A guide and tools for
maternal mortality programme assessment

MODULE 3

Designing and managing an evaluation

Version 2.0

List of Acronyms

BEOC	basic emergency obstetric care
PQOC	Perceptions of Quality of Care
SAFE	skilled attendance for everyone

Key Definitions/Concepts

Inputs:	**Inputs** are required to implement a programme. Examples of inputs are funds, time and human resources, including whether the programme is delivering or improving drugs, supplies, infrastructure, communications systems, training, health information systems, etc.
Process:	The **process** describes how a programme is implemented. This area deals with, for example, whether the skills employed for implementing the programme are appropriate, or the motivation levels of the implementers
Outputs:	**Outputs** are the intermediate purpose of a programme, examples being: improved service utilization, or better quality of services. Confusingly, what are commonly known as 'process measures' in maternal health (for example utilization levels of health services and quality of care) are actually output measures, or indeed a mixture of process and output measures.
Outcomes:	The **outcomes** of a programme are what are commonly known as the ultimate goal and are usually measured as health outcomes, for example morbidity or mortality.
Dissemination:	In the context of health services it is argued that .dissemination. has the meaning of communicating research findings to enhance their possibility of being translated into policy and practice, involving a two-way process of communication between the researcher and the policy maker and practitioner.
Causal net:	A model of the chains of events that lead to, or prevent, an adverse maternal outcome, e.g. a graphical representation of complex scenarios contributing to possible maternal death. (causal network and causal framework are used interchangeably).

Module 3: Table of Contents

3 INTRODUCTION

Module 2 has provided guidance on the Immpact process of identifying, refining and finalizing the evaluation question. This module (Module 3) guides you through the steps for moving from an evaluation question to the formulation of an evaluation protocol and the preparation of a workplan for a complex evaluation.

3.1 What is an evaluation protocol?

A protocol is a procedural document which defines the evaluation question and its specific objectives and sub-questions, and describes how the evaluation will be conducted. The complexity of composite evaluations requires careful preparation in order to develop a protocol which will bring the various components clearly together, and will become a useful resource document to co-ordinate the many different pieces of work required for the evaluation. This document will also be needed as part of your submission for technical and ethical review (see Step 8). The evaluation protocol should contain the following sections:

3.1.1 Background

a) The evaluation question and how it was developed (see Module 2).
b) Description of the programme and its situation and setting: including what useful information is already known, relevant literature, previous studies, situation analysis including any preliminary research undertaken, etc.
c) The rationale for the study or evaluation.

3.1.2 Methods

d) Study design: the approach to the overall design of the evaluation (see Technical Annexes in Module 5). Describe the methods (quantitative and qualitative) to be used, study design – how and where the evaluation will be conducted, comparisons, interventions, measures, data sources, recruitment strategy, sampling issues (justification for sample size, representativeness of the sample; and for quantitative aspects, power, statistical significance calculations).
e) Objectives: the objectives of the evaluation, related sub-questions and corresponding sub-studies.
f) Overview of sub-studies: details on each sub-study, its design, data sources, risks, potential limitations, biases, ethical matters, duration and resource needs.
g) Methods, tools and instruments for data collection.

Analysis

h) Analysis and synthesis plans.

Management

i) An overview of management arrangements: including logistics, data collection, data management, quality assessment, ethical issues, recruitment reimbursements for study participants, listing of key researchers.
j) Timeline: with consideration of such factors as dependence on other activities, availability of researchers, etc., as well as country governments' project planning cycles, key decision-making points, sequence and timing of obtaining approvals before start of the study, major country holidays, seasonal access and budget requirements.
k) Specific needs: such as co-operation with other projects/individuals, or need to influence stakeholders. Describe the degree to which any of the discussions required with other stakeholders have been conducted.

Dissemination

l) Dissemination and feedback plans: including whose needs will be addressed in the evaluation, who will find the results of interest, how it will be used to influence policy and practice.

3.2 Developing the evaluation protocol

The steps in figure 3.1 below guide you through the key activities needed to develop the evaluation protocol. As the activities within each of the steps are completed, relevant parts of the protocol should be written up, as shown in figure 3.1 rather than waiting for the completion of all steps before writing the protocol. Each section of your protocol can then be reviewed and used to conduct the actual evaluation. To support the design and management of the evaluation, checklists (for quantitative and qualitative studies) are also provided in Technical Annex E of Module 5 of this toolkit.

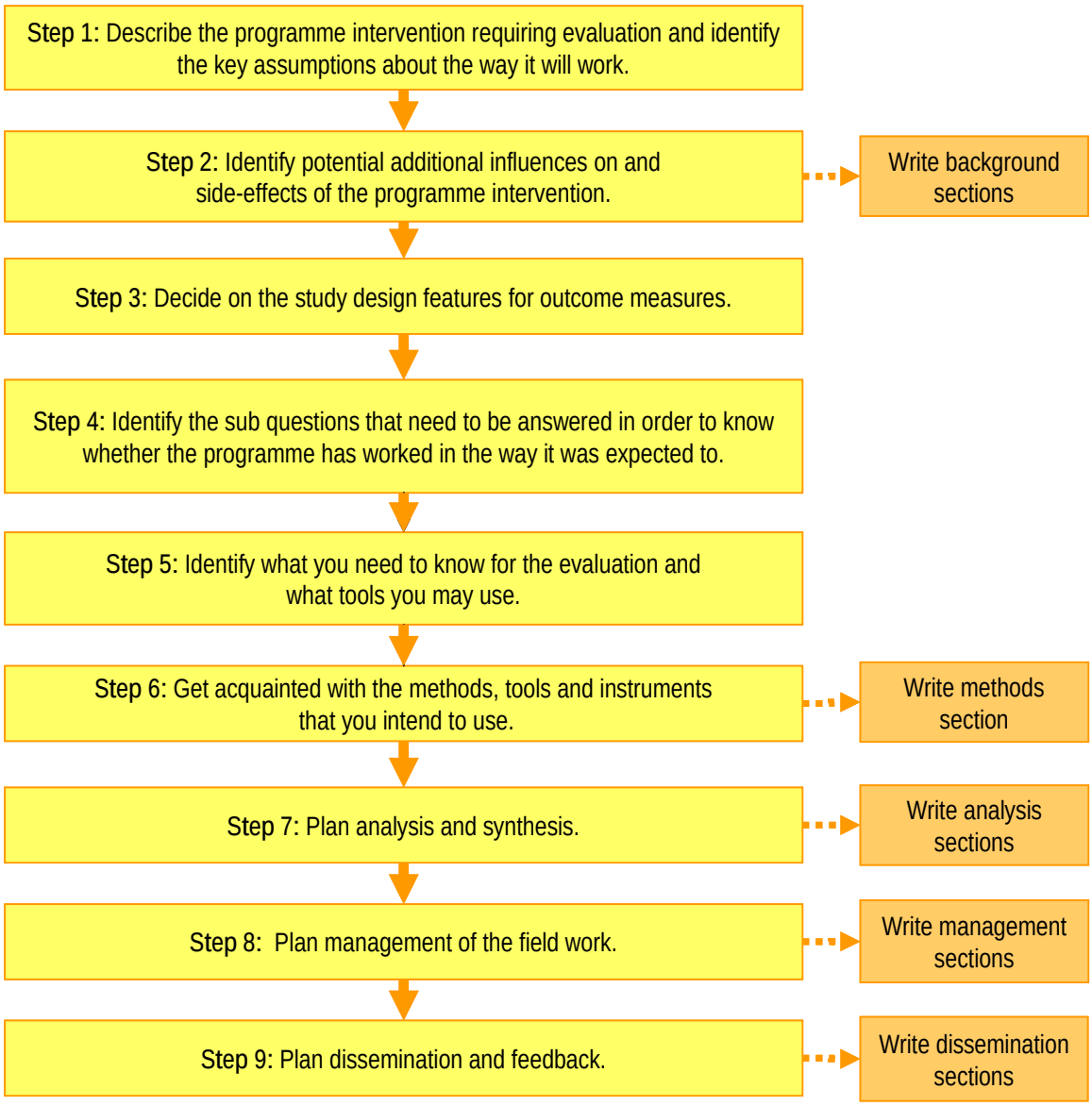


Figure 3.1: Steps in formulating an evaluation

Although the steps are presented in a linear fashion, you should be aware that the process of developing the protocol requires a reiterative and reflective process throughout, and you may find it useful to revisit and revise earlier steps several times before finalizing your protocol. This will mean that issues arising in the planning of the later steps will inform planning of earlier steps and the time-frame for initiating your evaluation activities. Therefore, by writing the protocol, you will have already anticipated problems and thought of ways to overcome potential barriers and difficulties.

Developing a protocol that addresses the key issues, and that is realistic and appropriate, usually requires the involvement of researchers who come from a range of backgrounds including quantitative and qualitative disciplines, policy and health systems analysts, social scientists, statisticians, epidemiologists, economists and practitioners. The involvement of individuals familiar with the realities of the local context being studied is also crucial. This will require regular interaction with the implementers of the programme as the protocol is being developed and as evaluation activities are conducted. This forms part of the iterative and reiterative process necessary to make the evaluation as fit for purpose as possible.

3.3 Step 1: Describe the programme and context

The main aim of this step is to obtain an accurate description of the main interventions within the programme and to determine how exactly the interventions within the programme were expected to achieve the desired results. You should aim to describe exactly what the programme was, where and when it happened, what it aimed to achieve and how it planned to reach its goals. A framework for conceptualizing the different types of component within a programme is provided in table 3.1.

Table 3.1: Framework for components within a programme

Types of programme component	***Examples of components***
INPUTS	• Funds • Time • Human resources
PROCESS	• Context of the health system • Appropriateness of skills • Motivation of implementers
OUTPUTS	• Increased utilization of services • More functioning facilities • Better health service quality
OUTCOMES	• Health • Death • Economic consequences

- **Inputs** are required to implement a programme. Examples of inputs are funds, time and human resources, including whether the programme is delivering or improving drugs, supplies, infrastructure, communications systems, training, health information systems, etc.
- The **process** describes how a programme is implemented. This area deals with, for example, whether the skills employed for implementing the programme are appropriate, or the motivation levels of the implementers.
- **Outputs** are the intermediate purpose of a programme, examples being: improved service utilization, or better quality of services. Confusingly, what are commonly known as 'process measures' in maternal health (for example utilization levels of health services and quality of care) are actually output measures, or indeed a mixture of process and output measures.
- The **outcomes** of a programme are what are commonly known as the ultimate goal and are usually measured as health outcomes, for example morbidity or mortality.

The programme activities can be mapped out using a programme logic chain (Leeuw, 2000). This process helps you to identify the assumptions about the programmatic intervention and the way it is expected to work that will need to be tested within your evaluation.

For example, a Ministry of Health might want to know if their village midwife programme has been effective in improving maternal and neonatal health at reasonable cost. Figure 3.2 depicts some key assumptions in that programme logic chain.

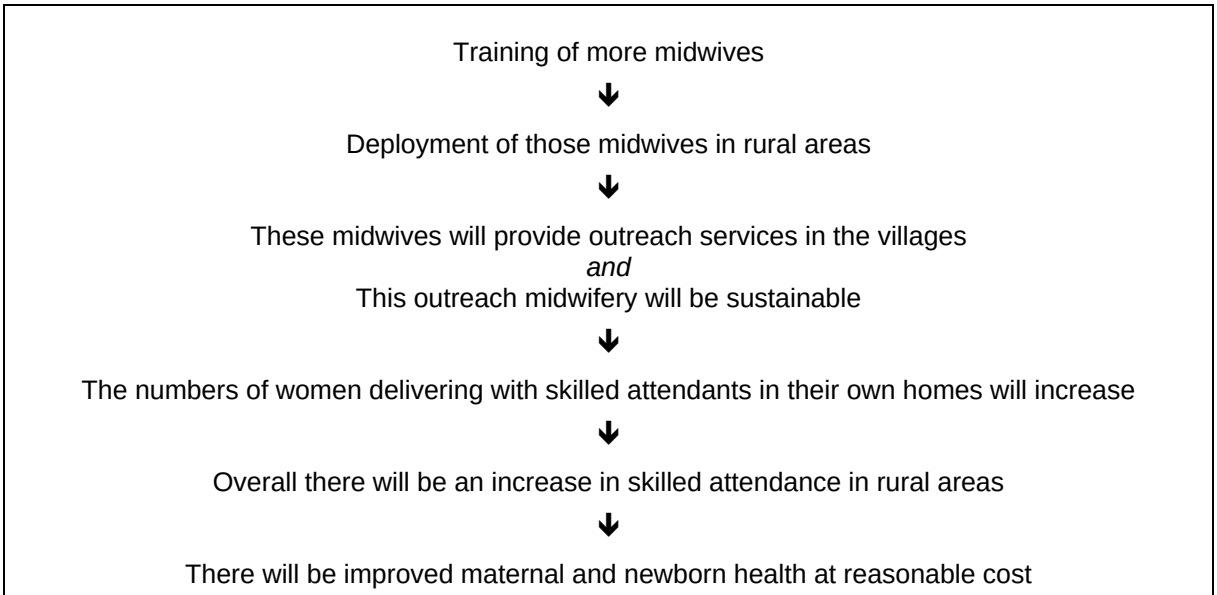


Figure 3.2: Logic chain for a village midwife programme

The logic chain is created by identifying elements that would need to be achieved for the expected result to occur. To inform the development of this logic chain, you can conduct some or all of the following activities:

- Hold discussions with people who have knowledge about the programme. These people may be external to the programme (for example, a ministry official responsible for co-ordinating all the safe motherhood programmes and projects in the country, or an external evaluator) or internal, such as those involved in managing, implementing or monitoring the programme. Both perspectives are useful to obtain.
- Visit or observe programme activities.
- Obtain and review documents. A useful guide for how to do this in a maternal health programme setting is available (SAFE 2003).

It may be useful to hold a workshop to bring together these different sources of information described above to formulate the logic chain. This can lead to the development of a logic chain that is agreed by all the key stakeholders.

As part of this exercise, you are likely to discover information about the context, or the situation and setting within which the programme was conceived and developed. Much of this information will be useful to summarize in the evaluation protocol for reference, and will inform Step 2. In addition, at this stage, a formal situation analysis can be conducted, which is particularly valuable as it may provide data and information which you can use for your evaluation, rather than having to collect it through primary data collection yourself. You may also find that you need to review the information on the existing situation again, and even supplement it with reanalysis of existing data to provide more specific information when planning the later steps described in Case Study 1. Guidelines on how to do a situation analysis and how to re-analyse existing data in a maternal health setting are available in SAFE 2003. This data may provide important baseline information, especially if, as is often the case, you are being asked to design an evaluation only after a programme has been implemented for a number of years.

The development of a programme logic chain is demonstrated in Case Study 1 from Burkina Faso.

Tip:
Don't underestimate the time required to fully understand and describe a programme. One single visit or interview is rarely sufficient and recurrent visits and discussions might be required as you go through the other steps which may raise specific questions that you had not thought of earlier. For this reason, it is vital to establish a good working relationship with the people working to implement the programme you are evaluating, especially if you are conducting an external evaluation. A good working relationship establishes trust, mutual respect and transparency.

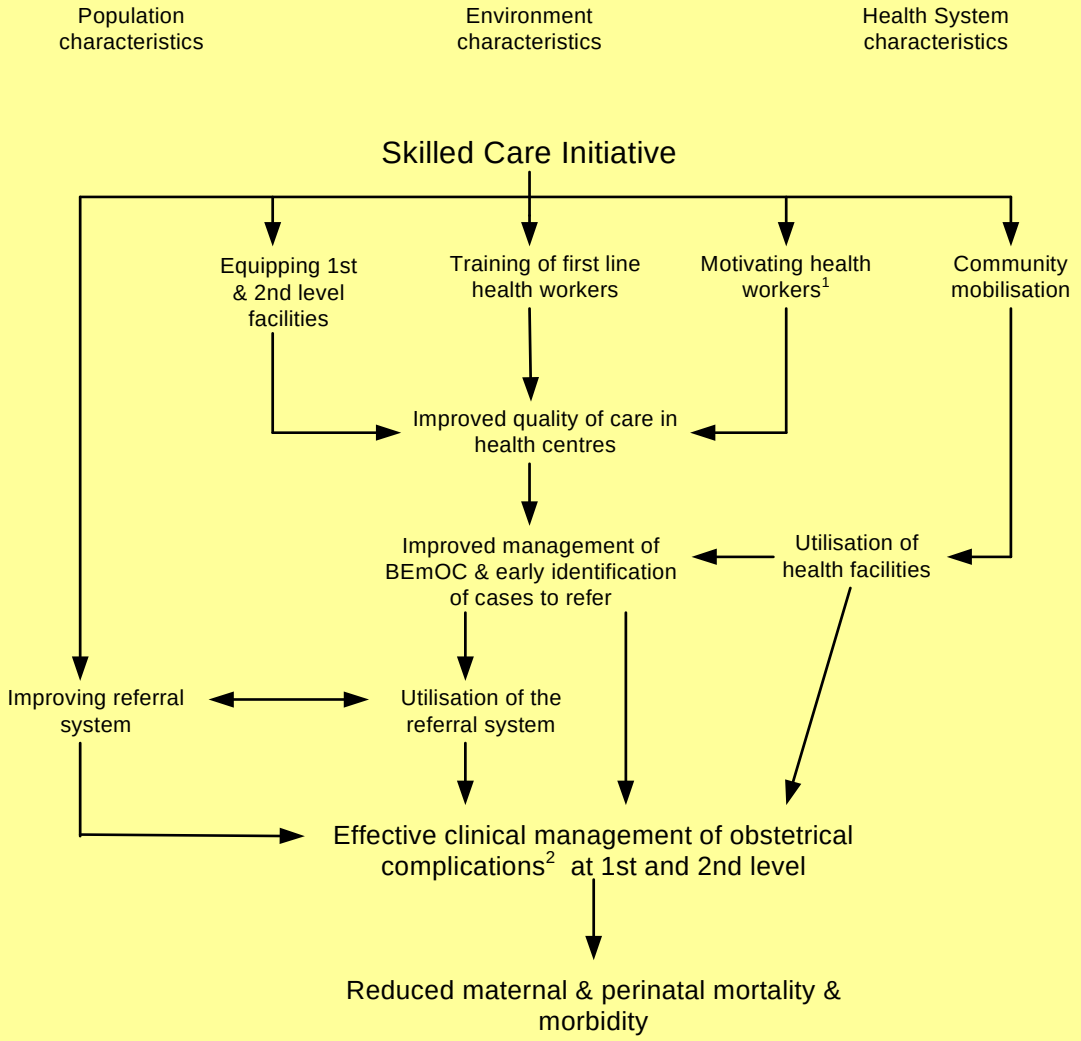
3.4 Step 2: Identify additional potential influences

In this step, the aim is to identify additional influences, aside from the programme interventions, which may affect the programme's implementation, or even generate the desired outcomes independently. Examples of this type of influence are political issues, general level of education, cultural norms and so on (sometimes these are referred to as distal determinants). The overall health system is also likely to be the source of other influencing factors, depending on how well it is functioning. There may, for example, be other health programmes being conducted, whether maternal or not, which might affect the processes and outcomes under evaluation.

It is important to identify these influences during the planning of the evaluation, in order to build in data collection that will help to account for their effect. However, it is equally important to keep looking for the unintended or unexpected throughout all the steps of the evaluation. As is shown in figure 3.3, it is possible to represent these influences diagrammatically, integrating them with your logic chain described in Step 1. Although this will necessarily make visual representation much more complex, creating such a *'causal net'* does help to ensure that key influencing factors are not forgotten about. We recommend that at the minimum, you develop a logic chain to identify the programme's underlying assumptions. Case Study 2 shows how influencing factors can be integrated with a logic chain into a causal net.

Case Study 1: Programme logic chain for the skilled care initiative in Burkina Faso

The Immpact evaluation carried out in Burkina Faso assessed the effectiveness of a strategy to ensure skilled attendance at delivery. This was done by conducting an evaluation of a programme of skilled care at delivery. The logic chain for the programme can be described as follows.



[1] Supervision, solar panels, working environment

[2] Antepartum & postpartum haemorrhage, obstructed labor, pre- and eclampsia, infection

Figure 3.3: Logic chain for evaluation in Burkina Faso

The programme aimed to improve conditions and supplies in health facilities and train health workers. The secondary aim was to improve staff retention and motivation. The underlying logic was that if maternity services at health facilities could be improved, that might lead to a greater uptake of those services, which in turn might lead to improvements in maternal survival.

Case Study 2: Causal net for the village midwife programme in Indonesia

Immpact Indonesian evaluation protocol contained a 'causal net' which was first developed during the consultative exercise with stakeholders to identify the links and relationships between various aspects of quality of care and provider performance (in an exercise similar to the one described for development of a logic chain in Step 1). The Ministry of Health in Indonesia wanted to know if their programme to post professional midwives in every village (the 'village' midwife) had been effective in improving maternal and neonatal health at reasonable cost. A causal net that examined midwives' contributions to the adequacy of care in the community was developed (see figure 3.4). For example, adequate obstetric care in the community (rectangular box) may be influenced by the physical availability of a midwife, her clinical ability and her workload. These were factors which were affected by the programme activities. However, apart from their role within the village midwife programme, village midwives are also multipurpose workers and their non-maternal workload may also be an important determinant of their obstetric capacity. The non-maternal workload was only identified as an important consideration in the Indonesian evaluation when the causal net was developed. It may not have been picked up if a logic chain alone had been developed, because the village midwife programme primarily focused on the role of midwives in providing pregnancy care.

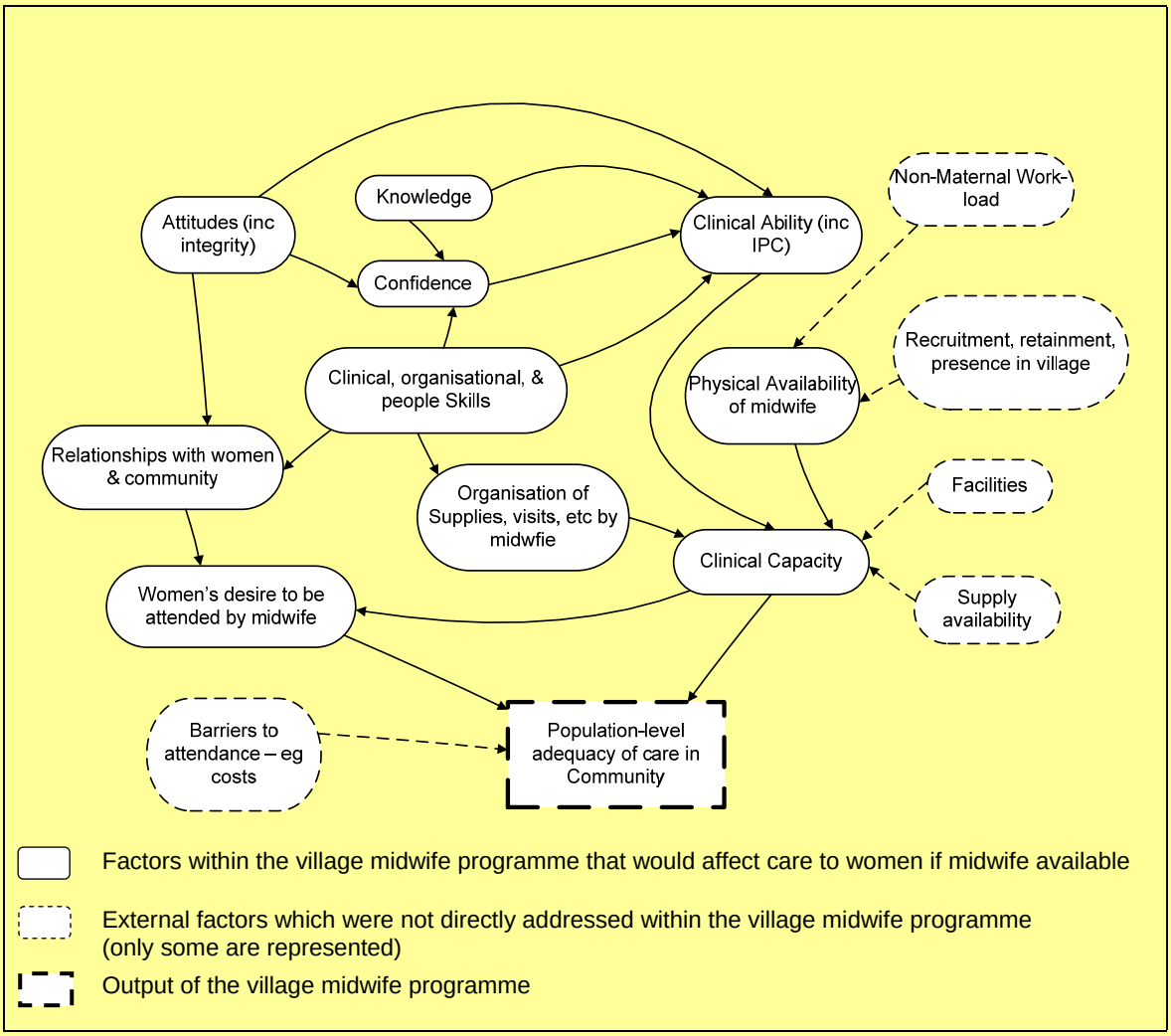


Figure 3.4: Causal net for the village midwife programme

The following activities can help you to identify the additional influencing factors on your programme:

- Conduct in-depth discussions with key stakeholders on programmes and contextual factors;
- Identify the related health programmes in the same geographical area;
- Describe in detail the most relevant ones which could affect your programme of interest;
- Map the key interventions of all these programmes by place and time.

3.5 Step 3: Decide on study design features for outcome measures

Once you have described the programme interventions and contextual information and have identified potential influences, it is time to start considering the possible study design(s) for your evaluation. The information collected to this point will be necessary to develop the study design. Some basic information on study design is available in Technical Annexes A and B in Module 4, but it is necessary to work with a competent epidemiologist or statistician to develop a good quality study design that will best fit your desired outcome measures. This will then form the foundation for the rest of the evaluation. It can be represented in a table as in the following Case Study 3 and Case Study 4.

Case Study 3: Study design in Burkina Faso

Let's take the example of the situation in the Immpact evaluation in Burkina Faso, which was described in Step 2. The evaluation assessed the effectiveness of a strategy to ensure Skilled Attendance at Delivery and can only be unequivocally answered by measuring maternal survival and relating it to areas covered by the skilled attendance programme which is being evaluated. Ideally this would mean measuring maternal survival in programme and non-programme areas before the programme interventions commenced, then measuring again in both areas after implementation. But the ideal design is not always possible. In this case, at the time the evaluation question was formulated, the programme had already been implemented. Retrospective assessment of maternal survival was the only option.

The study design for the outcome measure of maternal mortality is shown in table 3.2. The intervention area was the district where the programme under evaluation was implemented and the non-intervention area was the district where there were no specific programme inputs.

Table 3.2: Study design for the outcome measure

	Intervention area	***Non-intervention area***
Before programme interventions	Maternal mortality estimate 1	Maternal mortality estimate 2
After programme interventions	Maternal mortality estimate 3	Maternal mortality estimate 4

Case Study 4: Study design in Ghana

The initial evaluation question identified in Ghana was as follows: In the poorest regions of Ghana, what is the cost-effectiveness of health insurance versus fee exemption for delivery care in (a) increasing the uptake of care, and (b) reducing institutional perinatal and maternal mortality, in the context of alternative quality assurance mechanisms? The (ultimate) objective of the evaluation was to provide evidence of the impact of alternative financing mechanisms on maternal and perinatal outcomes, particularly with regard to the poor. Here, the safe motherhood intervention 'strategies' of interest were health insurance and user fee exemption. At the time of the evaluation request, there had been a recent introduction of blanket exemption (abolition) of delivery fees, and differential implementation between regions and districts. Ghana was also in an early stage of implementing a national health insurance scheme.

At this stage, the evaluation design proposed was a quasi-experimental/factorial design using time-series; with two interventions – abolition of delivery fees and health insurance, versus control. It was thought possible to match districts, based on a situation analysis. The evaluation design was represented as shown in Table 3.3.

Table 3.3: Evaluation design in Ghana

		Fee exemption	
		Yes	**No**
Health Insurance	**Yes**	6 districts (1–6)	6 districts (7–12)
	No	6 districts (1–6)	6 districts (7–12)

Such a design as the one proposed would be strong, particularly as it could be done prospectively (not all districts had experienced implementation of either intervention). However, as the evaluation was being planned, investigation into the actual situation through reiteration of Steps 2 and 3 revealed feasibility issues. Firstly, all districts had started implementation of delivery fee abolition, although some more recently than others; and secondly, the implementation of the health insurance scheme was under-funded and only patchily implemented. The study design illustrated above had to be rejected and a retrospective 'before and after' design adopted instead, investigating only the effects of the fee exemption strategy on delivery care, as shown in table 3.4.

Table 3.4: Retrospective 'before and after' design

	Fee exemption
Before	6 districts with recent implementation
After	6 districts with established implementation and with recent implementation

3.6 Step 4: Identify the sub-questions

Once the evaluation design has been framed around its outcome measures as described in Step 3, it is time to return to your logic chain and to decide what other programmatic aspects you wish to test and to measure. To do this you will need to look at each link in the logic chain that you have developed and think about the research questions generated by each assumption. For example, if we were to use the logic chain in figure 3.2 for the village midwife programme, the process would be as follows.

Assumption 1: Training of more midwives

Relevant sub-questions:

- Were more midwives trained in the time period of the intervention?
- How many?

Assumption 2: Deployment of those midwives in rural areas

Relevant sub-questions:

- What happened to the midwives after training was completed, where did they go to?
- Did the numbers of midwives deployed in rural areas increase?
- Were the extra midwives who were trained actually deployed in rural areas?

You can continue to develop evaluation sub-questions from assumptions all the way through the logic chain. Identifying these sub-questions helps you to find out what needs to be measured (or described or understood) in order to know whether the programme has worked in the way it was expected to work. Case Study 5 presents the sub-questions identified in the Ghana evaluation, together with their relationship to the causal net.

Case Study 5: Sub-questions of the Ghana Immpact evaluation, with organizational framework for the sub-questions

As described above, the evaluation that was eventually carried out looked for evidence of the effect of the fee exemption strategy on delivery care and institutional deaths in Ghana.

A list of sub-questions was developed for the Ghana evaluation question using a causal net.

1. How is the policy on free delivery being interpreted at different levels and within different sectors of the health system?
2. What is the extent and pattern of implementation?
3. How far has knowledge about the policy reached consumers?
4. What proportion of the cost does the exemption fee represent for facilities?
5. What is the effect of the policy on the costs of delivery care to society/households?
6. How are these consequences distributed between different socio-economic and geographical groups?
7. What non-monetary costs are experienced and avoided amongst the poor?
8. To what extent do consumers' knowledge, beliefs and attitudes to traditional and modern delivery care influence the use they make of the fee exemption?
9. What is the effect of the policy on utilization at various levels and types of health facility?
10. What is the effect of the policy on quality of care at various levels and types of health facility?
11. Can changes in maternal and perinatal mortality be linked to the policy?

The causal net is shown in figure 3.5 with the sub-questions mapped on.

These sub-questions were then grouped into **four main evaluation questions,** which are listed as evaluation objectives, as follows.

1. To examine the process of implementation of the fee exemption policy
2. To determine whether fee exemption changes women's care seeking behaviour
3. To determine the consequences of fee exemption of health care and outcomes
4. To quantify the costs of fee exemption

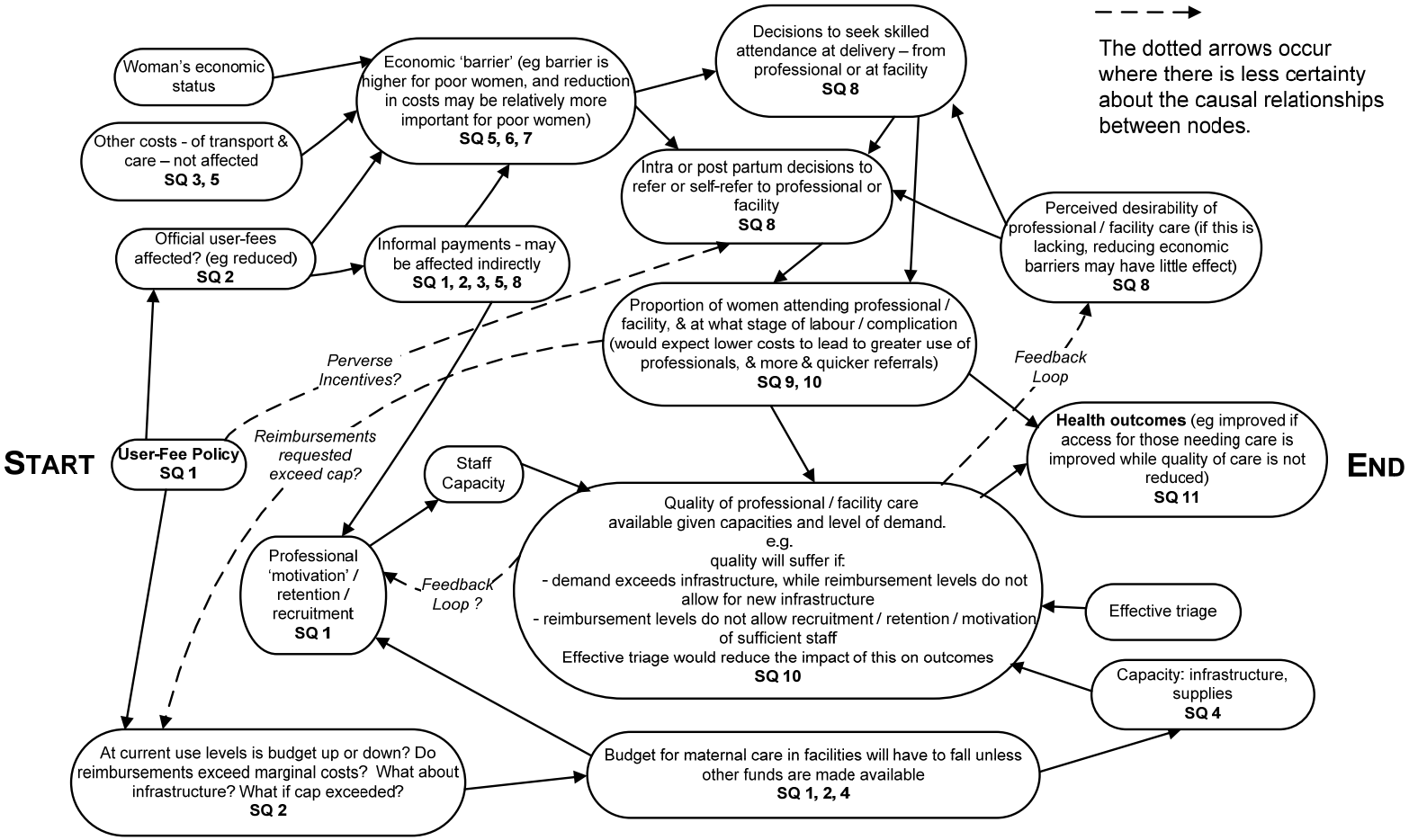


Figure 3.5: Causal net for Ghana evaluation question (EQ)

3.7 Step 5: Identify what you need to know and what tools you may need

Look again carefully at the examples of the sub-questions provided in Step 4. You will see that they are a range of types of questions. Think about the types of questions being asked e.g. is it asking 'why?' or 'how?' or 'how much?'? List each of your sub-questions in a table like the one table 3.5.

Table 3.5: Identification of data sources and prioritization of sub-questions

Sub-question	*Type of question*	*Possible data sources and tools*	*Level of priority*

Next, use the same table to identify the sources of information that could answer each sub-question (routine statistics, qualitative study, etc.). Have a look at the summary table of Immpact tools Module 1 (Tables 1.1, 1.2 and 1.3) to review the tools. You will also need to look at the existing literature on the topic and read up about other non-Immpact tools, (see Technical Annex F in Module 5) in order to assess comprehensively which tools might be best suited to your needs.

Tip:
Remember that the most efficient way of collecting data is to use existing data sources where possible (such as a census, demographic and health surveys, vital registration, health information systems, etc.), to minimize the costs of primary, or new, data collection. Familiarity with these existing data sources is important.

One way to decide between the different data sources, methods and tools is to make a list of the advantages and disadvantages of each. This is best done in a workshop and discursive setting, with a range of research expertise represented. Assess all the questions, and identify the pros and cons of each approach (see Case Study 6).

Finally, use Table 3.5 again to grade each question as high/medium/low priority for confirming or disproving your programme logic. You may not be able to test every assumption but it is important to make sure that you test the key ones.

Tip:
A good understanding of the programme to be evaluated is important to guide the selection of tools and methods. Immpact researchers have remarked that *'the causal net and the Immpact conceptual framework are inspirational'.*

In Ghana, when the research team developed the sub-questions and grouped them into four main evaluation objectives, the causal net from Step 4 was revisited and used to verify that the main areas of the evaluation were included correctly. The overall evaluation question was also revisited (Module 2) to ensure the four evaluation objectives contributed to the overall question. Much of this work was carried out in a workshop with the research team, and was not conducted by a single researcher alone. The main lesson is that it is important to go through the steps defined above systematically and revisit them reiteratively to guide the selections of outcomes and measurement tools.

Case Study 6 (part A): Choosing between different data sources, tools and approaches for measuring maternal mortality in the Immpact evaluation in Burkina Faso

In this situation we needed to know three things in order to calculate the maternal mortality ratio (numbers of maternal deaths per 100 000 live births).

- How to identify maternal deaths?
- How many deaths need to be identified?
- How to identify live births? (as a denominator)

There are several possible approaches, and each has its advantages and disadvantages as summarized in the table below.

Table 3.6: Approaches for measuring maternal mortality – advantages and disadvantages

Approach	*Advantages*	*Disadvantages*
Register deaths at health facilities	It could be easier logistically since we would only need to collect data at health facilities.	Unlikely to be representative, since the problem in the first place is lack of service uptake. Since we need retrospective data, it would depend on past records being available – unlikely.
Do a household survey	Needs a smaller survey effort than a full census.	Difficult to get timings on deaths, thus hard to separate before/after intervention groups.
Use an Immpact tool like SSS	Might be quicker and more efficient than more established methods.	This is a relatively new tool and has yet to be used in a wide range of different settings.
Conduct a household census	– Should give a robust answer. – Even effects like recall bias should cancel out between intervention/non-intervention areas.	Largest and logistically most complex approach.

Depending on your particular situation, and by identifying criteria most important in your case, you can decide which approach to choose. If it is a high priority to find a robust answer and if resource constraints are not high (as in the case of the Immpact study in Burkina Faso), you could make a decision to conduct a household census.

3.8 Step 6: Get to know the methods and tools you intend to use

Read carefully all the material that we have provided on the tools that you intend to use. The step-by-step guides provided in Module 4 give an overview of the different Immpact tools that can be utilized. As part of that process, also check the quality of any data sources you intend to use. This familiarization process will allow you to choose appropriate tools which you think can elicit useful findings (or answers) for each question.

Table 3.7 shows how, in the Ghana evaluation, the sub-questions, data collection tools, and data sources can be matched. For example, the first sub-question was to find out the extent of the pattern of implementation of the fee exemption policy: 'key informant interviews' had been identified as a priority methodology. To 'fit' the question to the tool in your case, ask yourself who might be able to give you information on the question, then decide how best to elicit that information from that source. Table 3.7 shows that it is likely that administrators and policy-makers will be able to supply this information (column 2), that these people are usually found in offices (column 3) and that the best way to elicit this information is by key informant interviews (column 4).

You will need to go through the list of your sub-questions and identify suitable tools in this way. To further specify your understanding of what the tool will do, you may find it helpful write down the objective(s) related to the sub-study using that particular tool, as shown under column 4 in table 3.7.

Next, it is useful to think through what sorts of indicator, or measure, that tool might provide. Examples are provided in the last column in table 3.7, indicating that the key informant interviews will give descriptions on fee exemption practices, understanding of the policy and so on.

Finally, to understand the way in which your selected tool works, think about how the tool will work in terms of the specific indicators you have identified. An example is provided from Case Study 6 (part B) below.

Case Study 6 (part B): Choosing between different data sources, tools and approaches in the Immpact Burkina Faso evaluation

Based on the 'advantages and disadvantages' table in part A of this case study, it was decided that a household census was the best tool to answer the sub-question. Ask the same three questions about the data needs as per the three bulleted questions in the case study above, but this time with reference to the selected tool, as follows:

- How to identify maternal deaths in a household census?
- How many deaths need to be identified?
- How to identify live births?

Consideration of these issues relating to a particular tool should then lead on to more in-depth considerations regarding what an acceptable recall period is for household enquiries about deaths; what eligible age range should be included for maternal deaths; and how to determine which adult female deaths were maternal. You can then ensure that you have the right tool or that certain adaptations may need to be made to the selected tool. More information on adapting tools is provided in Technical Annex C in Module 5.

Once you and your team are well acquainted with the methods, tools and instruments, it will then be necessary to write up descriptions of the intended sub-studies. To write this section of the protocol, it is useful to assign small groups of the whole research team to start writing up descriptions of each sub-study. Before they do so, it useful to provide researchers with an initial checklist of topic areas they should cover when describing their respective studies. The checklist of items should include:

- The title and hypothesis (which is a brief description of possible changes might occur due to the intervention and its effects) or research question
- Sub-study objective(s)
- How the objectives contributed to the overall evaluation question
- Sub-study design and comparisons
- What indicators the study would measure or issues it would explore
- Population for the sub-study
- How the sampling would be carried out, including the sample size
- Data collection methods
- Limitations and biases of the study and how these would be overcome
- Data analysis approach

As a form of internal peer review, each researcher should also present their sections to the whole team of researchers, who critique, ask for clarification and help to improve the hypotheses for each sub-study. This peer review process allows the whole team to understand each other's work and thus helps with the integration of the separate sub-studies, improving the researchers' ability to work toward the same overall evaluation objectives. The review process also encourages the individuals working with each particular sub-study to think and write more clearly about their own studies, thus improving their descriptions. Finally, it is a first stage toward developing better analysis and synthesis plans (Step 7).

It is important to consider all tools collectively in terms of the overall data that would need to be gathered. The evaluation team should see to find opportunities where instruments could be integrated, for example where the location or focus of data collection activities was complementary. Here again the importance of understanding

the other components of the evaluation was crucial for optimizing the value of these activities. Case Study 7 below discusses integration of tools in Ghana and Burkina Faso.

Case Study 7: To integrate instruments or not?

The possibility of integrating two separate sub-studies was explored at great length in Ghana. Referring to table 3.7, both the 'household costs' and 'utilization' surveys needed a random sample of women in the community. The 'household costs' survey required a larger sample size than the 'utilization' survey, so the intention was to try to integrate both studies by combining the two questionnaires as one and applying the combined questionnaire to part of the larger economic study sample. This would have been an ideal way of reducing costs for the data collection.

However, the 'household costs' study found that their sample size requirements were too large to make a house to house survey feasible. They decided to adopt a two stage sampling process, using community outreach child immunization points as their basic sampling frame to capture sufficient women experiencing different delivery modes for their study. This was possible as certain types of child immunization coverage in Ghana were of the magnitude of 95–98%.

Considerable debate followed over whether to delink the utilization survey from the economic study because of doubts about potential bias for the utilization study by using immunization points. In the end, a decision was made to separate the data collection for the two tools despite the higher costs. Findings from the 'utilization' study were fundamental to many of the hypothesis generated in other components and so the rigour of the findings from this study was paramount.

In Burkina Faso, integration of quantitative instruments was more successfully carried out. All instruments collecting data from health facilities were integrated, as were those collecting data from household surveys. This led to a longer preparatory stage to integrate the instruments, although once this was completed, data collection was rapidly completed (see 3.10.5 Timeline section under Step 8: Management).

Table 3.7: Sub-questions, data collection tools, and data sources from the Ghana evaluation

Sub-questions →	***Who/what can provide the information required from the sub-question?*** →	***Where can they be found?*** →	***Selected tool and related objectives*** →	***Summary of resulting indicators***
Objective 1: To examine the process of implementation of the fee exemption policy				
• What is the extent and pattern of implementation? • How is the policy on free delivery being interpreted at different levels and within different sectors of the health system?	Administrators/Policy makers	Offices	Key informant interviews • To supply context information on implementation of the policy • To assess impact of fee exemption on health financing at facility and individual health worker level	Fee-exemption practices, understanding of the policy, views on its impact, other context changes
• What proportion of the cost does the exemption fees constitute to facilities?	Financial records	Health facility	Financial flows tracking • To examine the financial impact of facilities of the exemption policy, setting that in the context of the overall health systems	Expenditure, financial flows, funding of exemptions
• How is the policy on free delivery being interpreted at different levels and within different sectors of the health system? • How far has knowledge about the policy reached consumers?	Clinicians and other providers	Health facility	Health worker incentives survey • To assess the financial impact of individual health workers of the exemptions policy and how that affects working practices • To determine factors that motivate health facility staff to provide quality care in relation to the policy of fee exemption PQOC facility survey • To determine factors that motivate health facility staff to provide quality care in relation to the policy of fee exemption	Provider characteristics, incentives, working practices, work load and work experience Provider motivation

Table 3.7 continued…

Sub-questions →	***Who/what can provide the information required from the sub-question?*** →	***Where can they be found?*** →	***Selected tool and related objectives*** →	***Summary of resulting indicators***
Objective 2: To determine whether fee exemption changes women's care seeking behaviour				
• How far has knowledge about the policy reached consumers? • To what extent do consumers' knowledge, beliefs and attitudes to traditional and modern delivery care influence how they make use of the fee exemption?	Women, men and other community representatives	Community	PQOC community survey • To investigate community knowledge and perceptions of free delivery care • To identify factors which may change care seeking behaviour	Women's characteristics and experiences
• What is the effect of the policy on utilization at various levels and types of health facilities?	Women	Community	Utilization Survey • To examine the effects of fee exemption on utilization of care. • To identify differences between utilization in different socio-economic groups	Utilization of health facilities and delivery attendant Poverty profile
Objective 3: To determine the consequences of fee exemption on health care and outcomes				
• Can changes in maternal and perinatal mortality be linked to the policy? • What is the effect of the policy on quality of care at various levels and types of health facilities?	Clinical registers and records	Health facility	RAPID • To quantify the effects of fee-exemption on hospital maternal mortality ratio especially during delivery Record review of quality of care • To quantify the effects of fee-exemption the quality of care provided in health facilities TRACE • To qualitatively assess the care provided in cases of maternal death	Maternal death Quality of care Adverse and favourable events in pregnancy care

Table 3.7 continued…

Sub-questions →	***Who/what can provide the information required from the sub-question?*** →	***Where can they be found?*** →	***Selected tool and related objectives*** →	***Summary of resulting indicators***
Objective 4: To quantify the costs of fee exemption				
• What is the effect of the policy on the costs of delivery care to society/ households? • What non-monetary costs are experienced and avoided amongst the poor? • How are these consequences distributed between different socio-economic and geographical groups?	Women. men and household members		Household costs survey • To determine the costs to the household of accessing delivery care, taking into account: – the impact of the exemption policy on household costs – the impact of the policy on equity – the social protection role of the policy for the poor	Societal, household and individual costs.

3.9 Step 7: Plan data entry, analysis and synthesis

Once you have decided on and familiarized yourself with the tools you are going to use for your evaluation and have brought them together, you will need to make plans for analysis and synthesis. The value of having a clear analysis and synthesis plan at this stage, and before any data collection is started, is related to ensuring that your evaluation will result in findings that fulfil the aims and objectives of the study.

A well thought out analysis plan helps to confirm that your study design is suitable and that your statistical techniques exist and are appropriate, and ultimately that your study provides useful results (Weinberg and Kleinman 2003). In qualitative analysis, it often true that the first stages of analysis are concurrent with data collection. However, this is not always the case and where appropriate, some initial planning for analysis, even for qualitative research, may be helpful. For the latter it may, for example, be appropriate to develop a provisional thematic framework for the analysis informed by a literature review on a topic (Pope et al 2000). More information on analysis for quantitative and qualitative research is available in Module 5. We also provide a proforma for an analysis plan in Appendix 1 of this Module.

Tips:

- Conducting a data analysis workshop is important in order to discuss and amend analysis plans. This workshop can pull complementary expertise together and generate fruitful discussion.
- Set responsibilities for analysis. During the analysis plan workshop, it is important to map analytical skills of the team and set responsibility within the team. This will speed up the data analysis stage.

The process of how to go about synthesising a complex evaluation should also start during the planning stage and not only after the data has been collected. At this planning stage, one of the early things to organize is the synthesis of the various sub-studies within the evaluation. The previous step of sharing information within the research team will already have brought out some key issues. When the analysis plans are drawn up, individual researchers should have a clear idea of what their study entails. While many potential sub-studies can be proposed at this stage, it is useful to call another meeting of the research team to discuss how to bring the whole evaluation study together. Here, the collective role of the evaluation team is to clarify the best fit of what was most relevant and practical for the evaluation. It will also help identify any 'gaps' in expertise or measurement tools, and indeed desirable duplications (situations where sub-studies working together will provide a more comprehensive explanation of change and relationships) as well as unwanted duplications, which should be avoided.

The technical and practical expertise within the evaluation team is an important consideration, given the somewhat dependent relationship between disciplinary knowledge and identifying relevant and practicable sub-studies.

To complement the meeting, it is also useful at this stage to formally identify priority data needs, particularly for the quantitative data. This can be done using the checklist provided in Appendix 2 'Identifying Priority Data Needs' and using the completed analysis forms discussed above. Although qualitative researchers follow similar principles, their approach tends to be more flexible and specific to their particular study. This process of prioritising studies is key to ensuring that a complex evaluation does not become unnecessarily large, complicated or costly.

Tips:

- Coordinating a multidisciplinary evaluation is a challenging but rewarding process. Some of the challenges we encountered in Immpact were as follows. Researchers who are experienced in particular areas are more likely to recognize opportunities for measurement, but at the same time, tend to only identify those relevant to their expertise. Furthermore, when compromises must take place to make best use of resources, the option chosen often depends strongly on what is understood by the evaluation team, key stakeholders and end users about the comparative value of each option. Yet, in Immpact we found that recognition and acceptance of these situations can lead to positive consequences. When decisions are made based on a sound understanding of the needs of stakeholders and on the practical opportunities and constraints to generating robust evidence, we found that it was not difficult to justify prioritisation of some areas above others.
- Since evaluations are fit-for-purpose, there is no one size-fit-all for tools and methods. Rather, existing tools, existing guidelines, previous evaluation practices must be revisited, sometime refined and adapted in order to answer the evaluation questions. The choice of evaluations tools is contingent to the type of evaluation, to the intervention to be evaluated, to the research questions to be answered but also to time and resources available.
- Although analysis plans, integration and synthesis of the different sub-studies are recommended during this step, it is recommended that the complementarity of sub-studies be continually revisited at various points in the study.

To synthesize findings, a research team meeting can be held to exchange information on preliminary findings between research teams. During this meeting, identify:

- findings from different studies which support each other (for example, the changes observed from two studies due to the intervention are both showing an increase in utilization);
- findings from different studies which contradict each other (for example, a qualitative study may show that women have experienced costs to access services, while an economic quantitative study has not detected any household expenditure to access services).

In this way, multiple perspectives (from different data sources and from different ways of measurement) of the effects of the intervention can be compared and contrasted. Findings from several different sub-studies which agree with each other improve the certainty of the finding, while findings which refute each other make findings less definite.

Case Study 8 presents the process that was undertaken in Ghana to synthesize research findings.

Tips:

- Set formats and responsibilities for synthesis and reporting. Standard formats for reporting facilitate the synthesis of findings from different axes of the evaluation. Reports are contingent to preliminary analyses and it is important to set realistic deadlines for reporting.
- The issues around how to synthesize findings in a complex evaluation is a science still in its infancy (Campbell et al 2000; Habicht et al 1999, Øvretveit 2002). Best practice and ideal means of bringing together findings from different studies, between qualitative and quantitative findings and within qualitative and quantitative research are yet to be identified. There are few other practical examples of how findings for complex evaluations have been synthesized (Lindsay 2002, Shediac-Rizkallah and Bone 1998).

Case Study 8 : Synthesis of research findings in Ghana

In Ghana, the change related to utilization of health facilities was obtained from a number of qualitative and quantitative sub-studies related to the following tools: key informant interviews, PQOC facility, PQOC community tools, utilization survey, household costs survey and record review of quality of care. The findings did not all agree with each other. Some of the studies indicated that utilization increased after fee exemption was introduced, while the utilization sub-study with the strongest design indicated that utilization had increased in one study area, but not in another. This led to the conclusion that it was not possible to indicate definitely that fee exemption would lead to increased utilization of services.

Conversely, the two studies on clinical quality of care, one qualitative and the other quantitative, suggested that there was no overall change in clinical care before or after the fee exemption policy was implemented. This was an unexpected finding, as one possible hypotheses of change was that the fee exemption policy would lead to lowering of care standards in health facilities because of over-utilization and stretching of capacity. The conclusion to the overall evaluation was that fee exemption did not change quality of care, in a situation where utilization rates of health services were not markedly increased.

In this case, it can be seen that the synthesis of findings was conducted in two ways. The first means of synthesis was conducted separately in two independent sub-questions – changes in utilization and changes in quality of care. The second means of synthesis was across the two sub-questions – in other words, the findings on uncertain changes in utilization rates was used to explain the reason for no change in quality of care.

3.10 Step 8: Management

Planning the management of a complex evaluation study is crucial to ensure that everything goes smoothly when the evaluation activities start. Assigning clear and effective leadership for the research team is even more important in complex evaluations. Field work also needs competent and experienced officers as part of the team to oversee the data collection logistics and arrangements. Continuity of the research team is essential, from the stage of planning and writing the protocol right up to the last step of dissemination and feedback.

The key areas which you need to cover in planning the management of the evaluation are:

1. Data collection
2. Data management and quality assurance
3. Resources
4. Timeline
5. Quality assessment
6. Ethical approval

3.10.1 Data collection

Having proceeded through the previous steps, the research team should already have started formulating ideas about how the data collection for each tool can be carried out. At this stage it is important to systematically plan the data collection. The ‘Standards and Principles for Primary Data Collection’ checklists in Module 5 (Technical Annex E) facilitate planning and standardization of data collection plans across different tools. Use these checklists to make sure that the sub-study teams are properly prepared. Data collection can usually be divided into collection at three main sites – health facilities, communities and using existing data (which may be found in statistical offices, ministry offices, etc).

In health facilities, teams of researchers can be based in each identified hospital or health centre to find, for example, records of maternal and perinatal deaths and complicated deliveries; to administer questionnaires, conduct interviews and discussions; to collect data from registers; to fill out case extraction forms; and to review health service data.

Tip:
An important consideration in health facility studies is the use of health providers' and others' time for data collection activities. These health personnel are often the most familiar with data collection in facilities, but using them for research purposes decreases their availability for clinical service provision activities. Adequate measures should be taken to ensure that service provision does not suffer. In studies which require significant use of health care providers' time, arrangements should be made to cover for time of those involved in the evaluation.

At the community level, research teams can carry out surveys, focus group discussions and individual interviews. These may be conducted from house to house or by targeting attendees at immunization and other clinics, or in places where the target population gathers in large numbers and where they may be representative of the community.

Tip:
Inevitably, on-the-spot decisions will have to be made during data collection activities. This needs to be taken into account when forming research teams. Either a senior researcher should be present to make such decisions or the research team should be able to communicate easily with a remote senior researcher.

As stated earlier, secondary analysis of existing data can be a valuable and cost-saving methodology. Data from previous Demographic and Health Surveys, and service provision assessment studies can be used to minimize and supplement primary data collection.

Co-ordination of data collection is important and the way in which this is planned will affect the resources and time needed. Immpact conducted data collection in two ways:

- In Burkina Faso data was collected in an integrated manner (see Step 6) which necessitated the development of integrated tools and instruments. This integrated data collection exercise can be grouped according to where data needs to be collected – for example, by combining all instruments with data required from household surveys in one, and data required from health facilities in another, and so on. This approach requires significant planning and preparation to combine different tools effectively, but it also means that the field work and resources for the data collection can be done very quickly and efficiently.

- In Indonesia and Ghana, there was some integration across tools using the same data site, but most of the data collection was conducted as individual sub-studies. This approach required more careful co-ordination of timing of studies and timetabling, but allowed each sub-study team to work in a manner which they felt was more manageable and flexible.

3.10.2 Training

Training and supervision of data collectors is necessary in order to ensure that data collection is actually carried out as planned; ensure that you have adequate resources (time, people and money) to conduct these activities. Training is key in managing fieldwork and should be completed just prior to the launching of a pilot test to avoid loss of skills and knowledge. The training should include:

- brief presentation of the research proposal
- detailed presentation of research objectives
- theoretical training of tools and data collection scenarios
- extensive practical training of tools and data collection scenarios
- pre and post evaluation.

Research training resources are provided in Technical Annex F in Module 5, and more information on aspects to consider for qualitative and quantitative data collection is available in Technical Annexes A and B in Module 5, and within the descriptions of each tool in Module 4.

> **Tip:**
> Allow sufficient time for training, which should be organized to include practical work in completing real data entry on samples. Monitoring of attendance of trainees helps to improve participation.

3.10.3 Data management and quality assurance

After planning for the data collection, it will be necessary to prepare and organize data management procedures. Again, Immpact recommends that these data management procedures are clearly in place before any data collection is started, so it is important to develop a data management procedure at this stage. The basic principles of this management procedure for quantitative data should include:

- Data capture in the field, dealing with training of data collections, monitoring and supervision of data collection, with inconsistencies identified, discussed and resolved and regular return of forms, which are logged into a notebook
- Data entry managed through procedures relating to computerized filing of tools, screen designing and means of data entry
- Data cleaning procedures such as merging of data sets and means of verification and correction of errors.
- Quality assurance procedures for approval of the final dataset
- Codebooks and file backups.

A simple example of such a procedural document is provided in Appendix 3: 'Quantitative data management procedure'. For qualitative data, management procedures are also necessary, but are more difficult to lay out in a structured way and will need to be addressed according to each individual study. General guidelines are provided in the Technical Annex B in Module 5 on qualitative research.

3.10.4 Resources

For a complex evaluation, it is clear that the human resources need to be multi disciplinary in nature. In Immpact evaluations, we have found that our teams have benefited from combined skills in health systems, epidemiology, medicine, midwifery, obstetrics, health economics, social science and qualitative methodologies. Field data collection is usually supervised by these researchers, but additional field staff will be required for data collection, often through temporary employment of suitably experienced individuals.

> **Tips:**
> - The recruitment of surveyors, supervisors and the evaluation core team can be time-consuming and costly, depending on the profile of human resources needed and the time of the year they are required (seasonal, holiday periods, semester terms). Do not recruit too early before the actual data collection.
> - Careful consideration should be given to legal issues surrounding recruitment of paid temporary staff. One should make sure surveyors, supervisors are fully aware of the benefit package before entering contract, and recruitment planning should take into account turnovers or firing. An average 20% extra staff can be pre-selected as a 'contingency'.

Clear data collection plans and identification of human resource needs are a vital contribution to setting a budget for the monetary costs involved. Cost categories are in the following areas.

- Resource use measures – time logs for staff, and travel
- Capital costs – buildings, equipment and vehicles
- Recurrent costs – personnel, office supplies, vehicles, building utilities, communication, dissemination activities, accommodation, venue hires, refreshments, maintenance, expenses and allowances

A full method and datasheet for costing of maternal health research studies is available as a 'costing manual' (SAFE 2003).

The actual costs of Immpact evaluations will clearly vary widely according to the evaluation needs, local costs, the number of tools used, the sample sizes required, and so on. As a very crude range, for field costs, you should expect to budget funds in the region of US$150,000 – $300,000 to conduct a complex intervention with multiple sub-studies. The costs may increase should you include the conduct of large household surveys in your evaluation for measurement of, for example, maternal mortality. The individual tools in Module 4 contain a section on 'Budget Implications' which in each case give more detail on the resource implications for conducting the study.

3.10.5 Timeline

A timeline needs to be developed in conjunction with the budget and resource planning: this will help with identification of resource requirements, timing of personnel needs, as well as overall co-ordination.

We provide two example timelines of data collection activities – one from Ghana (table 3.8) and the other from Burkina Faso (figure 3.6). As indicated earlier, Ghana did not integrate the data collection. The time for collecting data for all instruments stretched out over a period of 10 months. In Burkina Faso, the data collection was integrated and actual data collection took 4–5 months.

Tips:

- Set realistic timelines. When setting a timeline, one should be aware that there are steps we have control over and steps we do not. Taking time to establish a good working relationship with the intervention project officers, with health officials at ministry of health and with health care providers at district level can take considerable time.
- Other time-consuming areas includes contextually constructing survey instruments (integrated tool package), recruiting adequate personnel, devising a plan for data collection, supervision and quality control measures, managing data capture, identifying required logistic support and estimating costs.

3.10.6 Quality assessment

With the budget and timeline developed, you will be in the last stages of having a comprehensive, and well thought-out evaluation protocol. It is recommended to get this evaluation protocol peer-reviewed by colleagues who have an interest in evaluation. This peer review does not necessarily need to be 'external' at this stage, but can simply provide a fresh view, useful insights and recommendations for improvements. It should also identify any key gaps or concerns relating to the evaluation as a whole.

Tip:
In Immpact, we set up a panel for review of each protocol, comprising Immpact scientists and external individuals who were considered experts in particular disciplines or methodologies. The aim was to make the review group as multi-disciplinary as possible. International researchers with experience relevant to the evaluation objectives and with study design experience were sought. A feedback format was provided to the reviewers and you may find it useful as a checklist for your own quality assurance, as well as to provide guidance for your reviewers. (Appendix 4: Evaluation protocol feedback form).

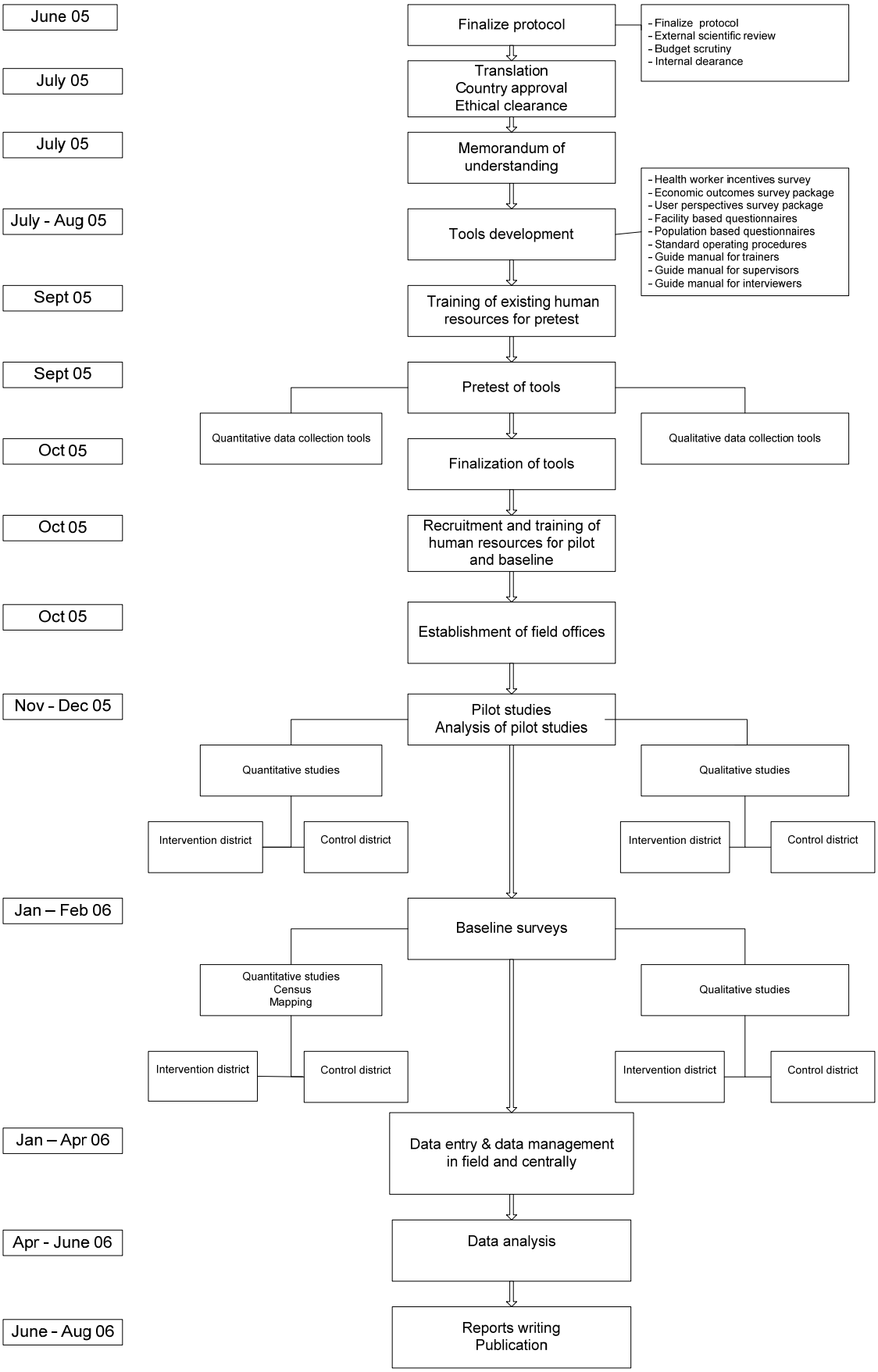
By working through both the negative and positive feedback from the reviewers, the evaluation team can clarify what is relevant to answering the evaluation question as well as what is practicable. Although some feedback may be technically correct and based on valuable experience it is not always practical in the specific context. Therefore, it is necessary for the research team leader to evaluate the comments and make decisions on changes to the protocol. A further safeguard can be that any major changes have to be resubmitted for review to ensure that any compromises are appropriate to the core needs of the evaluation.

Tips:

- Communication and feedback within the evaluation team and from the peer review were crucial influences. Face-to-face discussion was felt to be one of the most effective ways of communicating the best options. Constructively challenging the assumptions and expectations of the rationales for each sub-study was found to be needed in order to identify priority data needs.
- Sharing and constructive critique of each other's plans was also found to be important in order to ensure that the various sub-studies complemented each other, and together answered the evaluation question.

Table 3.8: Timetable for Ghana fee exemption evaluation study data collection

Timeline	*August 2005*	*Sept. 05 – Nov. 05*	*Dec 05 – Feb 06*	*Mar 06 – May 06*	*Jun 06 – Aug 06*
Key Informant Interview on policy implementation		(Oct) 14 Oct feedback Data Collection		2nd round data collection may be needed	Data entry, analysis & writing
Financial Flows Tracking	(Regions 1&2) Data Collection			Regions 1&2 (Second round may be required)	
Health worker incentives survey (HWIS)		Training & pilot; (start 7 Nov) data collection	Prelim analy.		
Perceptions of quality of care (PQOC, provider and community)		Train, pretest	(start Jan) Data Collection Regions 1&2	Transcription; Analysis	
Record review of quality of care (with perinate), RAPID and TRACE (data collection integrated as all are data from clinical case notes)		Training & pilot; (start 31 Oct) Region 1 Data Collection; (start 31 Oct) Region 2 Data Collection			
Utilization survey	Questionnaire Development		(1st Feb) Data Collection Regions 1&2		
Household costs survey		(Oct) listing and validation data collection; pretest, train 7 Nov	(Jan) Data Collection Regions 1&2	Analysis & write up	



3.10.7 Figure 3.6: Flow chart of timeline – example from Burkina Faso

3.10.8 Ethical approval

Depending on how the evaluation will be funded, you may need to submit your final protocol to funding bodies and you will need to submit it for ethical approvals. Your protocol should explain how this will be done.

In the planning phase, make sure that you are aware of the processes required for ethical approval, the formats required for submission of documentation and so on. Knowing this before you start writing your protocol may save you valuable time. No research undertaking, including pre-testing or pilot testing, should commence before full ethical approval is in place.

> **Tip:**
> Scheduling ethical clearance and getting feedback from the ethical committee can also take several months.

3.11 Step 9: Plan dissemination and feedback

Dissemination and feedback are commonly thought of as activities which need to be conducted at the end of an evaluation. As indicated earlier, Immpact found that having close links with the 'audience' for the evaluation throughout the process of planning is crucial. Sometimes, key contextual factors or policies change suddenly, and can affect the design and need for your evaluation. Policy and programme managers may not have completely understood the implications of the evaluation question they had formulated with the researchers (Module 2). The evaluation objectives, sub-questions and types of finding you expect to get from the use of various tools need to be fed back to the stakeholders during the various steps. Resources and time should be planned to ensure that dissemination activities are conducted both during and after the evaluation.

Dissemination and feedback mechanisms are many, but it is useful to classify these activities into community, national and international.

3.11.1 Community feedback

Firstly, a dissemination plan will be required to feedback information about the study to its participants. This may involve organizing a village meeting, going back to the health facilities where data was collected, contacting local/district organizations or local/district assemblies involved in, or having a stake in the study. Feedback of research findings can raise expectations and lead to questions from the community regarding action for increasing community amenities. It is important to have initial discussions with local government and other action-oriented organizations so that the responsibility and plans for future action is shared within the concerned individuals and organizations.

3.11.2 National feedback

The second level of feedback would be at national level. Often, it is national governments, or organizations working at national level, which might have requested or commissioned this study in the first place. Dissemination will be required not only to the commissioning agency, but also to other stakeholders who have an interest in the field of work, and particularly to the programme implementers (at local or national level) themselves.

3.11.3 International dissemination

International dissemination is also crucial. Global policy needs to be informed by knowledge and experience in the field in order to develop feasible and useful guidance for other countries. Many countries may be embarking on programme interventions similar to the one you evaluate – and your findings may be relevant despite differences in situation and context.

Tip:
Evaluation findings can be highly sensitive and care should be taken to consult widely and informally before formal release is arranged. Immpact recommends the conduct of evaluations in a participatory manner, from the formulation of the evaluation question to the final release of findings. Even if feedback and dissemination are conducted throughout the evaluation process, be aware that the people affected by the evaluation may not have fully considered how they might deal with adverse or negative findings (such as evidence that their programme has had no measurable effect. It could be useful to conduct a workshop with key stakeholders early in the evaluation process, giving concrete (fictitious) examples of both positive and negative findings and discussing how these will be dealt with. Preparation of this nature may help to mitigate against any adverse effects of the evaluation itself.

Useful forms of dissemination include the media, face-to-face presentations and workshops, policy briefs, written reports and publications. Many of these means of dissemination have specialized formats, and you may need extra expertise to use them effectively.

3.11.4 Authorship of papers for publication

As evaluators with a research grounding, it is expected that you will be familiar with the publication of papers in scientific journals. Be aware that authorship issues can be contentious and should also be planned before any data collection starts. It is advisable to set up authorship guidelines and name authorship in advance to avoid frustration, misunderstanding or conflict.

More guidance on Immpact's approach to good practice in collaborative partnerships, ethical conduct, quality assurance and authorship matters is described in Jentsch and Hussein (2006).

3.12 Conclusion

Having followed these steps, you will have a comprehensive evaluation protocol detailing a procedure for conducting an evaluation. You can now start implementing your planned evaluation activities. Key guiding principles for good evaluation practice have been described in this module and can be summarized thus:

- Be participatory in developing evaluations – interact widely, invest in good working relationships, feedback often and be inclusive;
- Spend time planning, consulting, describing interventions and context well, with attention to quality control at all stages of your evaluation;
- Invest in making your tools contextually relevant, and use existing data where possible;
- Make your evaluations efficient, and they may not necessarily be costly – prioritize, review and omit unnecessary duplications while valuing multiple perspectives;
- Set up your evaluation team with representation of multiple skills, clear roles, continuing and good leadership in mind;
- The formulation of a well-designed complex evaluation is not a linear process, so keep reiterating and revisiting all the steps in this module.

REFERENCES

Campbell M, Fitzpatrick R, Haines A, Kinmonth AL, Sandercock P, Spiegelhalter D, Tyrer P (2000) Framework for design and evaluation of complex interventions to improve health. BMJ 2000, 321 (7262), 694–696.

Habicht JP, Victora CG, Vaughan JP (1999) Evaluation designs for adequacy, plausibility and probability of public health programme performance and impact. International Journal of Epidemiology 1999, 28 (1) 10–18.

Jentsch B, Hussein J (2007) Developing a Guideline of Good Practice in a North-South health research collaboration: a participatory approach Social Policy and Society, 6 (2),179–191.

Leeuw FL (2003) Reconstructing Program Theories: Methods available and problems to be solved. American Journal of Evaluation, 24 (1), 5–20.

Lindsay AC (2002) Integrating quantitative and qualitative methods to assess the impact of child survival programs in developing countries: the case of a program evaluation in Ceara, Northeast Brazil. Health Education & Behavior, 29 (5), 570–584.

Øvretveit J (2002) Action evaluation of health programmes and changes. Abingdon, Radcliffe Medical Press.

Pope C, Ziebland S, Mays N (2000) Analysing qualitative data. BMJ, 320 (7227), 114–116.

SAFE International Research Partnership (2003) SAFE Strategy Development Tool: A guide for developing strategies to improve skilled attendance at delivery. The Dugald Baird Centre for Research on Women's Health, University of Aberdeen, Scotland. Available on line at http://www.abdn.ac.uk/dugaldbairdcentre/safe.

Shediac-Rizkallah MC, Bone LR (1998) Planning for the sustainability of community-based health programs: conceptual frameworks and future directions for research, practice and policy. Health Education Research, 13 (1), 87–108.

Weinberg JM, Kleinman KP (2003) Good study design and analysis plans as features of ethical research with humans. IRB: Ethics and Human Research, 25 (5),11–14.

Appendix 1: Analysis plan proforma

Complete the Proforma below for each sub-study within the evaluation.

Title of overall evaluation:	
Title of substudy:	
Title of Instrument:	*Attach the latest copy of your tool and its instruments*
Researcher name:	
Data entry method :	*e.g. single/double entry, what software for entry (not necessarily same as software for analysis), data management issues, data cleaning, coding, final documentation of data entry process*
Overall evaluation question:	
Research question(s) of this tool:	
Sample size:	*e.g. proposed numbers, rationale for this, types of cases, from which comparison areas, what power and confidence intervals are from expected change (e.g. 'at 80% power and 95% confidence interval able to detect changes of 20% in certain key indicators')*
Data Analysis	
Software:	
Level of analysis:	
Independent variable(s):	
Dependent variable(s):	
Prechecks:	*e.g. how will the variables be checked for normality, will there be any transformations (e.g. logarithmic)*
Manipulation:	
Missing Data:	*e.g. how will you assess the level of missing data, how will you deal with missing data*
Statistical Analysis:	*e.g. what statistical method and study design will be used, how will it be used, how will you deal with varying sample sizes from different data sites, will trends, correlation, causality, regression etc be analysed and how, how will the findings be presented (e.g. an index, proportion etc)*

Please include here either the tables that will be used in the analysis section of the report, or if you do not know at this stage what tables might be included, please give an example of a dummy table / graph as shown below.

Example of Dummy table
(for example if sample is number of cases seen in health facilities, ie caseload)

	Caseload (mean, sd)	
Central (before)		*This could be graphed e.g. by region and by time, showing both the raw and caseload index data*
Central (after)		
Upper Volta (before)		
Upper Volta (after)		

Appendix 2: Identifying priority data needs

This checklist is intended to help identify the priority data needs of each country-specific evaluation. These needs should be driven by the specific question (and sub-question) being addressed in these evaluations.

1. Rank the sub-questions within the overall evaluation question

2. Starting with the highest priority sub-question, make an inventory (list) of the principal (primary/essential/key) variables needed to answer this. Then list the subsidiary (helpful, interesting) variables. Repeat this process for the lower order sub-questions.

3. Prepare a tabulation plan for the principal variables, and so identify the required co-variates. This may be done either as a listing of frequencies or cross-tabulations, or by drawing-up dummy tables.

   For example, the key variable '% of free deliveries' may be included in the tabulation plan as:
   'Cross-tabulation of the number of deliveries and number of free deliveries, by district';
   or by drawing-up the following dummy table:

District	No. of deliveries (a)	No. of free deliveries (b)	% free (a/b)

4. Indicate on each dummy table, where you expect the necessary data to come from.
   - Facility-based: register review, routine statistics, case-note review, etc.
   - Population-based: survey; surveillance, etc.

   Give names to the data capture instruments to be used.

5. Amalgamate this information to create a second inventory – this time indicating all the variables which are expected to be gathered from each source and instrument.

6. Produce a listing to indicate which researcher is expected to provide guidance on the collection of the principal variables and by when

7. Identify the process and personnel required to produce integrated instruments/points of data capture, and the timeline.

8. Reflect on the checklist for standards and principles for data capture. (see checklist in Technical Annex E in Module 5)

Appendix 3: Example of quantitative data management procedure for Immpact Ghana

The following steps will be conducted for all quantitative components to ensure the highest quality and consistency of data.

1. Paper data management

1.1 Data capture in field

Quantitative data collection will include the following steps:

- Training of data collectors for relevant instrument;
- Monitoring of data collection in field (all forms completed by one person and checked by another);
- Inconsistencies identified discussed with data collector and corrected when necessary (by reassessing original form or revisiting interviewee)
- Data collection forms returned at regular intervals for filing.

1.2 Filing of forms

All questionnaires arriving in the office will be logged into a notebook by questionnaire type, number of questionnaires, date received and person receiving them. They will be grouped into batches of 100 and given batch numbers, which will be written on the questionnaires as well. A unique number will be assigned to every questionnaire to make it identifiable and also to compare double entry files (i.e. verification). A form, with date for data entry, initials of data entry clerks, and an indication of whether the batch was first or second entry will be placed on top of each batch. Each batch will be labeled and filed or packed in boxes appropriately for data entry.

2. Computerized data management

2.1 Filing on the computer

A separate folder will be created for every instrument/tool on every computer where data was entered. Each filename will have the first two initials of the questionnaire name, batch number and the entry number (i.e. the numbers ‘1’ or ‘2’ added to each filename) to indicate if the file was the first data entry or second (double entry, see section 3.1). The first and second entries will be used to run the verification (comparison) program to ensure data quality. Each batch of questionnaires will be entered into the file bearing its batch number and the questionnaire name using the following procedure:

- The first 2 digits will be derived from the questionnaire’s initial name (i.e.QC for Quality of Care.
- The next digit is the letter ‘B’ for batch and then the batch number, 01, 02, 03 (e.g. B01).
- The next 2 digits are the letter ‘E’ for entry number, which will be denoted by numbers ‘1’ or ‘2’ to indicate if the entry is the first or second (i.e. E1 and E2).
- A file ‘QCB01E1.REC’ implies the first batch of quality of care questionnaire and also the first data entry file

By following the naming pattern above the questionnaires will result in filenames as follows.

Sample Filename Chart:

Questionnaire	*Batch Number*	*Entry Number*	*File Name*
Quality of Care	01	1	QCB01E1
		2	QCB01E2.
	02	1	QCB02E1
		2	QCB02E2
Ghana Health Worker Incentives Survey	01	1	WSB01E1
		2	WSB01E2
	02	1	WSB02E1
		2	WSB02E2
Ghana Health Worker Incentives Survey, TBA	01	1	TBAB01E1
		2	TBAB01E2
	02	1	TBAB02E1
		2	TBAB02E2

2.2 Screen Designing

A screen for data entry and a check file will be designed for each type of questionnaire. The check files aims to limit errors during data entry by incorporating range and consistency checks, as well as to identify errors that might have escaped field workers and file editors. Such errors will be corrected and caution given to the field worker and the supervisors if possible.

3. Data entry

EPI6 will be used for data entry. The data entry clerks will be trained on each instrument and made to enter some sample questionnaires to ascertain their competence in every instrument before the actual data entry for that questionnaire commences.

The data will be entered in batches, and after each data entry the data entry clerks will append their initials on the form placed on top of the batch of forms. The data entry clerk will also indicate whether the entry was the first or second and the date the data was entered (see section 3.1).

3.1 Double data entry

A double data entry system will be used. Two data entry clerks will enter the same information, one after the other into separate data files. In this case a batch of every questionnaire is a file, and as described earlier, the first and second data entries will assign the suffix E1 and E2 respectively to the files. This means all questionnaires will be entered twice, by two different data entry clerks.

A verification program will be run for these two separate files, which compares every field entry for any difference. A difference between the two entries denotes an error in one of the files. The errors will be printed and given to the data entry clerks to crosscheck with questionnaires and correct them. This ensures clean and high-quality data.

4. Data cleaning

4.1 Merging and Final Data Cleaning

All the different batches of each instrument will be merged together after the data was verified and errors due to data entry will be corrected. A thorough cleaning program will be run on each merged final file. Field officers will be contacted for clarification on all inconsistencies from fieldwork for correction. All errors, problems or inconsistencies that needed the solution of field officers will be documented.

4.2 Variable and Value Labels

After all the corrections are made, (that is cleaning) the data will be exported to SPSS where variable and value labels will be added to all the variables. The variable labels describe the variable or the 'wording of the question' in detail and the value label assigns the coded responses (e.g. yes for '1' and no for '2') to the variables for easy references during report writing.

4.3 Quality assurance

In order to ensure the highest quality data before analysis, each dataset will undergo the following procedure (figure 3.7).

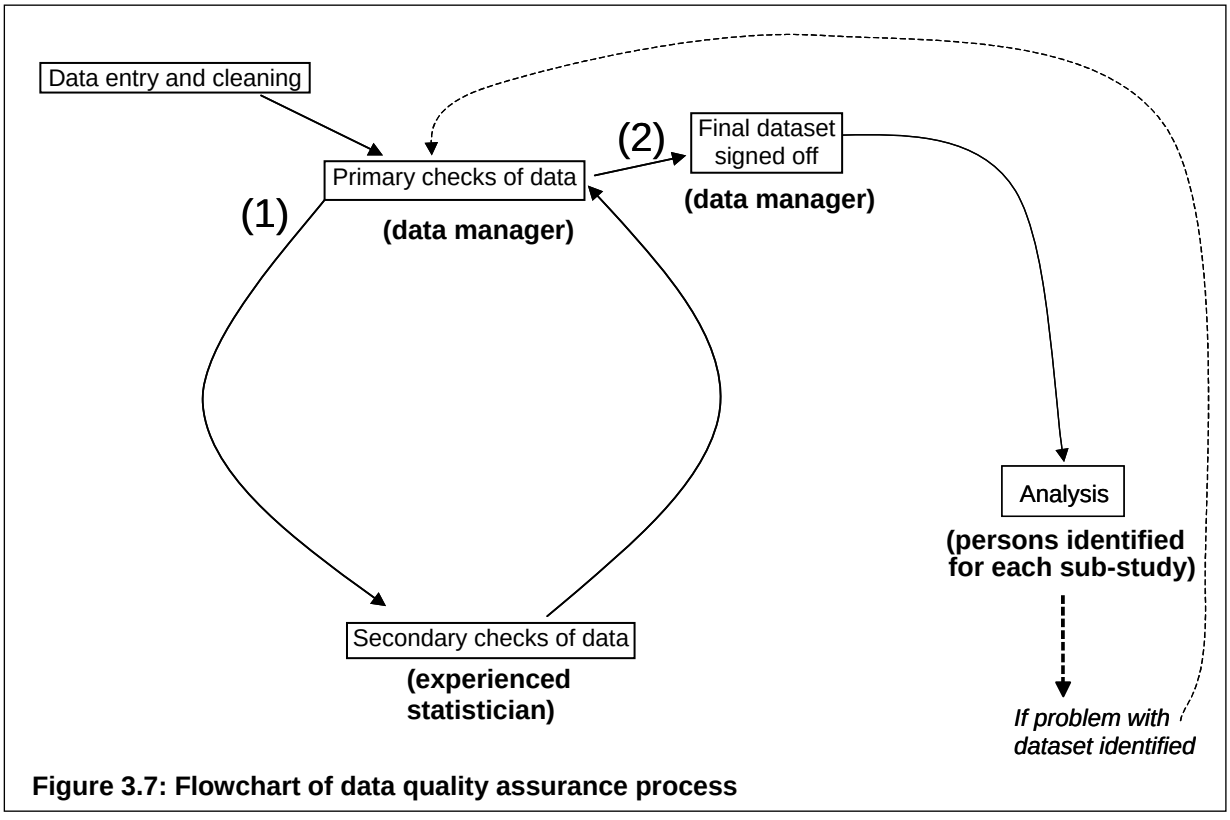


Figure 3.7: Flowchart of data quality assurance process

Clean data will be checked by the data manager. When s/he is satisfied that the data is clean from primary data checking, the dataset will be sent to the statistician for independent secondary checks. If any further areas that may need checking are found, the statistician will send queries to the data manager who will consult the original forms and conduct further cleaning. When data manager and statistician are both agreed that all queries are dealt with, data manager will sign off the final dataset. This dataset will be sent for analysis. If during analysis other errors imply some further data cleaning, a complete list of points will be sent to the data manager. If substantial changes are made to the dataset then the same two levels of data checking will be undertaken and the data manager will sign off a new version of the clean dataset and sent it out again for analysis.

Analysis should not be conducted on any dataset that has not been formally signed off as a final version by the data manager and statistician. This will allow tracking of the correct version of the data set if any analysis requires revisiting.

Key notes:

- Datasets should be issued and dealt with one at a time, rather than collected and sent to the statistician all at once.
- All documentation should contain the author's name(s) and version number
- All versions of datasets should be identified by a version number, date and responsible person(s). The final version of a dataset for analysis should be clearly signed off and marked as so. Care should be taken to ensure that if analysis has started and some final data cleaning is done subsequently, these changes to the dataset must be logged and the file identified as a different final version (a change control log).

5. Others

5.1 Code Book

A codebook will be developed for all the questionnaires. These codebooks give details of each variable, that is, the variable name, the variable label (wording of the question), and the value label (the possible responses). The codebook is similar to variable and value labels in the SPSS but is in Microsoft word so that those who could not use SPSS could easily understand the tool/questionnaires.

5.2 File Backup

Backups of data entry files are made daily on CDs to ensure file security. Backups of final datasets will also be made. These are stored in different premises.

Appendix 4: Evaluation protocol feedback form

Name of reviewer:	
Title of protocol reviewed:	

Reviewers' comments:
Please provide your critique of the protocol in the sections outlined below. If you think the section has not been well addressed, please suggest how it may be improved.

1. Introduction, background and preliminary work sections Do these sections explain the rationale and work up toward the evaluation question? Is there adequate evidence of adequate literature review and preliminary work?
2. Evaluation questions and objectives Are the questions and objectives clear and appropriate?
3. Study design Provide comments on the study design and the methods. For example: Is the study design clear? Are appropriate methods and measures used to answer the evaluation question? Are a sufficient range of methods used? Are the comparisons clear, and is the strategy being evaluated sufficiently described? Have biases and limitations been described well? Are relevant data sources identified? Have design issues for qualitative studies been well described?
4. Sample size Is the proposed sample representative of the target population (if appropriate)? Is there a justification for both qualitative and quantitative methods? Are the necessary statistical calculations completed? Has statistical advice been taken?
5. Data analysis Is the process and method of analysis been well described? Is it appropriate?
6. Utility Is sufficient justification given regarding the utility of the evaluation? Does the evaluation address and identified evaluation need or research gap? Does the evaluation have potential to improve policies and programmes?
7. Management of research Are the management processes adequate? Are sufficient details provided regarding recruitment, infrastructure requirements? Are the resources required sufficient for a timely, high quality evaluation to be completed? Is the budget adequate?
8. Ethics What ethical considerations might arise? Are ethical review processes adequate?
9. Any other comments?