Using the Lives Saved Tool to estimate the number of maternal and newborn lives saved by DFID programming: 2011–2015

Technical Note: Version 4

Ingrid K. Friberg[a,b], Angela Baschieri[c], Jo Abbotts[d]

[a] Johns Hopkins Bloomberg School of Public Health, Baltimore, USA

[b] Norwegian Institute of Public Health, Oslo, Norway

[c] University of Southampton, UK

[d] Department for International Development, East Kilbride, UK

Contents

Abbreviations

AIM: AIDS Impact Module
ANC: antenatal care
BEmONC: basic emergency obstetric and newborn care
CEmONC: comprehensive emergency obstetric and newborn care
CHERG: Child Health Epidemiology Reference Group
DFID: Department for International Development, Government of the United Kingdom
DHS: Demographic and Health Survey
E4A: Evidence for Action
FamPlan: Family Planning Module
FP: family planning
GAVI: Global Alliance for Vaccines and Immunization
IGME: Interagency Group on Mortality Estimation
IMR: infant mortality rate
IPTp: intermittent preventive treatment of malaria in pregnancy
ITN: insecticide treated bednet
JHBSPH: Johns Hopkins Bloomberg School of Public Health
LiST: Lives Saved Tool
MBB: Marginal Bottlenecks for Budgeting
MICS: Multiple Indicator Cluster Survey
MiH: Making it Happen
MMR: maternal mortality ratio
MNCH: maternal, neonatal and child health
MSI: Marie Stopes International
NMR: neonatal mortality rate
OHT: OneHealth Tool
PUMDUP: Prevention of Maternal Death from Unwanted Pregnancy
PPA: Program partnership arrangement
UK: United Kingdom
UNFPA: United Nations Population Fund
UNICEF: United Nations Children's Fund
U5MR: under-5 mortality rate

Introduction

In 2010, the United Kingdom (UK) Government Department for International Development (DFID) announced their 'Framework for results for reproductive, maternal and newborn health (RMNH)' which included several ambitious goals in relation to reproductive, maternal and newborn health including saving 250,000 newborn and 50,000 maternal lives by 2015 (1). DFID simultaneously committed to measuring its achievements in reaching these goals.

DFID supports health progress in developing countries through a number of different aid modalities. These include direct budget support to the government (general funds or health funds), bilateral/multi-donor programs, regional programs, support to multilateral organizations (such as UNFPA and GAVI) and program partnership arrangements (PPA) with civil society organizations.

Currently, there are no easy cost-effective methods available to directly measure the lives saved from the wide variety of types of health programs and funding streams. The most accurate method would be to directly measure mortality through before and after surveillance in the area where each individual program is implemented. However, this is time consuming, expensive and not feasible for all projects, especially those working at the national level or through budget support. There is also the possibility of double-counting the lives saved when programs overlap in place or time or over-ascribing impact to DFID programming when other programs exist in the area. In addition, measurement of maternal mortality at adequate precision to see statistically significant differences requires a very large population (100,000 or more pregnant women) that few programs reach.

Alternatively, child mortality rates or maternal mortality ratios can be collected through existing national surveys. They are limited in that most priority countries will not have such surveys at the exact points in time (2010 and 2015) that are necessary to estimate these impacts for the DFID results framework. In addition, these surveys are unable to adequately identify DFID programming impacts separately from non-health and non-DFID activities.

Selection of the modelling tool

Given the limitations of the aforementioned direct measurement methods, modelling has the potential to estimate lives saved. Tools considered included the Lives Saved Tool (*LiST*)(2), UNICEF's Marginal Budgeting for Bottlenecks (MBB)(3), the One Health Tool (OHT)(4), and the Marie Stopes International (MSI) Impact Calculator(5). These are discussed in turn, in relation to their potential benefits and limitations. An additional tool is available, called MANDATE, however this was not formally considered as its focus was intended to be on new technologies rather than existing ones(6).

LiST provides a structured format for users to combine the best scientific information about effectiveness of interventions for maternal, foetal, neonatal and child health with country specific information about cause of death and current coverage of interventions to inform their planning and decision-making as well as to help prioritize investments and evaluate existing programs(7). *LiST* utilizes publically available information on demography, family planning, HIV (incidence as well as preventions and treatments) and coverage of health interventions to estimate the number of lives saved by changes in these characteristics. It is built into the freely available Spectrum Policy Modelling Software(8), and has explicit linkages to Spectrum's AIDS Impact (AIM) and Family Planning (FamPlan) modules(9). *LiST* is the result of more than 10 years of work by the Child Health Epidemiology Reference Group (CHERG) for WHO and UNICEF and collaborators, who have completed a series of systematic reviews with a consistent methodology (10) on the effectiveness of interventions that impact newborn and maternal (and child and foetal) mortality. This body of work became the basis for the development of a systematic way of combining current knowledge of effective interventions into a single package. The CHERG published their work in the Lancet Series on Child Mortality (2003)(11, 12), Neonatal Mortality (2005(13) and 2014(14)), Nutrition (2008(15) and 2013(16)), and Stillbirths (2011)(17, 18). In addition, there have been three supplements which included updates and additional data on effectiveness of interventions (IJE 2010, BMC Public Health 2011, BMC Public Health 2013). Where the evidence of effectiveness of interventions was not available from robust epidemiological assessments, the CHERG applied the Delphi method – a formalized series of expert consultations – to generate an estimate of the required effectiveness parameters.

MBB, the second modelling tool considered, uses a similar methodology to calculate the lives saved, and derived from the 2003 Child Survival Series, as did *LiST*. The effect sizes have not been formally updated, but MBB can use the newest *LiST* estimates if needed. It requires extensive information on delivery channels (e.g. outreach, community, clinic, hospital) and costs as well as health intervention coverage at these channels, in order to work well. MBB also requires a significant investment in time to accurately model individual programs rather than a comprehensive health system since it models health capacity as well as delivered interventions. Thus, MBB is ideal for systems level analyses rather than sub-system level. It is also located within an Excel spreadsheet, making it relatively easy to see all of the elements, but can also be perceived to be difficult to work through each of the steps due to its great size.

The third option, OHT is based within the Spectrum software and uses *LiST* (and AIM and FamPlan) for the impact portion of its calculations. However, it has additional features, including costing and health systems, requiring significant additional inputs, similar to MBB. Again, it would require a significant investment of time to be able to do a single program analysis rather than the entire health system. In addition, this tool is somewhat less fully

developed than *LiST* or MBB, although it is already being used in countries and is supported by, and being developed by, the World Health Organization, among others.

These three tools are similar in that each uses a combination of health interventions, effect sizes and underlying mortality to estimate lives saved. Each has a situation in which they are more appropriate. Both MBB and OHT have bottleneck analyses which are useful for planning purposes. If one is interested in the costs associated with the impacts, then either MBB or OHT would be more appropriate than *LiST*, although *LiST* does have the ability to look at marginal costs or specific intervention costs while excluding the costs of health system changes that might be needed. If evidence emerged that the effectiveness of interventions differed by level of care (e.g. community, clinic, hospital) then all three of these tools would need to be modified to improve their estimates of effectiveness.

LiST is somewhat more nimble than MBB in that it primarily requires health intervention coverage, and does not require further disaggregation into the exact location of delivery (which may not be available). Nearly all of *LiST*'s effectiveness estimates have been updated since 2010 (and can even be used as inputs to MBB if necessary, essentially updating the default 2003 Child Survival Series effect sizes). Additionally, *LiST* is a quasi-cohort model, which allows for interactions between interventions. That means, if a child is saved from dying of a neonatal cause, then that child is not saved for the rest of that child's life, but is at population risk of dying of other causes; the child that is saved has the same risk of dying of the later cause as all other children(19). This feature somewhat mimics the reality of life in which interventions delivered earlier in life can have unintended consequences. Another benefit of the quasi-cohort model within *LiST* is that it allows the model to include risks associated with nutritional deficiencies.

The MSI Calculator is completely different and has a more comprehensive method of modelling impacts related to reproductive health. However, it is limited only to reproductive health interventions. It does not use effect sizes for calculating impact of interventions on mortality and other outcomes. It instead uses correlations to calculate impact rather than directly modelling the causal pathways as used by the other software tools.

No other tools have been identified which assess sector-wide mortality. Other tools have been developed by WHO (e.g.. CHOICE, et al.) which consider just costs, and not mortality, or just look at delivery care, but these are not currently in mainstream usage and were not considered. It should be noted that these tools are those which assess health impacts on mortality. They do not assess or include estimates of non-health related inputs to mortality reduction, such as changes in socio-economic status, education of women, secular trends in mortality reduction. Also, there are currently no tools available which can model health systems strengthening interventions directly; rather they can model their outputs if they are known.

Table 1. Summary of major similarities and differences between the software tools considered.

Topic	Lives Saved Tool (*LiST*)	Marginal Bottlenecks for Budgeting (MBB)	One Health Tool (OHT)	Marie Stopes Calculator (MSI)
Method of estimating Lives Saved	Interventions avert cause-specific mortality	Interventions avert cause-specific mortality	Interventions avert cause-specific mortality	Uses correlations of interventions with mortality
Effect sizes	Effect sizes updated regularly	Primarily uses 2003 effect sizes, or links directly to *LiST*	Uses *LiST* directly	Updated irregularly
Model type	Static quasi cohort model	Static model	Quasi cohort model	Static model
Additional features/ limitations/ discussion points	Has a separate costing module optional	Excel based	Uses additional health systems features	Focuses only on reproductive health interventions

After consulting with experts, DFID decided that the most appropriate choice for estimating its results on maternal and newborn lives saved was *LiST*. If additional tools become available before the end of the analysis period which afford the ability to look across the spectrum of health interventions, these will be considered for later comparative analyses.

Use of modelling

Modelling is a contentious tool within the global health monitoring and evaluation toolbox. It does evade the aforementioned issues with direct data collection, but has its own critical limitations which must be considered, and will be discussed below. Most importantly, modelling should not be allowed to replace critical thinking or to replace actual feasible data collection. This use of modelling is responding to DFID's two specific goals. The first was to assess whether or not the funding being appropriated and delivered to global health issues was adequate to the task of reducing mortality as desired. The second was to support DFID in identifying areas which could potentially benefit from additional DFID support.

A general issue with attribution relates to the knowledge gap. That is, not only are there multiple programs which co-exist, making attribution difficult, but there are multiple unknown programs which can affect health. This is in addition to activities which are non-health related having an impact. These can include education, poverty, agriculture, infrastructure, etc. There will also be unexplained synergies and antagonisms between all of these types of programs, as well as secular trends which, although they include the above issues, may also include many additional issues that are unknown and thus unquantifiable. As a result, when any attribution is done, it inevitably will only refer to those activities which can be measured, which is only a

small subset of the reality. The theory under which *LiST* works is that many of these interventions and changes which cannot be modelled will be accounted for indirectly through the requirement that they eventually affect (albeit indirectly) interventions which can be modelled.

As a direct result of these issues, (such as multiple partners, the knowledge gap, and additional related ideas) the results which will be discussed here should be termed as 'contributions' rather than attributions. That is, the results which will be derived from this exercise do not reflect what DFID has actually done. However, it does reflect what is likely that DFID has helped to support in the context of everything else that has occurred in the community and society. Further expansion upon the limitations of these type of analyses are discussed at the end of this document.

This technical note describes how DFID has chosen to use *LiST* to estimate lives saved from its bilateral health-related programs. The purpose of this note is to explain to internal DFID staff as well as external partners and the public the methodology used to support the calculations required by the Framework for results for RMNH and the limitations associated with it. It is aimed primarily at a specialist audience, and is one of three publications describing the methodology. An article for submission to an academic journal and a short overview for a non-specialist audience has also been prepared. The results will be published separately at the end of the project. Separate methodologies are being developed for including additional results from multi-country programs in countries where DFID does not have a bilateral health program. They are not described here, except in brief, since they have not yet been finalized.

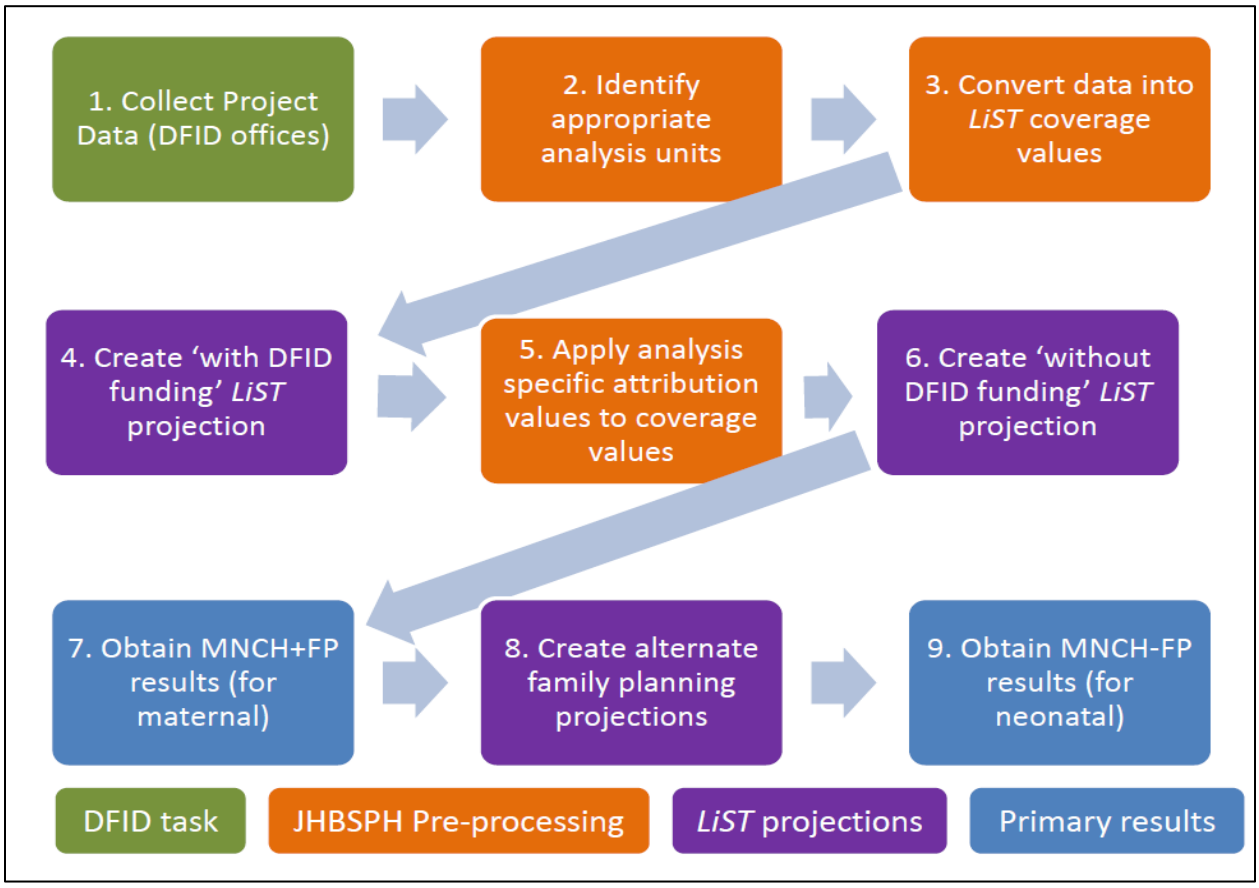
Methods

Modelling maternal and newborn lives saved between 2011 and 2015 is undertaken annually using the most up-to-date version of the *LiST* software, and the latest program data from DFID country offices relating to health or family planning. The DFID focus countries were all invited to participate in this analysis, with 23 participating in the 2014–2015 analysis round, the third year of the analysis. For a list of the specific countries participating, by year, please see Appendix A. For a list of the Spectrum versions which were used in each round, please see Appendix B. Details of the health interventions which are modelled by *LiST* are available in Appendix C.

Data on DFID attribution relating to each set of targets/achieved values are used to estimate what percentage of the program results are due to DFID funds when specific program data were not available. From these data, and the existing baseline data, two *LiST* projections were created (generally termed 'observed' or 'with DFID support' and 'hypothetical' or 'without DFID support', to reflect what occurred with DFID funding and what might have occurred

without it), the difference between which was considered to be the impact of health and family planning supported by DFID. These were summed across all programs to get a national value, and then across all countries to get a global estimate. Additional analyses were created to exclude the impact of family planning. A flow chart of tasks is in Figure 1. Each stage is explained in detail below.

Figure 1. Flow chart of major activities to generate maternal and neonatal lives saved for DFID



Data Collection Methods

Collect project data (DFID offices)

A data collection spreadsheet was created by the analysis team in consultation with health and statistics advisers from DFID country offices. It required three main pieces of information: 1) the complete list of programs in each country relating to health or family planning delivered between 2010 and 2015, 2) the populations targeted or reached by each of the programs, and 3) the interventions (or inputs) delivered, and targets or achieved results, for each year.

The requirement for population data ensured that the analysts knew the size of the geographical area served by the program. It was also used if the program used population estimates for planning purposes which were not based upon the United Nations Population Division medium variant population projections. This was needed to correctly estimate the underlying population upon which any program would be acting, and would include information to also help predict the number of living children and the number of pregnancies. Depending on the program, the population information could be total population, total women, total pregnancies, etc.

Country offices provided data on DFID attribution relating to each set of inputs, targets or achieved values. This enabled estimation of the percentage of program results which could be ascribed to DFID.

Identify appropriate analysis populations

Each project was assessed for three primary characteristics for a given calendar year: funding type (bilateral, multilateral…), geographic location (state/district/region within a country), and data available on the indicators. We then combined the individual projects' information as appropriate into *LiST* analyses ensuring that we did not double count the coverage estimates or their impacts. The exact combination of projects into analyses is available in each individual annual report, and can change from year to year, based upon how the data are reported.

Funding Types

DFID's bilateral programs were analysed using *LiST*, including both general or health sector budget support and sole and multi-donor programs. Programs funded via other aid modalities were generally excluded to avoid the likelihood of double counting the benefits, although some multilateral programs were included in specific countries where it was certain that the same interventions were not already being modelled. For example, funds given to GAVI support vaccination, while funds given to a government may also be used to support vaccination. Both may be paying for a part of the same vaccinations and would result in an over-estimate of lives saved if both were modelled independently and added together. These types of programs are being analysed separately with a different methodology and are not being considered here directly.

Combining data from individual programs into analyses

Projects were considered to overlap under a variety of circumstances. They could overlap based on location, such as two programs working in the same defined area on different interventions, where combined modelling is required since the saving of lives from one set of interventions could affect the second set of interventions. In this scenario, multiple programs would be combined into a single analysis. Alternatively, two projects could overlap on both

intervention and area. For example, a budget support program may consider delivery of ITNs as part of their program, since they are distributed at ANC clinics. Another DFID program might be responsible for purchasing all of the nets nationally. If both projects were modelled together, then each net would be counted twice. In this case, the program which purchased the nets would get the credit, but not the budget support program. All other activities relating to ANC would be considered as part of the budget support program and only the ITNs would be excluded.

Analysis generation

Once all of the projects for a country had been assessed for funding type, temporal and regional overlap and intervention overlap, a list of the analyses to be performed was created. There could be either more or fewer analyses than programs. For example, there could be two budget support programs, which when combined resulted in one national analysis. Alternatively, one program could focus on two regions with different baseline and target values and different underlying epidemiology and mortality rates. In this case, there would be two analyses, one for each region. This was done to minimize the number of analyses needed while also increasing the quality of the underlying data being used. At the end, the number of projects and the number of analyses in each country was summarized. The year 3 summary is below in Table 2. A key point to note is the drop-off between the projects reported and those analysed. Many projects without appropriate indicators, and many that are health systems strengthening types of projects could not be modelled.

Table 2. Comparison of data available for and included in the analysis

	Countries Reporting (N)	Analyses (N)	Projects Analysed	Projects Reported
2010–2011	19	40	87	127
2010–2012	20 (19)*	47	113	146
2010–2013	23	53	146	178

*One country reported but was unable to be analysed.

For example, in one country, there were two projects in which ITN coverage was an indicator of interest. The coverage data being used by each project was the same (e.g. the national targets), while the geographic location was different (one national and one provincial). We would not analyse these two projects exactly as described since the assumption is that some of the bed nets mentioned in the provincial area are the same as those mentioned in the national program. The project would be split into one provincial analysis and one which covered the remainder of the country. This ensured that the provincial nets were not counted twice.

Convert data into *LiST* coverage values

The *LiST* model requires that data on each health intervention be in the format of a coverage value, ranging from 0% to 100%. Thus, to be included, each analysis was converted

into a coverage value if not already reported as such for each calendar year of the analysis (2011–2015) as well as the baseline (2010) for insertion into a *LiST* projection.

Baseline coverage data (2010)

Whenever possible, project-specific data were used for baselines. Often countries reported their most recent national survey data. However in some cases, the underlying baseline data were not included; only the achieved/forecast values/targets might be available. In those cases, we used the best available data on coverage of health interventions to fill in the gaps. Typically, these came from Demographic and Health surveys (DHS) or Multiple Indicator Cluster Surveys (MICS). These surveys have data at both the national level as well as some sub-national levels. For example, in Pakistan, there is quality data available at the provincial level. Whenever appropriate, we used the sub-national data for the analysis, although this was often not possible for very small scale programs. In addition, if there were two surveys, one prior to 2010 and one after 2010, a linear interpolation between the values was used to generate the most likely 2010 estimate.

Intervention data type

Two types of data were available for each health intervention: coverage or numbers delivered. Coverage refers to the percentage of the population in need that receive the health intervention. For example, 95% coverage of measles vaccine indicates that 95% of all children 12–23 months of age have been vaccinated with measles vaccine. This was a common data type for budget support programs where the forecast values were actually national government targets of utilization. Coverage values were used directly as inputs into *LiST*.

Numbers of health interventions delivered was also a typical coverage indicator. Numbers were converted into a coverage value for use in *LiST* by dividing the annual value by the underlying population. One example is for a program reporting that it delivered 3.2 million ITNs. The population in need of ITNs is households. Thus, the number of nets distributed was divided by the number of households to get the coverage value which could then be entered into *LiST*. If multiple programs reported their data in numbers, and these were clearly unique numbers, then the values were added together to generate the total number delivered prior to calculating the coverage by dividing by the underlying population in need.

The underlying populations in need were generated from the default *LiST* projections when needed. The definition of population in need is the entire set of individuals who need an intervention, regardless of whether they are getting it currently or not. For example, the underlying population in need of a facility delivery is the number of births. There are default data within Spectrum which allow the projection of populations into the future based on the 2012 UN Population Division projections. From these default projections, underlying populations can be obtained. These include the number of births, pregnancies, children under

5, women of reproductive age and total population, for use as potential denominators. The total population can be divided by the most recent DHS estimates of the household size to estimate the number of households.

It was assumed that whenever supplies were reported as delivered, that they were actually received by beneficiaries. This is a notable assumption since in some cases it is known that the supplied item never left the in-country warehouses. In addition, wastage is a common description of the percentage of a supply that does not get used. To correctly assess wastage, rates would be needed for each individual intervention, project and country. These assumptions related to supply risk overestimating lives saved in the modelled results.

Not all programs included activities which could be modelled in *LiST*. Programs that reported only on knowledge shifts were typically not able to be included since there were no outcome data available. The exception was when handwashing or breastfeeding education programs were done and there was a known number of people reached. This was considered to be different than a program to improve the skills of a physician in a facility, with no clear explanation of the content, duration or impact of the program. Similarly, programs which only reported on adolescent impacts were also excluded. However, if adolescent programs reported on activities which could be modelled, such as family planning, then they were analysed although *LiST* does not automatically assume increased risk of maternal or neonatal mortality for adolescent mothers. If a program included at least one activity which could be modelled in *LiST*, it was included in the analysis. The effect of excluding certain activities is to risk underestimating lives saved.

Specific interventions

A list of all the interventions which were available in Spectrum 5.31, used at the time of the creation of this note, is available in Appendix C to this document. These are unlikely to change dramatically over time, while the effect sizes can vary from year to year as new data are available. In brief, the indicators which are included were identified as key in the Lancet Child Survival Series (2003), the Lancet Neonatal Series' (2005, 2013), the Lancet Nutrition Series (2008, 2014) the Lancet Stillbirth Series (2011) as well as from supplements published in BMC Public Health and the International Journal of Epidemiology, and other published sources, if needed. Although the majority of the data are peer reviewed, the quality of the effect estimates varies; some estimates are from Delphi studies, while there is equipoise on several estimates and they are consequently modified relatively frequently as new information becomes available. Several indicators required special consideration as described below. In all cases, if the values reported by the DFID country office were not reasonable for the ideal indicator and an alternative was available, the alternative was used and noted.

Insecticide treated bed nets (ITNs)

It was assumed that any ITNs delivered would be effective for 3 years(20), the average duration of the insecticide. In addition, to calculate the underlying bednet coverage, we assumed that the existing stock would be phased out over three years. That is, that one third of the existing nets would be not useful after each of three years, which assumed that they had been delivered evenly over three years prior. This assumption is probably incorrect, but since this assumption would be relevant for both the 'with DFID support and 'without DFID support scenarios, the impact on deaths averted is negligible. We only used the 'sleeping under an ITN' indicator if the 'ownership' indicator was not available.

Intermittent preventive treatment of malaria during pregnancy or sleeping under an ITN (IPTp)

The ideal indicator is a composite indicator, where either IPTp (2 doses in pregnancy) or a pregnant woman sleeping under an ITN is considered protective. Unfortunately, this was rarely available as a standard value and it is typically incorrect to add the two together since they are often being used in tandem by pregnant women. Whenever the ideal indicator was not available, but both of the individual values were available, we chose to use the higher of the two as our baseline value. The effect of this assumption is to risk underestimating lives saved.

In some countries, there were only data on the number of ITNs delivered to pregnant women. We added these additional nets to whichever appropriate baseline indicator was used. We used the same methods of decreasing coverage over three years, as described above, to estimate the duration of protection, although this risks overestimating lives saved because a woman is not continuously pregnant during that time period.

Antenatal care

The standard *LiST* indicator for antenatal care (ANC) is 4 or more visits to a health professional during pregnancy. Occasionally a non-standard 2 or more visit indicator was shared by the programs. In that case, we converted the 2+ indicator into a 4+ indicator, by assuming that the ratio between 2+ and 4+ was the unchanged from that seen in the most recent national survey. If there were no data on the 4+ indicator available in any situation, we simply used the 2+ indicator instead, which will result in a slight overestimate of the impact. This decision will be re-evaluated each year that the analysis is performed.

Breastfeeding

Typical indicators for breastfeeding programs do not match those within *LiST* since *LiST* attempts to match the risk of mortality by age groups and by levels of breastfeeding. Thus, programs often have an indicator of whether or not children less than 6 months of age are exclusively breastfeeding while *LiST* requires inputs for neonates separately from 1-5 month old children as well as for, in addition to exclusive breastfeeding, predominant, partial and no

breastfeeding. Each combination of age and breastfeeding type has a different mortality risk profile. Converting the available data into *LiST* coverage information is somewhat subjective. In reality although a single value is reported, it reflects the natural high initiation of breastfeeding and the reduction and dropouts over time to a lower value by the time the child is 6 months of age. Because of the natural history of this behaviour, when we applied the reported target to both age groups, frequently, the 0-1 month age group had already achieved the 0-6 month value. Thus, in these situations where the reported value for breastfeeding from 0-6 months of age is lower than the available data on neonatal breastfeeding rates, we could only apply the observed change to the children in the 1-5 month old category. The implications of that are that it is likely that in some countries, the impact of breastfeeding on neonatal mortality is being underestimated or completely ignored while the number of deaths averted in the 1-5 month category may be overestimated. Also, we had to determine which non-ideal category of breastfeeding was reduced as the exclusive was increased. This was done on a program-by-program basis, based upon the type of data available. If no data were available, we would assume the most conservative possible result, which was to assume that predominant breastfeeding was reduced and exclusive was increased. See the help menus internal to *LiST* for further details of the breastfeeding calculations.

Infrastructure

Several programs related to infrastructure were also modelled. These included improved sanitation, improved water, and quality improvements related to facilities for childbirth. Whenever the intervention delivered was reported in terms of infrastructure built, we assumed that this was a fixed intervention which had impacts which continued and remained through the entire analysis period (2011–2015). Thus, if the data were in terms of population covered in a given year, we assumed that that population number (not the percentage) was also covered in subsequent years.

Delivery care is notably complex. *LiST* makes assumptions about the proportion of facility deliveries which take place in very basic facilities, facilities that can theoretically perform all of the basic emergency obstetric and newborn care (BEmONC) functions and those that can theoretically perform all of the comprehensive emergency obstetric and newborn care (CEmONC) functions based upon the overall proportion of facility births (out of all births). These default assumptions are detailed in the *LiST* software help menus. One key assumption is that even facilities that are considered to be BEmONC or CEmONC do not always actually perform all of the expected signal functions. This assumption of sub-optimal quality is standard in *LiST*, and should be explored where evidence is available. Programs on childbirth care can vary greatly. Thus the modelling for these programs will also vary. One program might be a demand generation program, and the overall facility birth rate is increased. In that case, no adjustments to the proportions of types of facilities will be made, assuming that no changes in quality of

care will automatically occur. Another program might be upgrading existing facilities to BEmONC level quality (or upgrading all BEmONC-type facilities to actually provide all of the signal functions). In those situations, the overall percentage of facility births is not going to change. However, the quality (and in *LiST*, the type of facility or the specific interventions within the facility) will change. The decisions on exactly how to model this change in quality will depend on both the data as well as any additional information shared by the DFID country offices.

Family planning

The modelling of family planning occurs within the FamPlan module of Spectrum and not within *LiST* itself. Family planning must be considered for every single analysis since changes in family planning rates, or contraception, will affect the total fertility rate, with downstream effects on the number of births, pregnancies and ultimately the number of deaths and lives saved. For countries in which there was no family planning activity, the underlying total fertility rate and thus family planning levels needed to be determined. In countries with family planning activities, two questions needed to be assessed. The first was 'how has the national contraceptive prevalence rate changed because of a program implementing family planning activities?' The second was 'how has the method mix, or proportion of modern/traditional methods, changed over time as a result of the program?'

When family planning was part of a DFID country office program, contraceptive prevalence was obtained from of DFID's twice yearly results return. Contraceptive prevalence rate forecast values were considered to be those which DFID supported, while the attribution reported was used to calculate the alternative contraceptive prevalence rate for modelling without DFID support. If contraceptive prevalence was not reported there, then values from the DFID country office data collection template were used. If neither was available, then the default values within Spectrum were used. The defaults within Spectrum are based upon the United Nations Population Division estimates and are assumed to increase as a result of a projected decrease in total fertility rates, in most countries. In these cases, we assumed that the values with and without DFID support were identical to ensure that we were not ascribing any benefit to DFID due to family planning when it was not being supported.

For the first question when family planning activities were being implemented – how the contraceptive prevalence rate changed with a program that was implementing family planning – we assumed that all supplies reached the target beneficiaries, to be consistent with other supply driven intervention data that were reported by the countries. The effect of this assumption is to risk overestimating lives saved.

For the second question – how the method mix changed over time – we had to consider how family planning is modelled in the software. Spectrum uses data from the most recent

DHS/MICS survey on contraceptive methods to define the method mix within a country. This includes both modern (e.g. condom, sterilization) and traditional (e.g. withdrawal) methods. However, DFID only supplies modern methods. Thus, if a country reported that they were delivering family planning supplies, we assumed that the increase was all in modern contraceptives, even if the reported indicator was all contraceptives. For a more general program, such as partnering with practitioners or community workers, then we assumed that there could be some increase in non-modern methods as well.

It is important to note that, within *LiST*, since family planning has impacts on pregnancy rates, it will automatically impact both the number of maternal deaths (but not the mortality ratio) and the number of neonatal deaths (but not the neonatal mortality rate). For this analysis, only the impact of family planning on maternal mortality is considered. See below for the explanation of how family planning changes were used to develop the underlying fertility rate for neonatal mortality calculations.

Additional *LiST* data needed

Additional data were needed by *LiST* for each of these analyses. These included 2010 neonatal mortality rates (NMR), infant mortality rates (IMR), under-five mortality rates (U5MR) and maternal mortality ratios (MMR) as well as stunting rates, wasting rates, diarrheal incidence rates etc., as described in Walker et al(7). For some information, such as mortality rates and ratios, there are multiple possible sources. A single source was chosen for consistency across all countries, regardless of additional data available. Additional sources for key values were varied in the sensitivity analyses described below. For mortality rates (NMR, IMR, U5MR), the most recent IGME estimate for the year 2010 was used as the baseline rate. For the maternal mortality ratio (MMR) we used the WHO 2010 estimates published in 2012.

Analysis methods

Create 'with DFID support' *LiST* projection

Once all these data were compiled for each analysis, a *LiST* projection for the years 2010–2015 was created for each analysis. Each *LiST* projection includes all of the underlying data described above as well as the projected number of births and deaths in each calendar year. This projection was typically required to begin in the year 1970 to ensure that the 2010–2015 HIV data were correct, also explaining why the term 'projection' is used for the analysis unit. Each analysis had a single 'with DFID support projection. This projection was called the 'with DFID support' scenario and was saved as the first part of the analysis.

Generate attribution values

In order to estimate the lives saved over this time period, an alternative scenario, called 'without DFID support', was needed. The difference between these two scenarios generates the final results. In order to calculate the 'without DFID support' scenario, each project and analysis needed to have the appropriate attribution calculated. The calculation method depended on the type of data, coverage percentage or numbers, and whether the attribution related to the total coverage or the additional coverage.

In one example, the additional number of services (and thus an additional percent coverage) attributable to DFID funding and programming was available as described above in the coverage calculation estimates. It was assumed that DFID did not support the existing levels of coverage for that intervention, only the additional amount. If we assume that 100% of the additional amount was due to DFID funding, then the 'without DFID support' scenario assumed that there was no change in coverage from program baseline. Alternatively, if we assumed that only 50% of the services delivered were funded by DFID, then the 'without DFID support' scenario would have only half the coverage observed in the 'with DFID support' scenario. If multiple programs contributed, then the attribution was combined in the 'without DFID support' scenario (See Table 3).

An alternative method was used when the program was assumed to contribute to total health output rather than to the additional health output. Typically this would be a proportion of the health budget paid for by DFID. In that case, the 'with DFID support' scenario reflects the total coverage of a health intervention in the population. The 'without DFID support' scenario is calculated by taking away the DFID funded proportion of the overall health output. See table 3 below for examples.

If multiple programs were contributing in varying amounts to the support of the total health system, then variable annual attributions could be calculated. For example, in some countries, there could be two budget support programs, i.e. a health budget program running from 2008–2013 and a general budget program running from 2012–2017. In the years 2012–2013, the total attribution should be the sum of the two programs, while the year before (2011) would only include the attribution for the first program and the years after (2014–2015) would only include the attribution from the second program. Within each program, the annual attribution could be variable as well, depending on the data shared by the national office.

Create 'without DFID support' *LiST* projection

Once each of the attributions was defined, a second *LiST* projection was created for each of the initially created 'with DFID support' scenarios. It was critical that this was built on the exact existing projection and no changes were made to the year 2010 or earlier. This was

the only way to ensure that the 2010 values were identical in the 'with DFID support' and 'without DFID support' projections and allow appropriate comparisons.

Obtain MNCH+FP results (for maternal)

At this point, the number of maternal deaths in each year in each of the two scenarios was generated. The difference between the two scenarios in each year was reported as the total maternal deaths averted.

Table 3. General Description of how different data types were modelled

Data Type	*Additional* Services Delivered by DFID (number)*		Percentage of *Total* Health Budget Due to DFID‡	
Data Needs	Baseline Coverage, Services Delivered by DFID, Population		Baseline Coverage, Final Coverage, % of total budget from DFID	
'With DFID support' Coverage	Use Baseline and calculate % reached by services		Use Baseline and Target Coverages	
'With DFID support' Example	Baseline Coverage	50%	Baseline Coverage	50%
	Additional Services Delivered	3,000		
	Percent attributed to DFID	40%		
	Total Population in Need†	10,000		
	Calculation of additional coverage	40%*3,000/ 10,000=12%		
	Final Coverage	50%+12%=62%	Final Coverage	80%
'Without DFID support' Description	What would coverage be if the DFID supported services were not delivered			
'Without DFID support' Example	Baseline Coverage	50%	Baseline Coverage	50%
	Additional Services Delivered	0	Percentage attributed to DFID	50%
	Total Population in Need	10,000	Final Coverage	80%-(50%*80%) =40%
	Final Coverage	50%+0%=50%		
Deaths Averted Calculation	Deaths at 62% coverage minus Deaths at 50% coverage in the same year		Deaths at 80% coverage minus Deaths at 40% coverage in the same year	

*Assumes DFID only supported additional services; †Refers to both met and unmet need
‡Assumes DFID supported part of all services

Create alternate family planning projections

As described above, for each *LiST* analysis, multiple scenarios were created. The first (A, below) was the 'with DFID support' scenario. The second (D, below) was the 'without DFID support' scenario. Two additional scenarios not previously described were created (B and C, below) which were intended to strip out the impact of family planning and only calculate the impact of maternal and child health interventions. Scenario B modelled all of the health impacts that were in the 'with DFID support' scenario, but used the 'without DFID support' scenario for the family planning values. Scenario C, on the other hand, only included the 'with DFID support' family planning values and the 'without DFID support' maternal child health values. (Table 4)

Table 4. The *LiST* scenarios required for calculations for each analysis

		DFID support for Family Planning	
		Yes	No
DFID Support for MNCH	Yes	A*	B
	No	C	D‡

* A: 'with DFID support' scenario
‡ D: 'without DFID support' scenario

These four scenarios, described in table 4, were used to calculate the deaths averted due to DFID programming in three individual analyses. The first analysis modelled the total impact of both the MNCH programming and family planning (MNCH+FP). This was done by subtracting the total deaths observed in 2011–2015 in scenario A from those in scenario D for the same years (as described above). This was the primary result used for all of the maternal outcomes reported.

Obtain MNCH results (for neonatal)

The neonatal analysis was more complicated with respect to family planning. The UK Government target for newborn lives saved had been set on the basis of excluding the influence of family planning. Although family planning has a major influence on the number of newborn deaths at the epidemiological level, at the individual level we do not consider preventing a baby's conception as *saving* its life. Regardless, knowing the total fertility rate was critical to understanding the number of births and thus the number of neonatal deaths. As a result, additional analyses were designed in which both the 'with DIFD support' and 'without DFID support' had identical family planning rates in order to eliminate the family planning impact. It would be impossible to accurately determine which set of family planning rates were more likely to be correct over the time period of the analysis. Thus the second analysis of the three planned analyses looked at the impact of MNCH programming assuming that the contraceptive prevalence rate was the same as used in the scenario 'without DFID support'. This was done by subtracting the total deaths in scenario B from those in Scenario D. The third and final analysis looked at the impact of MNCH programming assuming that the contraceptive prevalence rate was the same as in the scenario 'with DFID support'. This was done by subtracting the total deaths in A from C. Since family planning programming data are integral in estimating the number of births, it was necessary to do two analyses to get the range of the likely change, given two different family planning scenarios. The range of results from the 2nd and 3rd analyses was reported as the primary result for the neonatal outcomes. For example, in one country, the total 'lives saved' from combining family planning increased from 16 to 28% with the health coverage increases, was 6,100. If one considered two different, but equally plausible estimates of family planning coverage (1: unchanged with approximately 16% of

married women of reproductive age using family planning or 2: increased from 16% to approximately 28% of married women using family planning), then the 'lives saved' was either 3,000 or 3,400. The lower range was the result reported for this analysis.

The primary result of the analysis was lives saved. That is, the number of deaths which are likely to not have occurred, at least partially because of DFID funding and inputs. In addition, we reported the deaths averted due to each health intervention within each country. The result of deaths averted does not take into consideration the fact that some of these children who are being saved due to one intervention can also die due to other causes at population risk. Thus, if one were to sum up the deaths averted by health interventions, this number will be slightly larger than the lives saved reported in the first analysis.

As many of the analyses were for multiple programs, or, inversely, multiple analyses were for one program, we chose not to report the deaths averted by individual program, but overall within the country.

Although not integral to the results of this particular analysis, since they were available by default, we also reported the range of stillbirths averted and the range of deaths averted in children 1-59 months of age.

Multi-country and multilateral programming

The main *LiST* analysis methods described here are only relevant for the bilateral funding streams. Additional projects have been funded which are being implemented in multiple countries, such as Prevention of Maternal Death from Unwanted Pregnancy (PUMDUP), Evidence for Action (E4A) and Making it Happen (MiH). There is also support being given to multilateral organizations such as GAVI. Since it will be almost impossible to separate out the effects of bilateral programming from multilateral programming, it has been decided to not pursue these impacts at this time within this *LiST* analysis. In addition, there is currently relatively little programming which has the potential to have a notable impact on maternal or newborn mortality.

Briefly though, the multi-country programming does have the ability to be analysed in certain situations. For each program with logframe data available, one question was asked: is there programming on the multi-country program topic from other programs in the same country which have already been included in the main *LiST* analysis. If the answer was no, then a new *LiST* analysis was done. In addition, if it could be determined that there was no programming on the multi-country program topic in a given calendar year, then the analysis was done only for those specific years. This second option required more forethought since many of the DFID programs actually reported non-calendar year results. This type of analysis

was tested starting in 2015 for the 2011–2013 results, with the intent to expand upon it in future years. Countries able to be included in the analysis are shown in Appendix A.

Quality control methods

The impact of this analysis is fully dependent on the results being credible, justifiable and reliable. This is a difficult task, given the reliance on secondary data collection from the DFID offices, and their reliance on selected data collection by either the country itself, or by the project implementers. Several steps were taken to mitigate the potential consequences of this reliance. First, any known additional data sources were included if needed, and general data quality checks were performed. Second, each country office was requested to validate their results. Third, alternate potential mortality rate data was used and presented for national analyses. Finally, two ranging analyses were designed to fit the analyses and reflect upon confidence in the resultant data.

The Johns Hopkins staff assessed each reported potential indicator and intervention which was reported. If additional data sources were available and appropriate, they were compared to the reported data. If no baseline data were available from the DFID country office data collection template, these additional sources were used. In all cases, the DFID office was made aware of the additional data used, and then asked to confirm this new additional data. Other data quality checks were done, dependent on the exact data available. These included converting all services delivered values into coverage estimates based on the default Spectrum populations. If the values were not consistent with other available data, or suggested coverage of greater than 100%, then the DFID country office was asked to investigate.

These data checks resulted in a series of direct questions which were posed to each of the DFID country offices. At that point, the country office was requested to validate or change the information which was prepared by the Johns Hopkins team within a short period. The responses from the DFID offices ranged from full acceptance of the results to disagreement with either some or all of the results. If possible, the comments from the country office were incorporated into the final results in the annual report. In situations where agreement or understanding could not be reached in a timely fashion, this was explicitly noted in the annual report. For example, Tanzania, in the second year of the analysis, felt that the family planning results were inappropriately high. This was noted and remedied in the third year of the analysis.

Uncertainty

In addition, it is clear that the results being reported in the primary analysis are subject to uncertainty. There is significant variability around several potential inputs to these analyses. For many of these it is impossible to determine the true confidence around the value, such as for number of ITNs delivered. Excluding those, several critical inputs of concern include the

mortality rates/ratios, the causes of death, the achieved/target values and the underlying population. In order to address these issues, several sensitivity analyses have been developed. The first is to rerun all of the analyses with alternate mortality rates. The standard mortality estimates used are from IGME(21). Alternate values could have been used from DHS/MICS surveys as well as IHME(23) or other local sources. As a result, the total results (number of deaths averted) could be very different. Each analysis was rerun with alternate values for the baseline mortality estimates, including IHME and DHS/MICS if there was a relatively recent (2008–2013) survey.

An additional important estimate of uncertainty relates to target coverage values. In many cases, the program targets are the same as the national targets, without reference to their achievability. We modelled additional options, such as if only one half of the goal or ten percentage points more than the goal was achieved, to obtain a potential range around the estimates.

It was anticipated that the underlying uncertainty within the Spectrum software will be able to be utilized for the 2014–15 round of analyses. This feature will include the ability to model uncertainty around the mortality rates, the coverage as reported by surveys, the cause of death structure and the effectiveness estimates.

Ranging analyses

Since it was not possible to calculate a true uncertainty estimate, two ranging analyses were created to generate a range of likely values that could be considered trustworthy around the maternal and neonatal deaths averted. The first simply asked whether or not the DFID country office had validated the results. If yes, then they were included as likely true results. The second asked about the quality and qualities of the data which was included. Only countries with multiple years of some achieved results and who had also checked their data were included in this analysis. In the early years of the project, relatively few countries were able to report actual data rather than forecast data, and thus, relatively few of the deaths averted were considered to have been achieved. By the third year of the analysis, the majority of countries were able to report at least some achieved results rather than forecasts.

Limitations

There are several limitations to this analysis that should be considered, many of which have already been mentioned. The types of limitations can be categorized three ways. The first relates to data availability, the second is technical, related to the software itself, while the third is related to overall uses of modelled results.

Data Limitations

Critical data elements for this analysis include both DFID-specific values as well as more general values. One important concern is that there is no good source for all of the achievements and targets which are being supported by DFID. The project monitoring systems do not necessarily include all critical activities; many only include a subset of activities and indicators. Thus it is likely that *LiST* is not modelling a comprehensive set of interventions for each program, and is underestimating the impacts of DFID programming.

Another concern surrounds the target values. In programs where the targets are delivery of specific interventions, it is relatively easy to estimate the outcomes and likely impacts. However, when the programming is distributed through third parties, national targets are often used. These are not necessarily based upon data, but upon the wishes of the national government. These become aspirational targets, rather than feasible and objective targets. As a result, *LiST* could be overestimating the potential future impacts. The second of the ranging analyses will hopefully take this into consideration, but it must still be considered in interpretation of the results.

Just as critical as the DFID specific program inputs are several of the default values. Many are modelled results themselves, including the causes of death and the IGME mortality rates and ratios. Plus, even though the newborn causes of death are available at the national level, the maternal causes of death are only available at regional levels. In addition, when extrapolating any of these values to the sub-regional areas for selected analysis, care must be taken. Although a methodology exists for these extrapolations (22), the variability around them is uncertain. In addition, even when a program is modelled nationally, it may be implemented sub-nationally, and in select areas, resulting in additional and inestimable uncertainty.

Technical Limitations

LiST in itself has additional limitations, due to functionalities which do not (yet) exist. This may include interventions which have an impact on mortality, but are not currently available within *LiST* or Spectrum. One example is birth spacing (24, 25). There is an impact on mortality, modulated through prematurity and small for gestational age, but at the time of the analysis in year 3, this feature was not 100% completed. Thus, there are additional deaths which can be prevented, which have not been counted. In this case, this is likely to be a small effect, but when it is finally included, these deaths averted will be able to be combined with the effects of maternal and child health interventions. Similar examples include the incomplete inclusion of hospital based care for newborns and for older children as well as for post-partum women.

Also of consideration is the translation between behaviour change communication programming and the outcome, which is inadequately accounted for within *LiST*. For example a

hygiene program can educate on handwashing, but it is currently unclear how well that translates into handwashing behaviours. Currently, we are likely to be overestimating the impact by incorrectly assuming that all people who are educated practice the behaviour. We are also concerned about supply wastage. For family planning supplies as well as other supply driven interventions (e.g., ITNs) we assume that all of the commodities are reaching their intended targets, although it is highly unlikely to be true. Thus, we believe that we are overestimating some of these impacts, but do not have a good estimate with which to modify this assumption.

The effectiveness assumptions around the interventions are key components of the technical software description above. Although these values are typically available in the peer reviewed literature, there are disagreements as to how appropriate each may be within a given context. Thus future analyses incorporating the wide ranges of uncertainty around some of these values will be critical to improving the quality of the results for this, and all, *LiST* analyses.

And finally, there are limitations around the assumptions relating to quality of care. Specifically during delivery care, there are standard algorithms which assume a specific quality relative to overall coverage levels. Changes in quality were modelled by changing the proportion of interventions available; however, these are, as described above, assumptions. There is little objective evidence on how quality changes, which can have large effects on newborn and maternal mortality. The impact of these changes on our estimations is unclear.

Interpretation Limitations

The use of modelling as a tool for monitoring and evaluation should also have its own serious caveats discussed. Models are simply that, they are models. They do not result in precise measurements of truth. They should not be used as stand-alone tools for assessing, program impact. They should be considered as only one tool in the armament of those interested in truly understanding the 'on-the-ground' situation and how to best affect it. Additional and disparate information sources should always be considered; modelling results which are not plausible should be critically assessed to understand their implications. In the first round of analysis, undertaken in 2013, antenatal corticosteroids for premature delivery can be considered a useful example. Antenatal corticosteroids were one of the most critical interventions for reducing newborn mortality. However, this was not considered to be credible in the environments where these results were being modelled. Thus, the report included the result, with the assessment that the result was not credible. This combination of showing the modelled results in combination with a critical and informed interpretation is key for generating useful results for future discussion. This is necessary to justify the use of modelled results, rather than taking them at face value without that critical thinking.

There are many factors which may lead to these results either over- or under-estimating mortality, some of which have been discussed here. *LiST* will not, as no modelling tool can, solve all of the issues. The estimates of lives saved produced using this methodology are considered indicative of the underlying true values in the absence of being able to directly measure mortality changes as a result of DFID programming. Several uncertainty or ranging analyses help to pinpoint the true level, based upon the available data and knowledge of the communities of interest.

As mentioned before, but which is also critical to remember is the distinction between attribution and contribution. Although the analysis is often termed as an assessment of attribution, the authors consider it much more of a contribution analysis. There are so many co-existing partners, projects, and trends related to health and as much, if not more, that are unrelated to health. Thus DFID is both implicitly and explicitly working with all of these and all of its effects are intertwined with these external forces. Nothing exists in a vacuum and the results of this analysis cannot adequately (or at all!) account for these forces. This fact must be considered in any presentation or use of the results: DFID contributes to saving lives in conjunction with all the known and unknown partners and changes.

Discussion

The results of this exercise should never be considered to be final exact results. There are many other factors in play which should be considered, but are not able to be fully mitigated. As mentioned, there are several limitations to this analysis, including the modelling tool used, the data available to populate it, and the need for modelling and interpretation of modelling results. Funding assumptions are critical – that is, that funding can directly be interpreted to relate to a change in coverage and mortality while not being aware of how others are acting in this environment. There is no fool-proof way of mitigating this problem, so it must be discussed in every use of the results. An additional consideration is in conflict situations or weak states. In these situations, the funding is often going to be used simply to maintain the current situation or prevent a more severe reduction in health status. In these situations, the use of funding to model impact may result in an overestimate of the additional lives saved, but may be a true reflection of the actual lives saved.

Another consideration is that the results of this exercise are not directly comparable with results from other sources, such as global analyses or other donors. The specific methodology used is likely to be different. For example, in these analyses, family planning is assumed to have no impact on neonatal mortality, while others may choose to include that. In addition, different specific time frames analysed will also affect the level of lives saved.

The results of this project, as with any modelling activity or situation where the data are incomplete, have the ability to be misused. The results, in terms of lives saved, could be published and used to indicate that some programs are inefficient, but that is not appropriate since in this exercise, we are not considering contextual factors such as civil strife, natural disasters, political changes and cultural history, which are all important to consider. These results are being calculated to support the framework for results, and will only be published in aggregate form to indicate how many lives saved it is likely that DFID has contributed to. We deliberately use the terminology 'contributed to' in all project documents, and in all reports referring to this project. In addition, the framework for results only includes maternal and newborn mortality: many of the projects being analysed will also have an impact on stillbirths or children 1-59 months of age. Thus the reported lives saved cannot be the total impact of the programming. This project was designed with the express purpose of measuring and evaluating the impact of DFID health programming on maternal and newborn mortality. A methodology was developed to attempt to reach that goal. As of now, *LiST* is an adequate piece of software which has the flexibility to respond to DFID's queries. It is hoped that as the end of the project approaches, the observed increase in data availability as well as improvements in the software will continue to meet the needs of DFID.

References

1. DFID. Choices for women: planned pregnancies, safe births and healthy newborns. The UK's Framework for Results for improving reproductive, maternal and newborn health in the developing world. 2010; Available from: www.gov.uk/government/publications/choices-for-women-planned-pregnancies-safe-births-and-healthy-newborns
2. CHERG. Available from: www.livessavedtool.org.
3. UNICEF. Available from: http://www.devinfolive.info/mbb/mbbsupport/.
4. Avenir Health. Available from: http://www.avenirhealth.org/software-onehealth.php.
5. Marie Stopes International. Available from: http://mariestopes.org/impact-2.
6. RTI. Available from: http://mnhtech.org/.
7. Walker N, Tam Y, Friberg IK. Overview of the Lives Saved Tool (LiST). BMC public health. 2013;13 Suppl 3:S1. Epub 2014/02/26.
8. Avenir Health. Spectrum. http://www.avenirhealth.org/software-spectrum
9. Stover J, McKinnon R, Winfrey B. Spectrum: a model platform for linking maternal and child survival interventions with AIDS, family planning and demographic projections. International journal of epidemiology. 2010;39 Suppl 1:i7-10. Epub 2010/04/02.
10. Boschi-Pinto C, Young M, Black RE. The Child Health Epidemiology Reference Group reviews of the effectiveness of interventions to reduce maternal, neonatal and child mortality. International journal of epidemiology. 2010;39 Suppl 1:i3-6. Epub 2010/04/02.
11. Black RE, Morris SS, Bryce J. Where and why are 10 million children dying every year? Lancet. 2003;361(9376):2226-34. Epub 2003/07/05.

12. Jones G, Steketee RW, Black RE, Bhutta ZA, Morris SS, Bellagio Child Survival Study G. How many child deaths can we prevent this year? Lancet. 2003;362(9377):65-71. Epub 2003/07/11.
13. Lawn JE, Cousens S, Zupan J, Lancet Neonatal Survival Steering T. 4 million neonatal deaths: when? Where? Why? Lancet. 2005;365(9462):891-900. Epub 2005/03/09.
14. Lawn JE, Blencowe H, Oza S, You D, Lee AC, Waiswa P, et al. Every Newborn: progress, priorities, and potential beyond survival. Lancet. 2014;384(9938):189-205. Epub 2014/05/24.
15. Bhutta ZA, Ahmed T, Black RE, Cousens S, Dewey K, Giugliani E, et al. What works? Interventions for maternal and child undernutrition and survival. Lancet. 2008;371(9610):417-40. Epub 2008/01/22.
16. Bhutta ZA, Das JK, Rizvi A, Gaffey MF, Walker N, Horton S, et al. Evidence-based interventions for improvement of maternal and child nutrition: what can be done and at what cost? Lancet. 2013;382(9890):452-77. Epub 2013/06/12.
17. Pattinson R, Kerber K, Buchmann E, Friberg IK, Belizan M, Lansky S, et al. Stillbirths: how can health systems deliver for mothers and babies? Lancet. 2011;377(9777):1610-23. Epub 2011/04/19.
18. Bhutta ZA, Yakoob MY, Lawn JE, Rizvi A, Friberg IK, Weissman E, et al. Stillbirths: what difference can we make and at what cost? Lancet. 2011;377(9776):1523-38. Epub 2011/04/19.
19. Winfrey W, McKinnon R, Stover J. Methods used in the Lives Saved Tool (LiST). BMC public health. 2011;11 Suppl 3:S32. Epub 2011/04/29.
20. Yukich J, Bennett A, Keating J, Yukich RK, Lynch M, Eisele TP, et al. Planning long lasting insecticide treated net campaigns: should households' existing nets be taken into account? Parasites & vectors. 2013;6:174. Epub 2013/06/15.
21. WHO and UNICEF. Available from: childmortality.org.
22. Amouzou A, Richard SA, Friberg IK, Bryce J, Baqui AH, El Arifeen S, et al. How well does LiST capture mortality by wealth quintile? A comparison of measured versus modelled mortality rates among children under-five in Bangladesh. International journal of epidemiology. 2010;39 Suppl 1:i186-92. Epub 2010/04/02.
23. Wang H, Liddell CA, Coates MM, Mooney MD, Levitz CE, Schumacher AE, et al. Global, regional, and national levels of neonatal, infant, and under-5 mortality during 1990 – 2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet. 2014;384(9947):957-79. Epub 2014/05/07.
24. Kozuki N, Sonneveldt E, Walker N. Residual confounding explains the association between high parity and child mortality. BMC public health. 2013;13 Suppl 3:S5. Epub 2014/02/26.
25. Kozuki N, Walker N. Exploring the association between short/long preceding birth intervals and child mortality: using reference birth interval children of the same mother as comparison. BMC public health. 2013;13 Suppl 3:S6. Epub 2014/02/26.

Acknowledgements

Production of this technical note was funded by the UK government. Thanks to Gillian Mann (senior health adviser, DFID) for helpful comments on several drafts of this document. Thank you also to the external peer reviewers for their helpful and insightful comments.

Appendices

These appendices (A and B) are designed to be updated annually as a new round of analyses is completed. There is a space for results in 2017 for completeness even though this has not yet been funded.

Appendix A: DFID focus Countries analysed

Table A. List of DFID focus countries included in the final analysis, by year in which the results were reported.

	Year reported	2013	2014	2015	2016	2017*
	Years analysed	2011	2011–2012	2011–2013	2011–2014	2011–2015
Region	Country					
Africa	Burundi					
	DRC**		X	X	X	
	Ethiopia	X	X	X	X	
	Ghana	X	X	X	X	
	Kenya	X	X	X	X	
	Liberia			X	X	
	Malawi	X	X	X	X	
	Mozambique	X	X	X	X	
	Nigeria	X	X	X	X	
	Rwanda	X	X‡	X‡	X‡	
	Sierra Leone	X	X	X	X	
	Somalia	X	X	X	X	
	South Africa				X	
	South Sudan			X	X	
	Sudan					
	Tanzania	X	X	X	X	
	Uganda	X	X	X	X	
	Zambia	X	X	X	X	
	Zimbabwe	X		X	X	
ASCOT	Bangladesh	X	X	X	X	
	Burma	X	X	X, M	X	
	India	X	X	X	X	
	Indonesia					
	Nepal	X	X	X	X	
	Vietnam					
Western Asia	Afghanistan			M		
	Central Asia			X	X	
	Pakistan	X	X	X	X	
Regional	Yemen	X	X	X	X	
Total	-	19	19†	23		

X: included in main analysis; M: included in multi-country analysis; *Will be updated when analysis is performed; **Democratic Republic of Congo; †South Africa also reported, but was unable to be analysed; ‡All data analysed was reported in the first year of the analysis.

Appendix B: Spectrum software

Each year a new and updated version of the Spectrum software is used to complete the analyses. The shift from version 4.x to 5.x is a large difference with notable changes in population structure and technical inputs. The differences noted within primary version numbers is typically not as pronounced as between version numbers. The intention is that the version used will be a publically available version, but it will also balance the need for having the most up-to-date data and format as well as the need for corrections to known errors. In the first two years of the analysis, it was possible to use a publically available version of the software. However, in the 3rd year, it was not possible to wait for the version to be released to complete the analysis. For more information on the current *LiST* version and software available, please visit the official website, livessavedtool.org.

2010–2011 results (reported in 2013): Spectrum 4.58

2010–2012 results (reported in 2014): Spectrum 4.71

2010–2013 results (reported in 2015): Spectrum 5.2 beta 12

2010–2014 results (reported in 2016): Spectrum 5.33

2010–2015 results (reported in 2017):

Appendix C: Interventions – for Spectrum version 5.31

This table lists all of the interventions which can be modelled by the Spectrum software, at the time this method note was finalized. For each link between an intervention and a type of death (maternal, neonatal, stillbirth and child) there is an effect size and a cause of death. It is possible for some of these values and linkages to change as frequently as every year. Some interventions have direct effects on mortality, while others work via changing risk factors, such as stunting, wasting or disease incidence. For the full details on how each intervention works, please see the help menus within the publically available version of the software, available at livessavedtool.org. This table does not list interventions modelled by either the HIV module (AIM) or the family planning module (FamPlan). Also note that the terminology in this table matches the *LiST* terminology and may not match the exact terminology used by specific interventions within the DFID system. Interventions were matched to project activities based on an understanding of the content, not the specific label.

Interventions	
Periconceptual	**Immunizations**
Folic acid supplementation/fortification	DPT
Safe abortion services	H. influenzae b
Post abortion case management	HepB
Ectopic pregnancy case management	Pneumococcal
Antenatal	Rotavirus
TT - Tetanus toxoid vaccination	Measles
IPTp - Pregnant women protected via intermittent preventive treatment of malaria during pregnancy or by sleeping under an ITN	**Postnatal through 5 years of age - Curative**
Syphilis detection and treatment	Maternal Sepsis case management
Calcium supplementation	Case management of premature babies
Micronutrient supplementation (multiple micronutrients + iron folate)	Thermal care
Iron supplementation	KMC - Kangaroo mother care
Multiple micronutrient supplementation	Full supportive care for prematurity
Balanced energy supplementation	Case management of severe neonatal infection
Hypertensive disorder case management	Oral antibiotics
Diabetes case management	Injectable antibiotics
Malaria case management	Full supportive care for sepsis/pneumonia
MgSO4 - Management of pre-eclampsia	ORS - oral rehydration solution
FGR - Fetal growth restriction detection and management	Antibiotics - for treatment of dysentery
Delivery care	Zinc - for treatment of diarrhoea
Clean birth practices	Oral antibiotics for pneumonia
Immediate assessment and stimulation	Vitamin A - for treatment of measles
Labour and delivery management	Antimalarials - Artemesinin compounds for malaria
Neonatal resuscitation	Therapeutic feeding - for severe wasting

Antenatal corticosteroids for preterm labour	Treatment for moderate acute malnutrition
Antibiotics for pPRoM	
MgSO4 management of eclampsia	
AMTSL--active management of the third stage of labour	
Induction of labour for pregnancies lasting 41+ weeks	
Postnatal through 5 years of age - Preventive	
Breastfeeding	
Clean postnatal practices	
Chlorhexidine	
Complementary feeding--education only	
Complementary feeding--supplementation and education	
Vitamin A supplementation	
Zinc supplementation	
Improved water source	
Water connection in the home	
Improved sanitation - Utilization of latrines or toilets	
Hand washing with soap	
Hygienic disposal of children's stools	
ITN/IRS - Ownership of insecticide treated nets (ITN/LLIN) or household protected with indoor residual spraying	