aid financing discourse following the 2002 Monterrey International Conference on Aid Effectiveness and the 2005 Paris Declaration on Aid Effectiveness (OPM, 2013).

While the intent of VfM analysis is generally accepted (informing decisions in a way that helps achieve the most given limited resources), what VfM is and how to measure it has led to much discussion.

Two of the core concepts are technical and allocative efficiency. Technical efficiency is a measure of how well a chosen set of outputs is achieved given a budget constraint, or, put another way, whether costs are minimised given expected output. Allocative efficiency, on the other hand, is a measure of whether or not those outputs are socially optimal basket to aim for. While analysis of technical efficiency is often done retrospectively to inform performance assessment (for example the work previously done by the Audit Commission and the Care Quality Commission in England to assess whether the unit costs of health care providers were excessive), analysis of allocative efficiency is generally prospective, used to inform policy makers where funds should be invested (such as the work done by the UK's National Institute for Health and Care Excellence (NICE)) (Smith, 2009).

VfM is often taken to mean a multidimensional balance between a number of outputs and an input (generally costs). Both the Organisation for Economic Cooperation and Development (OECD) and the World Health Organisation (WHO) have attempted to estimate and rank the value for money of each of their member states' health sectors – leading to significant methodological debate. The OECD study ranks countries according to ratios of expenditure and outcomes (proxied using life expectancy), and then the possible gains if they improve their efficiency (OECD, 2010). The WHO study was even more ambitious, and attempted to rank the total health system efficiency for each of its 191 member countries in 1997 by comparing health expenditure per capita to a weighted combination of health outcomes, inequalities in health outcomes, health sector responsiveness, inequalities in responsiveness and fair financing (WHO, 2000).

DFID's approach differs from that of the OECD and the WHO in that a single measure of performance or VfM is not calculated. Instead it evaluates VfM based on a reasoned assessment of three Es – economy, efficiency and effectiveness; defined as follows:

1. Economy: Are inputs being bought at the right prices? 'Are wages/fees too high or too low?' 'Are drugs and medical supplies being bought at the appropriate price?' 'Are text books being bought at the right price?' These are assessed through benchmarking against similar programmes in country for non-tradable goods such as labour, or internationally for tradable ones such as text books and drugs.

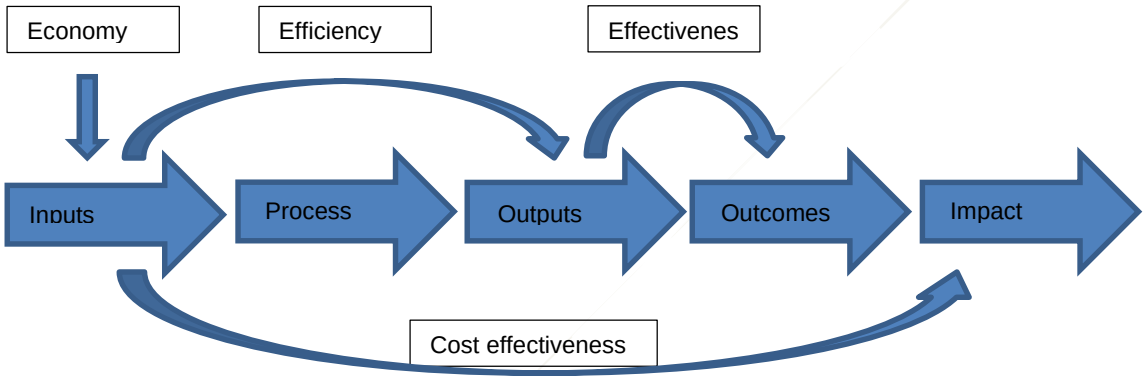
2. Efficiency: Are inputs being converted into outputs at a sufficient rate? 'Do more children have text books?' 'Are more people using drugs and medical supplies?' 'At what rate are these improvements being made given the inputs?' These are assessed through progress against output targets within a programme's logical framework.

3. Effectiveness: Are the outputs generating the desired outcomes? ‘Are children learning more at school?’ ‘Are fewer people sick or dying?’ ‘At what rate given the inputs?’ Again, these are assessed through progress against previously set outcome targets within the programme’s logical framework.

Outcomes are then hoped to have an impact. Cost effectiveness is theoretically the complete path – the rate at which expenditure on inputs can be converted to results at the impact level. So, the relationship between the three Es and cost effectiveness is that the three Es track the relationships between the inputs, outputs and outcomes that ultimately generate the intended impact of the programme (Figure 1).

The management board at DFID are then accountable for the VfM of the department’s programme expenditure. It is hoped that this has led to increased and more consistent/structured monitoring of VfM at the programme level, both internally and externally. All DFID funded programmes are required to report on their value for money as part of their regular reporting to DFID, and are assessed on it as part of their annual reviews. This means that programmes often have specific VfM logical frameworks, agreed with DFID, for which both a self-assessment and an external assessment must be completed on an annual basis.

Figure 1: 3Es in the input to impact pathway



1.2 Health economic evaluation and the Reference Case

Health economic evaluation has a long history of leading the field in the development of techniques for resource allocation decision making in the wider social sector (such as is being considered in this report). It is the use of a set of tools and approaches to inform resource allocation decisions relating to any technology or programme that has an impact on health (drugs, diagnostics, medical devises, public health interventions, purchasing strategies, benefits package provision etc.). The key objective in economic evaluation in health is to generate information about the health gains and other benefits (however defined) of interventions and programmes relative to their opportunity cost, in order to better inform health policy decisions. The most commonly used tools are cost-benefit analysis and cost-effectiveness analysis. However, in addition to these, tools from behavioural economics have been used to discuss the supply and demand of healthcare, and tools from inequality analysis have been used to discuss the distributional impacts of healthcare (Revill, Woods, & Sculpher, Forthcoming).

In recent years, the BMGF has begun to invest more heavily in the use of evidence to inform policy, including through the funding of health economic evaluations. The argument supporting active involvement in economic evaluation methodology is that priority-setting decisions in health

sectors are unavoidable, and that the decision making processes of local and national government institutions, as well as international organisations such as the GAVI Alliance, the Global Fund and the BMGF itself could be optimised through economic evaluation methodologies (iDSI, 2014). However, there is currently little consistency in methodology, quality or presentation of economic evaluations in low and middle income countries. This has reduced the comparability and usability of the analysis (Santatiwongchai, et al., 2015).

In response to this the BMGF commissioned iDSI to compile the case for an economic evaluation Reference Case (RC) to guide the planning, conduct and reporting of BMGF-funded economic evaluations and detail the standards that evaluations must meet if they are to be 'RC Compliant'. The idea is that if the BMGF adopt the RC, all future economic evaluations funded by them will have to comply with the RC principles. As they are currently funding such a significant proportion of the analysis, it is hoped that other funders will start to demand RC Compliance from their researchers as well, in a bid to increase comparability, usability and quality across the board.

The published RC includes eleven principles. They are based on existing good practice methodology and standardised methodologies or reference cases from the World Health Organization, UK's National Institute for Health and Care Excellence (NICE) and Thailand's Health Intervention and Technology Assessment Program (HITAP, Thailand) and also from the original reference case produced by the panel on cost-effectiveness in Health and Medicine convened by the United States Public Health Service in 1996. This incorporates extensive experience from making decisions on the VfM of technologies, and on a theoretical backing developed by those economists who have lead the development of cost effectiveness analysis in health over the last forty years. The eleven principles cover the core issues of transparency, comparators, generalizable impact measures, discounting, non-health sector costs and benefits, costs and effects for specific sub groups, uncertainty, fixed and non-fixed constraints and equity implications.

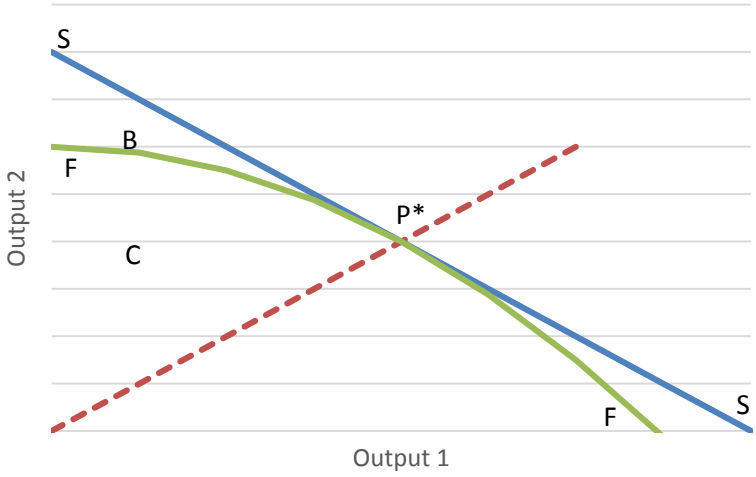
The argument is that if economic evaluations are transparent in the way they address the eleven key principles, and each of the principles are addressed, then the quality, comparability and usability of health economic evaluations will increase.

2 Technical and Allocative efficiency

Before addressing each of the RC principles in turn, this report comments on an overarching distinction between the RC and DFID's approach to VfM analysis. In its current form, DFID's approach is focused on something similar to technical efficiency, but does not address allocative efficiency.

The distinction between these two concepts is well outlined in an example given by (Smith, 2009). With current technologies and interventions available, imagine two outputs can be produced together to different extents with maximum possible production of each represented by the frontier FF (Figure 2). The relative social value of these outputs is represented by the line SS. Anywhere below SS represents a lower social value than along the line (allocative inefficiency), and anywhere below FF represents a lower production of either (or both) outputs than is possible given the constraints (technical inefficiency). While technical efficiency is concerned with how outputs are produced relative to the frontier FF, it does not capture the relative social value of the outputs, and so does not comment on whether the appropriate mix of outputs are produced given social preferences. The social value of the outputs is maximised where both technical and allocative efficiency are achieved – at P*.

Figure 2: Allocative vs Technical efficiency



In terms of DFID's approach to VfM, programmes may produce outputs, outcomes and impacts to different extents (all considered as outputs for the purposes of Figure 2). Some programmes, such as C, are below the production frontier FF, and so are technically inefficient. They could be producing more outputs given their constraints.

This is, to some extent, picked up in DFID's approach to VfM through the assessments of economy, effectiveness and efficiency. Programmes that are spending too much on their inputs (compared to other similar programmes) are identified. The same is true for those not achieving the outputs and outcomes that they planned for in their logical frameworks. Hopefully programme adjustments can be made, leading to an increase in future efficiency, improvement of programme performance and a movement towards the frontier FF.

On the other hand, some programmes, such as B, are already technically efficient. They should score well on economy, effectiveness and efficiency, and in turn be credited as good value for money through DFID's VfM analysis. They could not increase any of their outputs without

decreasing some of the others and are considered to be functioning well. However, these programmes are not necessarily producing outputs at the highest social value possible (only P* does this).

For example, if a programme is purchasing branded drugs, it could increase output for the same input by swapping to generic drugs. This would be identified through an assessment of the programme's economy. However, what if the programme is purchasing HIV treatment, when actually malaria is the real or bigger problem? It may be the case that the money used to buy the branded HIV treatment would have better been transferred to payments for malaria treatment. Just considering the economy, effectiveness and efficiency of the HIV programme in isolation does not identify this – it could be running perfectly, it is just that the programme is the wrong choice of programme.

It is important to note that DFID's VfM framework does not explicitly state that it is addressing technical efficiency. It is arguably more focussed on the practical elements of "spending money wisely" through, for example, good procurement. This may be interpreted as attempting to monitor whether desired outputs are delivered in the most technically efficient way, but the fact that incremental analysis and quality analysis are not done questions whether the approach to VfM is capturing technical efficiency either, or just doing a form of financial audit/monitoring and evaluation.

3 Can the RC principles contribute to DFID's VfM analysis?

This chapter is the heart of this report. It outlines each of the 11 principles in the RC, asking to what extent satisfying them would make DFID's VfM analysis more useful. Table 1 is a summary of the key findings. Ten of the 11 principles are found to not yet address in the guiding document *DFID's Approach to VfM*. It is argued that significant but feasible improvements are possible that could make the analysis more useful to DFID, and potentially lead to better outcomes in the future.

A key distinction to be clear on here is between *principles* and *methods*. Cost benefit analysis, cost effectiveness analysis and value for money analysis are all different methods for informing resource allocation in such a way that satisfies a maximisation problem. Where researchers have access to different resources, time or information available to conduct a study, or where the informational needs of a decision maker varies, different methods may be appropriate. There is a suggestion, however, that principles (such as those in the RC), should hold across all methodologies looking to inform resource allocation (Revill, Woods, & Sculpher, Forthcoming). It this is true, then each of the principles for health economic evaluation should also apply to VfM analysis across a range of sectors as well as health.

Table 1: Summary of the key feasible improvements

Principle:	Proposed feasible improvements on current practice:
1. Transparency	Include a decision rule for whether or not a programme is considered good value for money at the programme level, prior to VfM analysis being conducted – 'these are the circumstances under which we will consider a programme good VfM…' State the perspective from which the VfM analysis is being done: DFID, DFID country office, national government, etc.
2. Comparators	Where programmes are compiled of many interventions and it is impossible to identify a programme comparator, it should be possible to identify comparators for the key interventions that make up the programme – make the most of this. Plan programmes with difference in difference analysis in mind. Identify comparator states/districts from the start of implementation and track throughout implementation.
3. Evidence	Define the scope of relevant evidence. What evidence is relevant to an assessment of VfM and what is not? Consistency is needed if results are to be comparable over time and between programmes. Explicitly state that infeasibility of data collection does not mean irrelevance of data. The relevant data that is not available should be clearly labelled 'missing' in VfM analysis.
4. Measure of outcome	Impacts should be routinely measured in the most generalizable unit for their sector. Impacts should be reported on disaggregated by sector/measure.

Principle:	Proposed feasible improvements on current practice:
	If authors feel that comparison between sector specific benefits is possible, this may be done in addition.
5. Measure of costs	Define the scope of relevant costs. In general, all costs of the programme not also faced by comparators programmes should be included. Clarify whether costs incurred by other organisations should be included or not. This decision should be based on intended role of DFID in driving improved resource allocation of its own investments, of recipient country investments and of its partner organisation investments. Do not neglect the expenditures incurred by DFID and implementing partners due to exchange rate fluctuations and inflation.
6. Time horizon for costs and effects	Incorporate explicit judgements on relevant time horizons for programmes on an individual programme basis according to duration of new costs and benefits. Deflate future costs and benefits using country specific discount rates for both costs and benefits. Inflate past programme costs and benefits with the same logic that future costs and benefits are deflated.
7. Costs and effects outside health	Costs and benefits outside the primary jurisdiction (in the case of the RC this is health, but for DFID's VfM analysis it could be anything) of a programme should be presented in the same manner as those in the primary jurisdiction – disaggregated by sector in the most generalizable metric possible. It is likely that information on costs and benefits outside a programme's primary jurisdictions will be missing. In this case it should be clearly reported as missing.
8. Heterogeneity	VfM reports should explicitly attempt to describe target population heterogeneity and comment on potential differences in VfM across different sub groups – even if this is to say that nothing is known about the population heterogeneity.
9. Uncertainty	Include deterministic sensitivity analysis of the key variables and assumptions in the VfM analysis. Where possible, probabilistic sensitivity analysis is preferable. However, in most cases it is acknowledged that this will be too data intensive to be feasible.

Principle:	Proposed feasible improvements on current practice:
10. Impact on other constraints	Programme costs should be put in the context of wider budgets – in particular the DFID country office budget, total DFID budget and, where local handover is an option, domestic government budgets. The opportunity costs of the major non-financial resources (such as human resources and health system infrastructure) should be described and, where possible, quantified.
11. Equity implications	Provide clarity on how equity implications of programmes should be presented in VfM analysis. One possible suggestion is to present the outputs and outcomes by the different groups identified during the discussion of target population heterogeneity. This would enable discussion of the programme's impact on different groups within a society.

3.1 Principle one: Transparency

RC: An economic evaluation should be communicated clearly and transparently to enable the decision makers(s) to interpret the methods and results.

Principle outline

Being transparent means being clear about the decision problem (i.e. what intervention and outcome are being assessed? In what population and context? What are they being compared to? How are they being compared?) It underpins a number of the other principles, particularly decisions about appropriate evidence (section 3.3). The scope of relevant evidence will be directly informed by the decision problem. For example, if the decision problem states context as "Sub Saharan Africa" then the evidence requirements under principle three are broad. If the context is stated as "primary care facilities in rural Tanzania" then the evidence requirements are narrow.

The objective of economic evaluation is to inform policy. If the methodology and results of an evaluation are not reported clearly and transparently, even the most robust analysis of the soundest evidence will not be informative. Furthermore, transparent reporting of analysis increases the transparency of decision-making processes (thereby increasing accountability). It also increases the transferability of results (even when overall results are not generalizable, aspects within the work may be) and the procedural quality of the work (it can be repeated and tested).

Relevance/Value to VfM

The objective of DFID VfM analysis is to inform decision making regarding financial allocation between DFID funded programmes in order to maximise impact. As information generated through more transparent methodologies is more informative, this is both a relevant and valuable principle for DFID's approach to VfM.

Current reflection in VfM

To some extent this is already embedded in DFID's VfM analysis. DFID's approach to VfM analysis is available online, and the VfM of DFID programmes is internally and externally assessed through predefined VfM logical frameworks on an annual basis. These logical frameworks outline clearly the goals and objectives as well as intended outputs and outcomes of programmes.

However, the complete VfM decision making process it is not transparent. While the criteria against which programmes should be assessed are clear, the scores they must achieve in order to be considered good VfM are not. There is no hurdle value for money that a programme must exceed in order to be funded. This is justified through reference to DFID's diverse portfolio – making a generalizable VfM unit unfeasible. If a generalizable unit of VfM is not feasible, how can programmes be scored in a comparable manner? For example, how should the value for money of a programme to increase access to safe maternal health care be compared to the value for money of a programme to increase competition in private sector markets?

The problem DFID highlight here is an outcome issue, further discussed in section 3.4. But it creates a transparency issue because without an outcome measure it becomes impossible to know how the VfM of one programme, and the corresponding recommendations should be compared to another. Moreover, it is plausible that the analysis of one programme could be repeated by different assessors, and, with the same evidence, different conclusions could be drawn about whether a programme needs to improve its VfM or not. The lack of a usable outcome measure means the decision problem cannot be properly specified, and decreases the transparency of the process.

Feasible improvement

The feasibility of a general outcome measure of VfM across DFID programmes is a significant problem, and discussed in section 3.4. Aside from this, DFID should be commended for the high level of transparency in objectives, goals, and outputs associated with its funded programmes.

3.2 Principle two: Comparators

RC: The comparator(s) against which costs and effects are measured should accurately reflect the decision problem.

Principle outline

This principle is central to current approaches to economic evaluation of health interventions. The RC defines a good decision as one that involves comparing the additional health benefits of an intervention with the health likely to be lost (health benefits forgone) elsewhere as a consequence of additional costs. The underlying assumption is that resources spent on one intervention cannot then be spent on another, so there is an opportunity cost to any resource use. If this opportunity cost (health benefits forgone) is greater than the actualised health benefits, then resources could have been better spent on the alternative intervention.

Specifically, the correct selection of a comparator programme is necessary for accurately reflecting the incremental cost-effectiveness ratio of a programme. This identifies the cost of the benefits on the margin. It is the ratio of the extra money that is spent because the programme was implemented to the extra benefits that are generated. How this could be otherwise spent is the true opportunity cost of the decision.

It is also important to highlight that analysis of comparators is only useful once the correct comparators are chosen. The correct comparators are those which are realistic alternatives that the money would likely be spent on, were it not spent on the programme under evaluation.

Relevance/value for VfM

Each year DFID has an expenditure limit set by the treasury. In the 2014/15 budget DFID was given a total expenditure limit of £10.3 billion. This acts as a fixed budget constraint. Because the budget is fixed, there is an opportunity cost to all programme funding, and, in turn, understanding the alternative possible uses of the money (comparator programmes) is necessary for monitoring VfM.

Current reflection in VfM

Currently, comparator programmes are only loosely addressed, and only to assess the first of the three Es – economy. VfM analysis makes use of benchmarking to assess whether a programme is paying too much for any of its inputs – particularly labour. This works by looking at similar programmes with similar inputs, and seeing if the programme under evaluation is paying more or less than the comparator programme for that input. For example, what is the average daily rate for a local or an international consultant? And how does this compare to other programmes?

This is problematic for at least five reasons. First, assessing economy separate from outputs can become a cost containment exercise, which does not necessarily lead good VfM – the objective is just to spend less on an input than the comparator programme, not linked to a corresponding measure of output quality.

Second, programmes are only compared between each other (within a small set) and not against a standard of economy. This means that the economy of a programme is considered good once it is cheaper than the next programme, with no assessment of the economy of that comparator programme. If that programme represents bad economy, then having better economy should not in itself mean good economy.

Third, the assessment is done after expenditures have been actualised. This means that it cannot inform resource allocation of expenditure between comparators. It can be used to inform future decision making in similar circumstances, and give a push for costs to be contained in continuing programmes. But, for this, VfM analysis needs to be explicitly linked to planning processes of future programmes, which is not clearly outlined in *DFID's approach to VfM*.

Fourth, it does not take into account the overall budget implications of programme expenditure on specific inputs (for example, it is assessed whether consultants are hired at competitive unit prices, but total expenditure on consultants is not considered). This means that the total opportunity cost of expenditure is never actually quantified – which means alternative uses of money cannot be meaningfully considered.

Fifth, even if total budget implications were assessed, this benchmarking is done between similar programmes in the same country. The decision problem for DFID is not restricted to any one country. DFID is allocating funds between many programmes in many countries. If DFID is looking to maximise its impacts globally, it needs to understand how the benefit of a programme in Sierra Leone relates to the cost of a programme in Ethiopia. If, on the other hand, VfM analysis is from the perspective of individual countries, then comparisons can be more restricted. This relates to the transparency principle, outlined in section 3.1.

When assessing the effectiveness or efficiency of programmes (relating to outputs and outcomes) benchmarking is not used. Instead, programmes are assessed against targets agreed between programme implementers and DFID, independent of explicit reference to comparator programmes.

As such, DFID's VfM analysis, in its current form, does not adequately address appropriate comparators, the incremental cost effectiveness of programmes or the opportunity cost of their expenditures.

Feasibility

The clearest way to make progress on this may be to separate the issue by unit of evaluation – whole programme or individual intervention (programmes are often compiled of a number of different interventions). At present, this report acknowledges that in situations where many interventions are being implemented simultaneously (as part of an overall programme), and the sets of interventions and the extent to which they have been implemented vary significantly in different areas, it is very difficult to isolate two programmes for unbiased comparison.

The RC recommends identifying 'all possible interventions available', and comparing the treatment to either 'do nothing', 'best practice' or 'the intervention most likely to be replaced if the treatment is adopted'. Because programmes are often made up of numerous interventions, the number of permutations for a set of interventions means the set of 'all possible programmes' may be prohibitively larger than is the case in evaluations of individual interventions, making the recommendation too demanding to be applicable.

Pursuing analysis of individual interventions and their comparators may be a more feasible way forward. This could be considered a kin to a process of identifying 'best buy lists' that look at interventions individually rather than as part of an overall strategy – such as the WHO choice model. While information on how interventions interact with each other would not be captured, this would be an improvement on current practice – which only inconsistently sees this sort of information used. Assessing the VfM of individual components of programmes may also be a good way of highlighting inefficient aspects of overall programmes, which is a useful part of improving overall programme efficiency (Smith, 2009).

Nonetheless, as DFID is generally assessing programmes rather than interventions, whole programme evaluation should not be neglected for seeming impossible. Possibilities may arise from the common district/local government structures of many of the countries in which DFID operates. Many have devolved significant amounts of power to district or local governments. One consequence of this is that it is not uncommon for a programme to be implemented in some districts, but not others. This creates the potential for natural experiments that use similar states as comparators. There may be cases where difference in difference analysis of panel data could then be used to compare districts with and without a programme.

3.3 Principle three: Evidence

RC: An economic evaluation should consider all the available evidence relevant to the decision problem.

Principle outline

Failure to draw on all available evidence may introduce bias in an unknown direction and limit the capacity of an economic evaluation to inform good decision-making. What constitutes all available evidence comes down to the author's judgement, but should be reported transparently.

The reference case states that the budget and time available for an evaluation are relevant to the assessment of the feasibility of an evaluation and scope of a decision problem, but that they should not influence the scope of the relevant evidence. A key aspect of this principle is that infeasibility of data collection does not mean irrelevance of data. While it may be the case that the collection and synthesis of all relevant evidence is prohibitively expensive or time consuming, transparent judgements should be made about the implications of not including the missing information. This is further developed in section 3.9, where the characterisation of uncertainty is addressed.

Relevance/value for VfM

VfM analysis is an evidence informed activity. In the same way that excluding data may bias results of an economic evaluation, excluding data may bias results of VfM analysis. Without unbiased estimates of costs and benefits, any evaluation of VfM may be unreliable. If the scope of relevant evidence is not defined pre analysis, there may be a tendency to focus only on the available evidence, rather than contextualising that evidence within the total set of relevant evidence and then drawing conclusions. It also means that different programmes in different years may include more or less evidence as relevant – which decreases the comparability of results, both between programmes and over time.

Current reflection in VfM

In demanding that VfM is measured through an assessment of economy, effectiveness and efficiency, DFID implicitly demand evidence on inputs, costs, outputs and outcomes. However, the scope of this is not defined. It is not clear which costs (discussed further in section 3.5), outputs and outcomes should be included. For example, should indirect costs and benefits be included? Should future costs and benefits be included? Should costs and benefits to those outside the target population be included?

DFID's approach to VfM analysis is clear that analysts should be confident in the strength of the evidence base and explicit in stating the underlying assumptions relied on to arrive at any conclusions. However, the RC does not offer guidance on how to assess the strength of the evidence base, or make any comment about the scope of an evidence base other than that staff should have access to the latest evidence and systematic reviews of the evidence. In the cases where there is only limited evidence regarding what works, DFID requires that strong monitoring and evaluation plans are in place, and the need for impact evaluations should be considered – with partners encouraged to do the same.

Feasibility:

Accounting systems that generally already exist mean expenditure data should be available (although compiling it is a problem). Sub units within a sector can feasibly be expected to record their activities; for example a health facility recording deliveries, a training school recording graduates, a primary school recording enrolment. Large scale household surveys, however, are very expensive, and generally only carried out in lengthy time cycles. The Demographic and Health Survey, for example, aims to be completed and published every five years. For this reason, it should generally be feasible to include evidence on costs, inputs and outputs in VfM analysis. However, evidence of outcomes and impacts may be less feasible, as large surveys are currently the only national scale window into the impacts at the household level of the programmes DFID fund.

There is a large literature on value of information analysis which should be used to guide where investment in further evidence is advisable. For a brief snapshot see (Fenwick, Claxton, &

Sculpher, 2008). The general message is that some evidence is prohibitively expensive to gather because the opportunity cost of gathering it is greater than the benefit the information brings. When budget constraints are low, opportunity costs are high and missing data is an inevitable part of evidence based decision making.

However, as required by the RC, DFID should be explicit in VfM that the scope of relevant evidence is independent of the feasibility of data collection, and that where data is missing this should be clearly stated and the uncertainty that arises due to this missing data should be characterised (further explored in section 3.9).

3.4 Principle four: Measure of outcome

RC: The measure of health outcome should be appropriate to the decision problem, should capture positive and negative effects on length of life and quality of life and should be generalizable across disease states.

Principle outline:

Using a measure of health outcome that is broad enough to capture all socially valued aspects of health and is applicable across various investment types enables full consideration of the opportunity costs of expenditure on any intervention.

Using a disease specific health outcome measure, on the other hand, limits the ability of the decision maker to make informed trade-offs between competing options for how to spend money across different disease states, undermining comparability and consistency of decision-making.

Relevance/value for VfM:

DFID programmes are implemented in a large number of settings – health, education, financing, private sector development, water, sanitation, employment development, etc. Making trade-offs between the impacts achieved in different sectors and settings is unavoidable and accountability in this sort of trade-off is one of the objectives of VfM analysis. Ideally, the measure of impact should capture positive and negative effects of a programme and be generalizable.

Current reflection in VfM:

Returning to the issue highlighted in principle one – DFID have elected not to adopt a generalizable measure or level of VfM because their portfolio is too diverse for a generic measure to accurately rank all programmes.

What is generalizable in DFID's approach to VfM analysis is that economy, efficiency and effectiveness should always be considered. However, without a common unit to measure the three aspects of a programme across a range of settings, some degree of accountability is lost.

Feasibility:

The feasibility problem outlined by DFID is difficult. DFID have thought about a standardised VfM hurdle that programmes must pass in order to be considered for funding, but rejected the idea because the notion of a general measure of VfM seemed unlikely, due to the lack of a general measure of outcome.

While metrics have been developed for the health sector – the Disability Adjusted Life Year (DALY) and the Quality Adjusted Life Year (QALY) are two examples – equivalents in other sectors, or general equivalents across sectors are less well developed. Work on how to incorporate non health

benefits fully into health economic evaluations has struggled with the same question. The core problem is that including benefits from a variety of sectors within a single policy evaluation requires a specified social welfare function – where the relative values of different aspects of livelihood are defined and quantified in a comparable manner (Revill, Woods, & Sculpher, Forthcoming).

There are a number of ongoing attempts at this. Perhaps that most well documented and developed is the Human Development Index. This has grown over the last 25 years, managed by the United Nations Human Development Programme, and is now published alongside similar indices adjusted for equality, gender and gender inequality in the Human Development Report. It is a combined measure of expected income, health and education in a country, and was designed to capture the capacity of individuals to define their own lives. A more recent index, developed by the Oxford Poverty and Human Development Initiative and also published in the same Human Development Report, is the Multidimensional Poverty Index. This adds to the mix a measurement of general living standards, estimated through asset ownership, flooring type and access to water, fuel, electricity and sanitation (Alkire, Conconi, & Seth, 2014). A third approach, currently being pursued by the Overseas Development Initiative, takes influence from the methods used to value different health states to inform health spending in England and Wales (Melamed, 2011). The EuroQol group's generic survey describes possible health states according to five dimensions of health and three health outcomes – leading to 243 possible health states. Stated preference games can then be played with members of the public to reveal how they rank each possible health state, which in turn can guide health expenditure allocation. A similar process aiming to describe poverty and rank the different states may enable decision makers at institutions such as DFID to make informed decisions about allocating their money between programmes in health, education, market development etc. – if poverty is agreed to be the underlying or key issue.

However, capturing information necessary to track changes in these indices and specific enough to inform programme evaluation may be problematic. It would generally demand survey data, which is prohibitively expensive for programme evaluation.

While work continues in this field, it should at least be possible to have a common outcome measure in some fields, for example health. This report recommends that where there are expected health impacts of a programme, these can and should be routinely measured either in terms of DALYs averted or QALYs gained. Non health impacts should be routinely measured in the most generalizable unit available for the specific sector, and reported disaggregated by sector. If comparisons between sector specific benefits are then considered possible, this can then be pursued in addition. This is further explored in section 3.7 – where the RC addresses the problem of comparing benefits measured in non-comparable metrics.

3.5 Principle five: Measurement of costs

RC: All differences between the intervention and the comparator in expected resource use and costs of delivery to the target population(s) should be incorporated into the evaluation.

Principle outline:

Overall costs of interventions should be reported, including costs of resource inputs. In addition, quantities of resources should be reported separately from their unit costs to enable decision makers to assess total opportunity costs. Costs and resource use do not need to be included where they do not differ between evaluated alternatives, as these will not contribute to any difference in cost-effectiveness. Costs and resources used by other organisations should be included, because there is still an opportunity cost to this expenditure, even if not to the primary organisation or even jurisdiction. (Dis)economies of scale and scope should be highlighted, so that decision makers will know how unit costs are likely to rise or fall if implemented in a different jurisdiction or with a different scope – however information on this is likely to be scanty.

The main emphasis should be placed on transparency of an attempt to include all costs associated with an intervention that differ from the evaluated comparator.

Relevance/value for VfM:

Measurement of costs is central to VfM analysis. In addition to capturing all costs of a programme that are different to its comparators, there are a number of particular issues faced when measuring costs through DFID's approach to VfM. First, DFID aims to maximise the impact of both its own expenditure and non-DFID aid. This significantly expands the scope of relevant expenditure information. Second, many DFID funded programmes have the potential to be expanded from local to national interventions. It is not always clear how costs will change as programmes are rolled out. Third, many of the countries in which DFID operates experience high inflation rates – including for the prices of non-tradable goods such as labour. Coupled with this, while DFID spends in pounds sterling, its programmes spend in domestic currencies and the exchange rates between these currencies fluctuate. If not monitored and planned for properly, inflation and exchange rate fluctuations can effect large expenditures on a programme. All of these costs need to be captured in both prospective planning and retrospective VfM analysis.

Current reflection in VfM:

As DFID's VfM analysis is retrospective, it can include total expenditure data. However, the scope of total expenditure is not defined. While one of the stated objectives is to improve the VfM of all aid, there is no guidance on how to capture expenditure by other organisations and how it relates to the VfM of the DFID expenditure. Moreover, there is no mention of how to analyse the impacts on a programme's VfM if its scope, scale or jurisdiction are changed, or on how to deal with inflation and exchange rates.

Feasibility:

Analysis of direct expenditure on a programme is achievable – both by DFID and any counterpart funders. However, the total cost of a programme may go way beyond this. The true total opportunity cost of a programme includes indirect expenditure, such as when a programme uses the time of someone whose salary is not paid through the programme. If it is to have an impact, a programme to improve education through building classrooms, for example, relies on teachers turning up to teach (paid by government) and parents sending their children to school instead of to work. While these are both part of the total cost of educating a child, particularly the second is hard to value.

As costs that do not differ across interventions do not need to be included in analysis, if DFID were only comparing education based interventions these costs may not be relevant. However, if DFID is comparing interventions across a wide range of settings, exclusion of costs will rarely be appropriate. In order to take a stance on this, DFID needs to be clear about its objectives. If it is concerned with funding programmes that are cost-effective on the margin (cost effective aspects of an overall strategy with unknown cost effectiveness) it only needs to consider more direct costs. If it is really concerned with improving the VfM of all aid, and, conceivably, all social expenditure, it needs to consider the wider spectrum of indirect costs such as teacher salaries as well.

Capturing and preparing for expenditures due to inflation and exchange rate fluctuations is manageable as both are internationally well monitored economic indicators. Also, with some rudimentary assumptions, particularly relating to populations, demographics and geography, the costs associated with scaling programmes up and down should be possible.

3.6 Principle six: Time horizon for costs and effects

RC: The time horizon used in an economic evaluation should be of sufficient length to capture all costs and effects relevant to the decision problem; an appropriate discount rate should be used to discount cost and effects to present value.

Principle outline:

The time horizon is the period over which the costs and effects of the intervention and comparators are calculated. This is important in economic evaluation because interventions have costs and benefits that occur at different times, and often a long time into the future. All benefits and all costs associated with an intervention (but that differ between comparators) should be incorporated into an evaluation, no matter when they occur.

It is important that the costs and benefits occurring in the future are discounted at a relevant rate. Most social investment programmes discount future costs and benefits to reflect three core factors – the extent to which people want to bring forward benefits and delay costs, the extent to which future costs and benefits may not occur and the extent to which future costs and benefits may be less important due to some form of growth (Revill, Woods, & Sculpher, Forthcoming). Discounting costs and benefits is a method for estimating their present values (what are they worth in terms of the same costs and benefits today) so that they can all be compared.

Relevance/value for VfM:

DFID funded programmes may include medium or long-term development expenditure plans or establish the need for recurrent expenditure indefinitely. They may also lead to healthier, more educated, higher earners over the next 20 to 30 years. Properly assessing the value for money of a programme demands that these future costs and benefits be addressed. As such, accurately setting the time horizon of VfM analysis to capture all costs and effects of a programme is both valuable and relevant.

Similarly, estimates of these future costs and benefits should be discounted and expressed in their present value so that they can be compared.

It is also important to recognise that VfM analysis is retrospective. Where VfM looks back over a number of years, for the same reason that future costs and benefits should be discounted, past costs and benefits should be inflated. This is about expressing the present value of costs and benefits to enable more accurate comparison of information between years.

Current reflection in VfM:

The DFID approach to VfM makes no mention of discounting, inflating, present value or time horizon. As mentioned in section 3.3 and 3.5, the scope of relevant evidence and expenditure is not defined. No guidance is offered on how to compare expenditures and benefits occurring in different years. Moreover, no mention is given on how/whether to incorporate future costs and benefits at all.

Feasibility:

Setting a time horizon and attempting to capture total costs and benefits (both past and future) in terms of their present value is feasible, and guidance on doing this should be explicit in DFID's approach.

The choice at which rate costs and benefits should be discounted/inflated can be problematic however. The WHO-CHOICE guidelines recommend that both benefits and costs should be

discounted at 3% per year, and a recent review of economic evaluations in low and middle income countries found 3% to be a very common discount rate for both costs and effects (Santatiwongchai, et al., 2015).

It is not clear, however, what this recommendation is based on, whether the same discount rate should be applied to both costs and benefits, or whether the same discount rate should be applied in different countries and different sectors. A higher discount rate for benefits favours options where benefits are realised in earlier periods, whereas a higher discount rate for costs favours options where costs are faced in later periods (Revill, Woods, & Sculpher, Forthcoming). If budget constraints are expected to be significantly relaxed in the medium to long term, for example, it may be more appropriate to employ a higher discount rate for costs than benefits. This is obviously something that is context and time dependant, and using inappropriate rates may inappropriately inform decisions.

How then should DFID set discount rates? Should they be different depending on the context? Should health or education benefits be discounted at a different rate in Pakistan or Malawi? If they should be set according to the domestic health and education sectors then the answer is yes. However, this may be politically challenging, as it will result in DFID explicitly valuing future benefits in different countries at different rates. It is also significantly increases the complexity of DFID's approach to VfM, as separate rates will need to be set for each sector and each country and regularly reviewed. This is not necessarily an insurmountable task. The International Monetary Fund, for example, use different discount rates for different countries when calculating the present value of debt as part of their regular debt sustainability analysis.

3.7 Principle seven: Costs and effects outside health

RC: Non-health costs and effects associated with gaining or providing access to health interventions that do not accrue to the health budget should be identified where relevant to the decision problem. All costs and effects should be disaggregated, either by sector of the economy or to whom they accrue.

Principle outline:

Most health economic evaluations are concerned with how available healthcare resources are allocated to generate health gains. In addition, however, non-health benefits and costs may be relevant to a decision problem. In particular, such impacts include wider costs and benefits incurred by families and communities. For example, when a parent is healthy, just measuring their health does not measure the large non-health household benefits this can bring, particularly in terms of income. When a parent attends a health clinic, just measuring the cost of care does not capture the cost of transport and lost income that brings. If overall health benefits outweigh direct health costs, but not total costs, implementing an intervention may not be the optimal decision.

It is not clear how health benefits should be traded with non-health benefits, so the RC recommends that non-health costs and benefits should be reported, but disaggregated by sector.

Relevance/value for VfM:

Livelihood is multidimensional, and its quality reflects many things – health, opportunities, life-style, income, family etc. As such, ideally DFID would capture costs and effects beyond the primary arena of its programmes. An agricultural programme designed to decrease reliance on imported foods may also have implications on population health through improved diets. This has a social value, and is relevant to VfM analysis.

The challenge highlighted in sections 3.1 and 3.4 – that it there is currently no general measure of benefit across all of the arenas in which DFID operates – is the same problem that the RC faces in trying to incorporate costs and effects outside health.

Current reflection in VfM:

As mentioned already, DFID's approach to VfM does not offer guidance on the scope of the relevant cost or benefit data. It is unclear whether non-aid expenditure on a programme should be considered or how to incorporate non-DFID expenditure. It is also unclear what to do with indirect programme benefits. The lack of a generalised measure of benefit has meant that DFID do not compare benefits between sectors. It is also not clear from whose perspective VfM is being done. Can something be good VfM for DFID, but not for the society in which it is implemented? This could be the case if a DFID programme generates the outputs it intends at a low cost, but unintentionally causes increased costs on a third party.

Feasibility:

Including non-programme costs and benefits will be very demanding. However, as outlined in section 3.3, availability of evidence should not impact the scope of relevant evidence – infeasibility does not mean irrelevance. Where this data is available it should be incorporated into VfM. Where it is not, it should be identified, and the uncertainty generated by the missing data characterised (as discussed in section 3.93.10).

The question of how to prioritise benefits in different sectors in one programme poses the same challenge as is outlined in section 3.4 of how to prioritise different programmes being implemented in different sectors. There is currently not an established metric for doing so.

This report recommends that DFID adopt the same solution as is proposed in the RC. Benefits from different sectors should be disaggregated, and presented according to their social sector. Where sector specific generalizable measures are available (for example the QALY or the DALY), they should be used. If authors feel that comparison between sector specific benefits is possible, this may be done in addition.

3.8 Principle eight: Heterogeneity

RC: The costs and effects of the intervention on sub-populations within the decision problem should be explored and the implications appropriately characterised.

Principle outline:

Heterogeneity refers to the situation where sub groups within a larger group exhibit different characteristics that tell us something about the way they are likely to respond to an intervention – men and women, adults and children, underweight and overweight people. The more we know about a population, the more we can describe and understand its heterogeneity, and the more accurate our understanding/prediction of an intervention's impact can be.

An exploration of heterogeneity enables decision makers to consider whether interventions should be made available to total populations or specific groups within a population. It means different decisions can be made for different groups of individuals, leading to potentially more cost-effective allocations of resources and better outcomes.

Relevance/value for VfM:

This is important in the context of VfM because it can easily be the case that a programme is good value for money when targeted at specific groups of people, but not when targeted at others. An overall assessment of VfM without a discussion of the heterogeneity of the population the programme is targeting may undervalue improvements in the VfM that would be possible if the programme was targeted differently.

Current reflection in VfM:

The DFID approach to VfM makes no mention of exploring population heterogeneity within programmes or how the VfM of programmes may differ if targeted at different sub groups. It is stated that development results should be targeted at the poorest, and should include sufficient targeting of women and girls, however this is presented as a moral position rather than one based on increasing VfM.

Feasibility:

A discussion of heterogeneity is possible even if an evidence based description of heterogeneity is not. This should be explicit in DFID's approach to VfM. Not only may this encourage further data collection, but, as suggested in section 3.11, it may assist understanding of the equity implications of programmes.

3.9 Principle nine: Uncertainty

RC: The uncertainty associated with an economic evaluation should be appropriately characterised.

Principle outline:

Because responses to an intervention are variable, even among individuals with similar characteristics, we are uncertain about the impacts of any intervention. Because future direct costs can change and because indirect costs are variable, we are uncertain about the inputs of any intervention. This uncertainty can be characterised, and in doing so information can be revealed that makes economic evaluations more useful. Characterising uncertainty enables decision-makers to make judgements about how confident to be that expected results will materialise, and what may happen if they do not. Uncertainty matters most where unexpected results are likely and cause very different results. It matters less when unexpected results are not likely or do not impact on the results. Characterising this allows decision makers to focus on the important aspects of uncertainty.

Relevance/value for VfM:

This is particularly applicable to *DFID's approach to VfM analysis* because there is so much uncertainty – due to both the low quantity and quality of the available data and the unpredictable future. Lower quantity and quality of evidence makes it more difficult to understand the heterogeneity of a population, thereby increasing the uncertainty we face. In addition, significant uncertainty is caused by a higher probability of shock events such as political instability, disease outbreaks, natural disasters and wars coupled with larger vulnerability in terms of the destruction these events can cause. This significant degree of uncertainty faced during the implementation of DFID programmes means that it is all the more important to characterise the uncertainty of the VfM analysis and results.

Specifically, characterising uncertainty in VfM analysis would enable decision makers to engage with how confident we can be that a programme represents good or bad value for money. If a programme is considered good value for money, but the uncertainty of this advice is characterised and significant, it may be worth investing in further research to inform the decision. The programme may continue in the meantime or may be put on hold depending on the expected opportunity cost of the programme (Revill, Woods, & Sculpher, Forthcoming).

Characterising uncertainty is also relevant for DFID because of the preference donors have for funding capital expenditures and because of the demand for capital expenditure in countries without basic infrastructure. Capital expenditures are irreversible, and where costs are irreversible

there is an interplay between the decision to make expenditures and the decision to conduct further research to decrease uncertainty. Where uncertainty is high, delaying capital expenditure until further research has decreased the uncertainty may generate more benefit overall, because it avoids costly start up and construction expenditure that may ultimately be abandoned if it turns out to the wrong decision (Revill, Woods, & Sculpher, Forthcoming).

Current reflection in VfM:

DFID's approach to VfM currently does not engage with this. It is expressed that analysts should be confident in the strength of evidence, and explicit in stating any underlying assumptions; but how to evaluate confidence in the strength of evidence is not outlined. Moreover, what to do where analysts are not confident, as will commonly be the case, is not mentioned.

Feasibility:

Generalised cost-effectiveness analysis (G-CEA), as used by the WHO-CHOICE programme, attempts to represent uncertainty through stochastic league tables. These tables present the probabilities with which interventions will be part of an optimal set of interventions (benefits package) conditional on available health care budgets (WHO, 2003). For example; it may be concluded that given a health care budget of $100 per capita, there is a 95% chance that the optimal benefits package includes intermittent preventative treatment of malaria during pregnancy. However, this approach demands information on the distribution of costs and effects for all interventions, as well as their correlations both within and across interventions. It also does not offer guidance on the likelihood that a given combination of interventions is optimal. Given the significant (probably impossible) data demands, and dubious benefits, this may not be a cost-effective strategy for characterising uncertainty (Revill, Woods, & Sculpher, Forthcoming).

A simpler alternative could be to include deterministic sensitivity analysis of the indicators. Deterministic sensitivity analysis involves holding all bar one variable constant and assessing the impact of changing the remaining variable (by 5%, 25%, 50% and 100% for example). If changing variables for which we are highly uncertain significantly changes our conclusion regarding VfM, then we conclude that our results are not robust with respect to the uncertainty we face. However, if we are uncertain about something, and changing it does not impact our assessment of VfM, then our conclusions are robust to the uncertainty. This should be explicitly stated in the conclusions.

There are, however, significant problems with deterministic sensitivity analysis (Ades, Claxton, & Sculpher, 2006). Two main problems are that it does not capture correlation between variables (holding all else constant while one variable is manipulated is not realistic) and that different variables have different distributions (so certain errors may be more likely than others). However, at this stage it is not clear how the preferable probabilistic sensitivity analysis could be done given the lack of information available. This is an important area for further research.

Given available tools, deterministic sensitivity analysis is recommended. Not characterising uncertainty at all means that a significant component of any decision DFID makes is ignored. This is feasible and is a useful method for highlighting areas of risk due to uncertainty. However, it is a long way off the methods for characterising uncertainty available and recommended where more data is available (such as probabilistic sensitivity analysis as a method for capturing the propagation of uncertainty when multiple sources of evidence are synthesised). If not treated with caution, for example, without an understanding that correlations between and distributions of variables have not been considered, it will not always generate the most appropriate recommendations for a decision problem.

3.10 Principle ten: Impact on other constraints

RC: The impact of implementing the intervention on the health budget and other constraints should be identified clearly and separately.

Principle outline:

The net total costs of a health intervention should be determined to assess impacts on actual budgets – as only this will determine the total opportunity cost. Cost effectiveness is a measure of costs relative to effects, and says nothing in itself about budget impact, which is why the ability of the health budget to absorb this cost needs to be assessed as well. This is particularly true in low and middle income countries, where health budgets are smaller and it is more conceivable that individual interventions cost significant proportions of total budgets.

From a resource allocation decision making perspective, the reason it is important to measure budget impact is because, while many interventions have only marginal budget implications, some interventions (or programmes) are so expensive that they have a non-marginal impact on the budget constraint. Through displacing a significant amount of what the sector could otherwise provide, after the new intervention is rolled out the threshold at which interventions in the sector are considered cost-effective is different to what is was before the intervention was rolled out. One example of this is the provision of antiretroviral therapy (ART) for HIV/AIDS in many countries in sub-Saharan Africa. The total cost of ART represents such a significant proportion of the total budget constraint that the opportunity cost of marginal expenditure is greater than if ART was not provided (Revill, Woods, & Sculpher, Forthcoming). When this happens, there is a risk that new programmes displace expenditure that is more cost-effective than the new programme.

A second aspect of this principle is that the challenges to health service delivery extend well beyond financing alone, particularly in low and middle income countries and particularly with donor funding. Other constraints, such as human resources and infrastructure are also often in short supply. Implications on non-financial resources should also be stated explicitly.

Relevance/value for VfM:

Both of these aspects are relevant to VfM analysis. Understanding the true opportunity cost of DFID funded programmes demands knowledge of total costs, and how these total costs relate to total budgets – at sector, national and international levels.

Many DFID programmes are not implemented nationwide, and expansion is considered an option. In addition to total programme expenditure, DFID may want to comment on the expected budget implication of expanding a programme to national coverage. If it is expected that the national government will take over the programme, then the implication on government budgets should also be assessed. In addition, given the decision making structure of country offices (which have a certain amount of control over a budget), the implications of programme expenditures should be expressed in terms of total country office budgets as well. This is to give a sense of how the programme sits in relation to total DFID expenditure in the relevant country or region.

The second aspect of the principle is particularly relevant. Financing is not the only resource challenge. Due to the multiple sources of financing present in many of the countries in which DFID works, the opportunity cost of budget expenditure is perhaps the least clear of all the opportunity costs. One of the clearer resource constraints is the human resource stock. Staff working on one programme have a limited ability to work on another. If DFID funds a maternal and child health programme, this is likely to take up a significant amount of time for many of the countries skilled health workforce. This means less time spent on non-reproductive and child health matters, which is a cost that needs to be considered if the true net value of the programme is to be estimated. Thus the cost of non-financial resources used as a result of DFID funded programmes needs to be reflected in DFID's VfM analysis.

Current reflection in VfM:

VfM analysis looks retrospectively at total programme expenditure. However, this is not looked at in relation to any other budget (such as total DFID budget, total DFID in country budget, total

sector budget or total government budget). Because of this, while total costs are known, they are not explicitly put in the context of other programmes.

Moreover, there is no recommendation to explore total budget implications of programme expansion or adoption by a new organisation (maybe the domestic government).

There is also no guidance on how to assess total implications on non-financial resource stocks, such as human resources.

Feasibility:

Crude estimates of programme expansion cost, impact of expenditure on government budgets and opportunity costs of non-financial resources (such as human resources) are all very feasible. Once all assumptions are made clear, the impact of total resource use by programmes on a number of different constraints are feasible.

3.11 Principle eleven: Equity implications

RC: The economic evaluation should explore the equity implications of implementing the intervention.

Principle outline:

Decisions concerning resource allocation in health should also reflect considerations other than efficiency – for example equity. Economic evaluations should incorporate equity implications of an intervention as well as cost-effectiveness. It may be appropriate to prioritise an intervention with a lower cost-effectiveness ratio if it also has greater equity implications. Making the trade-off between efficiency and equity clear enables transparent decisions to be made that reflect social preferences.

Relevance/value for VfM:

Given DFID's stated objective with VfM, analysis of a programme's equity implications is imperative. The premise upon which money is allocated to DFID in the first place is that equity considerations are a legitimate cause for prioritisation.

Current reflection in VfM:

As outlined in section one, DFID's approach to VfM involves addressing economy, efficiency and effectiveness. Equity is then stated, almost as an add-on, without a clear guideline on how to address the trade-off between equity and any of the other three Es. Equity is stated to mean making sure DFID's development results are targeted at the poorest, and include sufficient targeting of women and girls.

Feasibility:

Equity is a difficult phenomenon to include in this sort of analysis because it is expensive to track. Numerous different metrics are available to measure equity, but they demand information on outcomes and impacts linked to information about individual respondents (income, age, gender etc.) i.e. survey data is needed.

Three broad approaches have been applied to health economic evaluation, and may prove fruitful for VfM analysis. The first, and perhaps the most common, is to consider equity as one of a number of broad social value judgements made at policy level, influencing the valuation of health

benefits depending on individual patient characteristics – judging that health benefits are worth more when the patient has a high burden of disease, for example, or is young. The second approach is really a formalisation of the first approach. Multi Criteria Decision Analysis attributes values to efficiency and equity, and aggregates them. It makes the first approach explicit, and may encourage consistency and transparency. The third approach is to attempt to quantify a health loss equivalent to a reduction in inequality. It is essentially to express health benefits and equity benefits in a common measure, and is referred to as distributional cost-effectiveness analysis (Revill, Woods, & Sculpher, Forthcoming).However, none of these options negotiates the problem that monitoring implications on equity is often prohibitively expensive.

A possible option that may be worth developing is to link an evaluation of a programme's equity implications with the discussion of heterogeneity. If the heterogeneity of a target population (or general population) has been properly outlined, then it should be possible to quantify beneficiaries of a programme by different sub groups within a population. Even a crude discussion of this would be a step forward, and may highlight important issues, such as particular groups so far not benefiting from programmes targeted at them.

4 Discussion

This report has explored the opportunities and challenges that come out of a comparison between DFID's approach to VfM analysis and iDSI's RC for economic evaluation. While they are not the same, both guidelines are intended to help generate information useful for maximising the impact of expenditure. For this reason it was considered a worthwhile task.

DFID's approach to VfM is a great step in the right direction. It is accessible online, and logical frameworks, including sections specifically focussing on VfM, are prepared in advance of implementation and then monitored throughout. How much was spent on a programme? What were the goals and objectives? What was achieved? These are all questions that, to a large extent, can now be answered about DFID funded activities. They are crucial first steps in making resource allocation a systematic and accountable process, increasing efficiency and, hopefully, raising the quality of outcomes. However, improvements are always possible, and, based on this report, the RC for economic evaluation is a document that DFID can draw on.

To a large extent, *DFID's Approach to VfM* could be immediately and feasibly updated. First, the scope of relevant evidence should be defined – this is drawn from consideration of a number of the RC principles. VfM analysis is evidence informed, but which evidence is relevant is not prescribed. Which benefits should be included? Which costs should be included? Over what time period should costs and benefits be included, and how should costs and benefits in different years be compared? What about indirect costs and benefits? Or non-financial costs? Making a decision on each of these questions is unavoidable during VfM analysis – it happens whether authors know it or not. However, as there is no guidance, decisions may be made inconsistently and implicitly – decreasing the quality, comparability and usability of the analysis. Authors conducting VfM analysis should be required to actively engage with each of these questions and explicitly justify their answers. In some cases DFID may want to prescribe general standards, but in others it may want country specific answers reflecting individual programme characteristics.

Under the status quo, the heterogeneity of target populations and the uncertainty of the evidence used may be ignored altogether. *DFID's Approach to VfM* makes no mention of either. Authors should be required to describe what is known about the target population heterogeneity, even if that is to say that nothing is known. This is important because further targeting of programmes could improve value for money. It may also be that DFID wants to understand how the target population compares to other populations in case it is worth scaling up a programme or repeating it in a second area, and may contribute to a better understanding of the equity implications of a programme. As mentioned at the beginning of this report, all of these things may in fact already be being done in some instances, but it is not a DFID requirement, and so will not be happening generally.

Characterising uncertainty is particularly important for DFID's VfM analysis because of the huge level of uncertainty faced during the implementation of DFID funded programmes. First, the quality and quantity of information is often low and, second, the probability of shock events with devastating repercussions is high. Any conclusion regarding the VfM of a DFID funded programme is subject to significant uncertainty, and understanding this is necessary if recommendations are to be appropriately drawn. A feasible first step would be to require authors to conduct deterministic sensitivity analysis on key variables and assumptions. This would highlight where unexpected events have the most damaging results and, in turn, where uncertainty is most important.

Requiring that outcomes are presented disaggregated by sector may improve the transparency of trade-offs where no general measure of benefit across sectors is available. At present it is not clear how to incorporate benefits measured in different units into one analysis. This is problematic for two main reasons. First, DFID funds programmes in many different sectors between which measures of benefit are incomparable and, second, many individual programmes have benefits in numerous sectors. Without guidance, there is a risk that some authors will simply not include benefits outside the primary sector of concern, or will incorrectly attempt to make comparisons in a way that is not transparent or accurate.

Methodological specifications currently exist to guide authors on how to meet each of these demands. Where it is not possible because the data is not available, this should be clearly stated, following the tenet that infeasibility of data collection does not mean irrelevance of data. A large part of improving the evidence based nature of VfM analysis should be in being open about what data is missing, and making the most of the data that is there. An updated approach to VfM that addressed each of these issues would provide DFID with more useful analysis to inform its decision making, and in doing so could ultimately lead to better outcomes in the sectors in which it works.

There remain, however, a number of unresolved issues. There is still no decision rule for whether or not a programme represents good VfM, and this reduces the transparency of the process. For reasons discussed in this report, this is a hugely challenging task relating to the lack of a general measure of benefit that DFID aims to generate, but it is also central to monitoring allocative efficiency, and cannot be perennially put off because it is too difficult. If DFID is able to incorporate an assessment of allocative efficiency into its VfM analysis, then the decision rule should be based on a threshold of cost effectiveness that reflects the opportunity cost of displaced expenditure (Revill, et al., 2014). However, even if DFID do not manage to incorporate allocative efficiency, some decision rule needs to be explicit that can be challenged in order to maintain the accountability of the process.

There is also no adequate methodology for incorporating equity considerations into resource allocation decision making. Even in the RC this is specified to a lesser extent that the first ten principles. If equity is to play a significant part in the distribution of DFID money, it needs to outline explicitly how this should be monitored. One avenue recommended by this report is to combine the evaluation of population heterogeneity with equity implications – so beneficiaries of a programme are quantified disaggregated by population sub group identified during the description of heterogeneity – thereby highlighting specific groups not benefiting.

A number of further areas for discussion emerged during informal peer review processes of this report. The first was an expansion on the distinction between evaluating performance ex post (DFID's approach to VfM) or ex ante (the RC). Evaluating performance in retrospect has a number of strengths and weaknesses. The first and most obvious strength is that analysis is attempting to evaluate what has actually happened, rather than attempting to evaluate something in a sample, and assuming it will work to a similar degree in a wider context. Given the high levels of uncertainty, this may lead to more accurate information on both the costs and the benefits of programmes. On the downside, by the time poor VfM is noticed, it has already been bought. The damage from this is limited if there is a smooth link between VfM and planning processes, so that lessons can quickly be incorporated into programme functionality, but it is unclear how well this works.

A second area for further discussion was the extent to which the RC could learn from DFID's attempts to interrogate technical efficiency. The problem with focusing on allocative efficiency is that it does not offer practical guidance to programme implementers on how to improve what they are currently doing. Rather, it focuses in establishing whether or not they are working towards the right outputs. An analysis of technical efficiency identifies specific areas where improvements in VfM are possible, and what needs to be done – for example the substitution of branded for generic drugs. For a policy maker who knows what they want, technical efficiency analysis useful to make sure they are going about getting it in the best way – this is perhaps why is it more commonly used for audit and quality control purposes (Smith, 2009).

Finally, questions were raised as to whether an updated version of the RC could be explicitly tailored to programme evaluations by the addition of principles such as 'assessment of potential for future development', 'training opportunities' or even 'scale and scope'. This sort of question highlights the challenge of capturing all the benefits of a programme. Direct benefits that are part of the logical framework from the onset will probably not accurately capture all of the positive externalities that occur as a result of a programme, just like financial costs will probably not capture all of the negative externalities. Adding these principles to the reference case might help by reminding researchers to consider certain aspects of costs and benefits, but there will always be

extras, and so researchers should in general be encouraged to look beyond the more obvious direct costs and benefits for the otherwise unobserved effects.

5 Recommendations

Recommendations are split into three sections – recommendations for DFID; recommendations for iDSI and the BMGF; and recommendations for future research.

5.1 Recommendations for DFID

Given the current level of technical progress in the field of resource allocation decision making and impact maximisation, DFID should update its approach to VfM analysis and generate information that is more useful in terms of maximising the impact of its expenditure.

Recommendation 1

Assign a team with the responsibility of updating *DFID's Approach to VfM* to reflect the most up to date technical analysis available.

Specifically, feasible improvements are to:

- Include a decision rule for whether or not a programme is considered good value for money at the programme level, prior to VfM analysis being conducted. *From principle one.*
- State the perspective from which the VfM analysis is being done: DFID, DFID country office, national government, etc. *From principle one*.
- Define the scope of relevant evidence. This means making explicit which costs and which benefits are relevant to an assessment of VfM – which direct, indirect, future, past, target and non-target costs and benefits should be included? Part of this means defining more clearly the perspective from which VfM should be done. Is it about good value for DFID? For the development sector? For society as a whole? Each has different implications for the relevant scope of evidence. *From principle three.*
- Be explicit that infeasibility of data collection does not mean irrelevance of data. Data that is relevant but not available should be made explicit (labelled as missing), and the uncertainty generated by missing data should be characterised. *From principle three and nine.*
- Impacts should be routinely measured in the most generalizable unit for their sector and reported on disaggregated by sector/measure. If authors feel that comparison between sector specific benefits is possible, this may be done in addition. *From principle four and seven.*
- Define the scope of relevant costs. In general, all costs of the programme not also faced by comparators programmes should be included. *From principle five.*
- Clarify whether costs incurred by other organisations should be included or not. This decision should be based on intended role of DFID in driving improved resource allocation of its own investments, of recipient country investments, and of its partner organisation investments. *From principle five.*
- Offer guidance on the incorporation of future costs and benefits, and how to inflate and deflate costs and benefits to reflect their present value. *From principle six.*
- The heterogeneity of target populations should be described. This is to explore whether the VfM of programmes could be increased if programmes were further targeted. It may also inform decisions to expand programmes or repeat them in new locations, and contribute to the evaluation of a programme's equity implications. *From principle eight.*
- Require that the uncertainty of the conclusions due to the low quality and quantity of data used to inform them be characterised. At least deterministic parameter sensitivity analysis should be done on all key assumptions, as well as evidence informed variables for which

there is considered to be a high degree of uncertainty. This is to highlight where conclusions are least robust. *From principle nine.*

- Require that the implications of total programme costs on all relevant budgets be expressed. This should at least include DFID's in country budget. Where handover to another organisation is possible (particularly to a domestic government) the implication on that organisation's budget should also be presented. *From principle ten.*
- Require that the implications of programmes on non-financial constraints (such as the stock of skilled labour) be presented. *From principle ten.*
- Provide clarity on how equity implications of programmes should be presented in VfM analysis. One suggestion is to present the outputs and outcomes by the different groups identified during the discussion of target population heterogeneity. This would enable discussion of the programme's impact on different groups within a society. *From principle eleven.*

Defining guidelines on each of these issues would increase the quality, consistency and usability of DFID's VfM analysis, and allow more accurate estimations of the technical efficiency of programs.

Recommendation 2

Expand the remit of VfM analysis from technical efficiency to both technical and allocative efficiency.

This means:

- Making comparisons between the total costs, total benefits and cost to benefit ratios of appropriate comparator programmes.
- Where an appropriate comparator is not available, include comparisons between individual components/interventions of programmes.
- Plan programmes with difference in difference analysis in mind. Identify comparator states/districts from the start. Offer guidance on this in VfM guidelines.

Recommendation 3

Outside of VfM analysis of funded programmes, require that funded entities like Global Fund and GAVI use the RC to guide their analysis when deciding best buys and investments.

5.2 Recommendations for iDSI and the BMGF

The RC is a clear, but it is so far targeted at health economic evaluations and does not capture all of the specifications that can best inform decision making.

Recommendation 1

The reference case should be more easily applicable to complex programme evaluation and other investment decision making tools used by policy makers beyond full economic evaluation – currently it is designed with health economic evaluation of individual interventions in mind. This report has shown how it can be relevant and useful when assessing the standard of VfM analysis done by DFID across a wide range of sectors, not just health. It may also be useful for a range of

other methodologies across the wide spectrum of analysis that is designed to inform policy maker's decisions.

This means:

- Consult with donors and governments on the current decision making processes. From this, attempt to make future iterations of the RC clearly relevant to a wide range of decision making tools and processes – from guidelines for data intensive economic evaluation to ten simple questions for policy discussion.

Recommendation 2

Specify in greater detail how equity can be factored into transparent resource allocation and impact maximisation.

This means:

- Offer guidance on how equity can be quantified, particularly in low income settings.
- Offer guidance on how trade-offs between technical or allocative efficiency and equity can be made in context specific but transparent ways.

5.3 Recommendations for future research

Adherence to each of the 11 principles could be improved if analytical techniques and available data were better. This is particularly true in low income settings. Future research is necessary across the board, but is particularly pressing in a number of areas.

Recommendation 1

Research into a generalizable measure of benefit across the many dimensions of livelihood could enable consistent and transparent trade-offs between the many sectors in which organisations such as DFID work. As discussed, work on this is well on its way – in particular the various iterations of the Human Development Index, the Multidimensional Poverty Index and survey based rankings of different states and dimensions of poverty. However, these techniques are still far from satisfactory measures that can be affordably tracked as part of regular programme evaluation.

There is also a risk that looking for general measures of benefit on a global scale to be too rigid. The measures used in the UK health sector (that the survey based rankings have been based on) reflect the preferences of the population in England and Wales and are used to inform allocation decisions in England and Wales. Should such work on poverty look to generate different rankings based on different preferences of people all around the world? Conceivably preferences over different states of poverty are culturally varied? If people in Solomon Islands want different things to people in Liberia, is a general global measure of benefit actually appropriate?

Recommendation 2

A significant problem with evidence informed analysis in low income countries is the low quality and quantity of data available. Improved methods for characterising the uncertainty this generates are crucial for understanding what we can actually learn from the data. The deterministic sensitivity analysis that is recommended in this report is crude. It does not take into account the distribution of variables or the correlation between variables, and so, while it is an improvement on no sensitivity analysis, it is a long way from perfect. More comprehensive techniques, such as probabilistic sensitivity analysis, are generally too data demanding. A middle ground is needed – a transparent method for making the most out of the low quantity and quality of data available. This is currently

part of iDSI's research agenda, and progress on the question may significantly enhance the usefulness of evidence in low and middle income countries.

Recommendation 3

Equity is generally considered to be a legitimate reason for prioritisation, but specific methods for doing this are not available. Consequently, a brief discussion of equity is often included at the end of an analysis, merely as an add-on discussion. Methods for tracking and quantifying equity, specifically in low income settings, would enable DFID to systematically incorporate equity into its measures of VfM. Lots of work is done using concentration indices in richer countries, but less has been done using these methods in low income countries. Further research should be done on quantifying equity in low income settings. Again, iDSI is working to address this, and progress in the field could significantly improve the quality and usefulness of potential analysis in low and middle income countries.

References

Ades, A., Claxton, K., & Sculpher, M. (2006). Evidence synthesis, parameter correlation and probabilistic sensitivity analysis. *Health Economics*.

Alkire, S., Conconi, A., & Seth, S. (2014). *Multidimensional Poverty Index 2014: Brief Methodological Note and Results.*

DFID. (2011). *DFID's approach to Value for Money.*

Fenwick, E., Claxton, K., & Sculpher, M. (2008). The value of implementation and the value of information: combined and uneven development.

HM Treasury. (2014). *Public Spending Statistics Release: November 2014 - part of HMT Public Spending Statistics.*

iDSI. (2014). *Methods for Economic Evaluation Project.* BMGF.

Melamed, C. (2011). *Does development give poor people what they want?*

OECD. (2010). *Health care systems: Getting more value for money.* OECD.

OPM. (2013). *Consultancy for literature review for measuring value for money in health.*

Revill, P., Walker, P., Madan, J., Ciaranello, A., Mwase, T., Gibb, D., . . . Schulpher, M. (2014). *Using cost-effectiveness thresholds to determine value for money in low- and middle-income country healthcare systems: Are current international norms fit for purpose?* CHE Research Paper 98.

Revill, P., Woods, B., & Sculpher, M. (Forthcoming). Economic evaluation of healthcare programmes and interventions: applications to low and middle income countries. In *Global Health Economics and Public Policy.* World Scientific.

Santatiwongchai, B., Chantarastapornchit, V., Wilkinson, T., Thiboonboon, K., Rattanavipapong, W., Walker, D. G., . . . Teerawattananon, Y. (2015). Methodological Variation in Economic Evaluations Conducted in Low- and Middle Income Countries: Information for Reference Case Development. *PLoS ONE*.

Smith, P. (2009). *Measuring value for Money in Healthcare: concepts and tools.* QQUIP.

USAID. (2012). *The health system assessment approach: a how-to manual.* USAID.

WHO. (2000). *World Health Report - Health Systems: Improving Performance.*

WHO. (2003). *WHO guide to cost effectiveness analysis.*

WHO. (2007). *Strengthening health systems to improve health outcomes.* WHO.