

# Operational Research to Improve Health Services

## A Guide for Proposal Development

2011

*Desmond Tutu TB Centre*  
*International Union Against Tuberculosis and Lung Disease*

Nulda Beyers  
Mareli Claassens  
Donald A Enarson  
Pren Naidoo  
I.D. Rusen  
Brenda Smuts

This manual was produced from materials developed for the Operational Research Assistance Project, in partnership with The Union through TREAT TB, with funding from the United States Agency for International Development (USAID). The contents do not necessarily reflect the views of USAID or the United States Government.

*Suggested citation*

Beyers N., Claassens M., Enarson D., Naidoo P., Rusen I., and Smuts, B. Operational Research to Improve Health Services: A Guide for Proposal Development. Paris, France: International Union Against Tuberculosis and Lung Disease, 2011.

© International Union Against Tuberculosis and Lung Disease (The Union)  
68 Boulevard Saint-Michel, 750006. Paris, France.  
September 2011.

All rights reserved.

No part of this publication may be reproduced without the permission of the authors and the publishers.

ISBN: 978-0-615-57180-5

## **PREFACE**

This Guide was developed for use in the Operational Research Assistance Programme (ORAP) of the Desmond Tutu TB Centre (DTTC), Stellenbosch University, South Africa. The ORAP was developed by the DTTC in collaboration with the International Union Against Tuberculosis and Lung Disease (The Union) and funded through the TREAT TB programme with a grant from the United States Agency for International Development (USAID).

The Guide contains course material used in the ORAP; this course, taught over a period of five days, has its objective to develop a research proposal that can be carried out in the following year in the health facilities that participate in the National Tuberculosis Control and Prevention Programme of the Republic of South Africa.

The ORAP initiative is based on the firm belief that local health services personnel involved in delivering tuberculosis services are those most likely to know the challenges to delivering high quality services. Moreover, the initiative aims to partner personnel from the health services sector with a local academic institution enabling the two sectors to work together to develop the research proposal.

In order to carry out the research contained in the proposals developed during this workshop, the DTTC organises access to experts in statistics and ethics review processes and assigns each group a senior mentor to assist in the implementation of the research. The partnership of all stakeholders is a main focus in developing, carrying out and reporting the research. This is particularly important to achieve the ultimate objective of the research, which is to change policy and/ or practice to enable provision of the highest quality of services possible to those affected by tuberculosis.

The text is a compilation of our experience in this initiative and is shared with others, not in a definitive sense but instead as a practical and humble record of our experience. There are undoubtedly many other approaches that are equally valid in aiming to achieve the objectives of the initiative but we offer this record of our experience and hope that it might be useful for those interested to operationalize the critical thinking that is so crucial to improvement in our health services.

<b>Table of contents</b>	<b>page</b>
<b>1. What is operational research?</b>	
1.1 Operational research – a working definition	1
1.2 Importance of operational research to health programmes	1
1.3 Who conducts operational research?	2
1.4 Operational research – what support is needed at country level?	2
1.5 Operational research – how do we measure success?	3
<b>2. What is the stage?</b>	
2.1 Justifying the need for research	4
2.2 Overview of operational research	5
<b>3. What is needed to get started</b>	
3.1 Research question	9
3.2 Proposal outline	11
3.2 Using routine data for research	12
<b>4. What are the basic components?</b>	
4.1 Study design	18
4.2 Population	23
4.3 Selecting a sample	25
4.4 Variables and definition	27
4.5 Measurements	29
<b>5. What are the building blocks?</b>	
5.1 Data management	32
5.2 Comparison (Analysis plan)	37
5.3 Error and bias	39
<b>6. What is the justification?</b>	
6.1 Literature review and referencing	43
6.2 Strengths and limitations	45
6.3 Significance and impact	46
<b>7. What is the context?</b>	
7.1 Ethical issues and completing ethics forms	48
7.2 Regulatory aspects of research	51
<b>8. What does it take to move ahead?</b>	
8.1 Work plan and timelines	52
8.2 Roles and responsibilities	53
8.3 Budget	55
8.4 Getting support for research	59

## **1. What is operational research?**

### **1.1 Operational research – a working definition**

Operational research has been defined variously depending on the setting, the researcher and the nature of the research. The International Union Against Tuberculosis and Lung Disease (The Union) and many of its research partners adhere to a simple, straightforward definition of operational research as follows:

“research into strategies, interventions, tools or knowledge that can enhance the quality, coverage, effectiveness or performance of the health system or programme in which the research is being conducted”<sup>1</sup>

Supporting this practical definition are three basic principles that guide operational research efforts:

1. The health programme/system in question should have well-defined goals/objectives.
2. Constraints and obstacles that prevent these objectives from being achieved must be identified, prioritized and articulated
3. Research questions need to be asked to address these constraints.

A common understanding of what is meant by operational research as well as agreement on these key principles is necessary starting points for the successful development and implementation of relevant operational research as outlined in this guide.

### **1.2 Importance of operational research to health programmes**

Operational research has been increasingly recognized as vital to the strengthening of health programmes. For example, the expanded Stop TB Strategy explicitly includes operational research as one of the key components for successful global tuberculosis control. The Global Fund to Fight AIDS, Tuberculosis and Malaria (GFATM) recommends that programs spend between 5 and 10% of their grant budget on monitoring and evaluation, including relevant operational research. The percentage of approved GFATM grants that included an operational research component increased from 19% during funding rounds 1 to 5, to 58% in round 7. On average, US \$400,000 was requested for operational research per proposal, accounting for 3%-4% of the total requested budget.

The true value of operational research for health programmes is not the inclusion in global plans or dedication of resources; rather it is the impact of research results on programmatic and policy decisions and practice. Several recent examples highlight this direct relevance.

For example, in a basic, low-cost operations research study, researchers from The Union’s operational research course in Paris demonstrated that the third sputum sample collected from patients with suspected TB added relatively little to the yield in

---

<sup>1</sup> Zachariah R, Harries D, Ishikawa N, Rieder HL, Bissell K, Laserson K, Massaquoi M, Van Herp M, Reid T. Operational research in low income countries: what, why and how? *Lancet Infect Dis* 2009; 9: 711-717.

detecting new TB cases<sup>2</sup>. This research led to national and global policy changes recommending two sputum samples instead of three in many settings, resulting in significant savings in terms of costs and human resources.

Not every research effort can be expected to influence global policy. Indeed, much of the added value of operational research lies in its ability to address and solve local problems in the delivery of high quality health services. A necessary starting point is the focus on topics of direct relevance to health programmes with the potential goal of adding to the evidence base upon which programmatic and policy decisions are made.

### **1.3 Who conducts operational research?**

While the ultimate responsibility for development and maintenance of a strong operational research component lies with government health authorities, operational research may be undertaken by many people, including front-line health workers, local academic institutions, non-governmental organizations, community organizations and international partners. Ensuring the relevance and quality of the work undertaken as well as the coordination of the various partners, must remain the role of the government health programme (e.g., the national tuberculosis programme or the national AIDS programme).

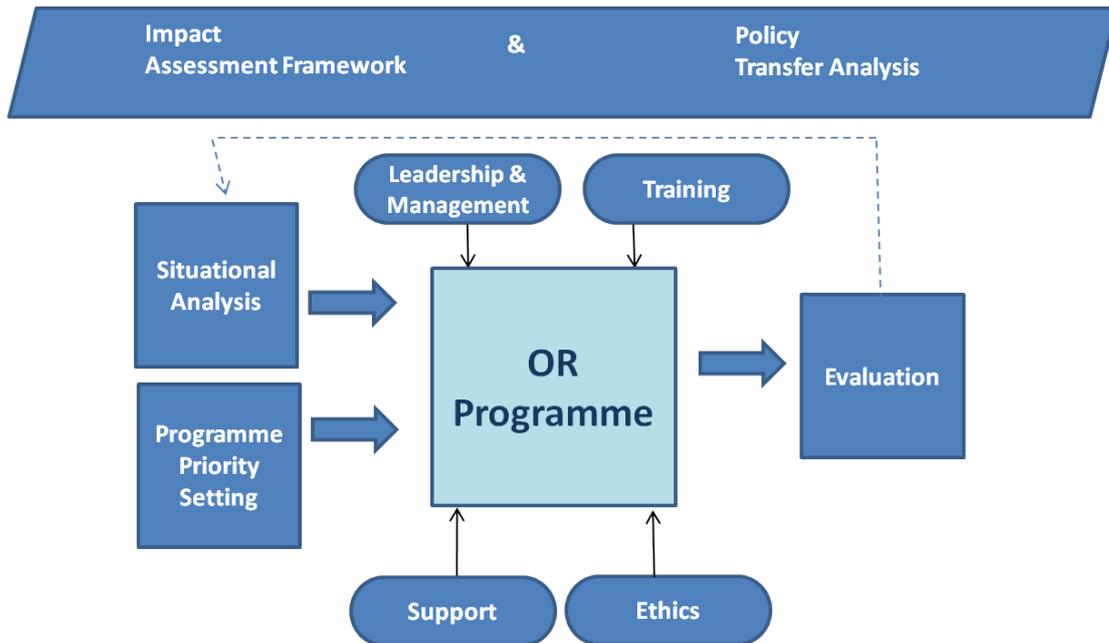
### **1.4 Operational research – what support is needed at country level?**

While the concept of operational research as an essential tool for health programmes is widely accepted, challenges to successful implementation of comprehensive operational research (OR) activities at the country level are numerous. In particular, many countries still operate in the absence of a detailed, systematic plan of research with clear linkages to programme priorities, thus the impact of any research efforts is clearly limited. Similarly, the implementation of OR training or actual research projects, in the absence of carefully conducted situational analysis prevents many countries from achieving their desired goals. In addition, the appropriate external sources of support, both financial and technical, must be in place at all stages of planning and implementation. These resources must be sufficiently flexible to allow local partners (rather than those providing the funds or external experts) to set priorities. Such resources are insufficient or absent at some or all stages of OR implementation in many countries. Finally, the Impact Assessment Framework and Policy Transfer Analysis are particularly useful tools to support countries in the planning, implementation, and evaluation of OR.

A comprehensive framework of country level OR support, as developed by The Union, is outlined in **Figure 1**. A document devoted to describing the details of country support for operational support has recently been prepared and is available from The Union.

---

<sup>2</sup> Katamba A, Laticevschi D, Rieder HL. Efficiency of a third serial sputum smear examination in the diagnosis of tuberculosis in Moldova and Uganda. *Int J Tuberc Lung Dis* 2007; 11: 659-664.



**Figure 1. A comprehensive framework of country-level OR support**

### **1.5 Operational research – how do we measure success?**

As with any investment in the health system, clear and measurable indicators of success are necessary to assess performance and guide future activities. Within a programme of operational research, the most important measure of success is performance of the health programme, e.g., improved indicators of tuberculosis control. However, this ultimate aim is not always easily or immediately attributed to the operational research programme. An intermediate marker of impact of operational research is an analysis of policy uptake in relation to research results. Finally, publication of peer-reviewed articles related to the research is an objective, measurable outcome of the programme which assures the research has been carried through to completion, is of sufficient quality to successfully complete the peer-review process and the results have been disseminated within the broader health community.

## 2. What is the stage?

### 2.1 Justifying the need for research

According to the Oxford dictionary, research is defined as “the systematic investigation into, and study of, materials and sources in order to establish facts and reach new conclusions”. “Operational research (OR) consists of the application of analytical methods to help organizations make better decisions when solving complex problems”<sup>3</sup>. Operational research is therefore also called “decision science”.

In health services, there is a need for research with the purpose of creating new knowledge that focuses on problems or challenges within the field of interest. These problems or challenges should be locally relevant and the new knowledge must lead to action which will change policy, practice or management thus leading to improved delivery of health services.

Often, the success of research is measured is by the number of resultant publications. Although there is no doubt that research and research results must lead to publications and the number of publications is one outcome measure of research, results and new knowledge must not be a dead end; results must also lead to changes in clinical, laboratory or services management and possibly policy.

Researchers and health care workers in low-income countries (LIC) have an additional responsibility – they must ensure that: (1) authorship is fair, (2) research becomes sustainable and (3) the quality of research and data are optimal.

Number (%) editorial board members stratified by country human development index (HDI)			
Number members on Editorial Board	High	Medium	Low
315	223 (71%)	76 (24%)	16 (5%)

Number of articles	Number of authors	All authors (%)			Last author (%)		
		High	Medium	Low	High	Medium	Low
2,384	12,737	48	38	14	59	32	9

In an article published in the British Medical Journal in 2004, Keiser et al<sup>4</sup> (**Tables 1 and 2**) conducted a systematic review to evaluate representation of editorial board members and authors in 12 tropical medicine journals referenced by the Institute of Scientific Information (ISI) in 2003. These tropical medicine journals publish manuscripts of research done on diseases that occur mostly in low- and middle-

<sup>3</sup> [http://www.ehow.com/about\\_5127052\\_definition-operations-research.html#ixzz1IYh3Nvvd](http://www.ehow.com/about_5127052_definition-operations-research.html#ixzz1IYh3Nvvd)

<sup>4</sup> Keiser J et al, BMJ 2004;328:1229-32

income countries. Editorial board members and authors were classified according to the human development index (HDI) of their country. The authors found that only 5.1% of the editorial board members were affiliated with countries with a low human development index. Of the 2,384 full articles published in the 6 highest ranking tropical medicine journals, only 13.7% of authors were affiliated with countries with a low HDI.

Apart from publications, investigators should also ensure that health care workers in low- and middle-income countries are empowered to: conduct research; become authors; learn how to write; and, publish articles in peer review journals. Therefore mechanisms for ensuring the sustainability of research and the quality of data should be put in place.

The new knowledge created by research should thus lead to and be measured by:

- a) Publications
- b) Change in management
- c) Change in policy
- d) Empowering of individuals in LICs including:
  - i) Researchers
  - ii) Community members
  - iii) Government/Department of Health workers

## 2.2 Overview of operational research

### 2.2.1 Operational research in Industry

As stated in section 2.1, Operational research (OR), also called decision science, involves the application of “analytical methods to help organizations make better decisions when solving complex problems”<sup>5</sup>.

Examples of operational research in industry<sup>6</sup>:

- designing the layout of a factory for efficient flow of materials
- constructing a telecommunications network at low cost while still guaranteeing quality service if particular connections become very busy or get damaged
- determining the routes of school buses so that as few buses are needed as possible
- managing the flow of raw materials and products in a supply chain based on uncertain demand for the finished products

<sup>5</sup> [http://www.ehow.com/about\\_5127052\\_definition-operations-research.html#ixzz1IYh3Nvvd](http://www.ehow.com/about_5127052_definition-operations-research.html#ixzz1IYh3Nvvd)

<sup>6</sup> [http://www.wordiq.com/definition/Operations\\_research](http://www.wordiq.com/definition/Operations_research)

### 2.2.2 Operational Research in Health

Examples of operational research pertaining to health services:

- documentation of initial defaulters in clinics which led to new systems for management of sputum samples and implementation of research results
- documentation of low percentage of people receiving HIV test when being screened for TB – new systems developed to enable screening
- assessing and improving quality of health data

Operational research is different from clinical or epidemiological research in that it examines an entire system (in this case the health care system) rather than focusing on an individual or a group of individuals (as in clinical or epidemiological research where patients or specific populations are examined). In addition, operational research has at its core, the goal of improvement of a system (the health care system) through identification of challenges and making recommendations for solutions in the system.

Sometimes it is difficult to start in research and to know what to research. But if we change the way we look at the health services field and develop an enquiring mind, it is then not so difficult to know what to research – we must just identify the challenges and develop a much more disciplined way of thinking.

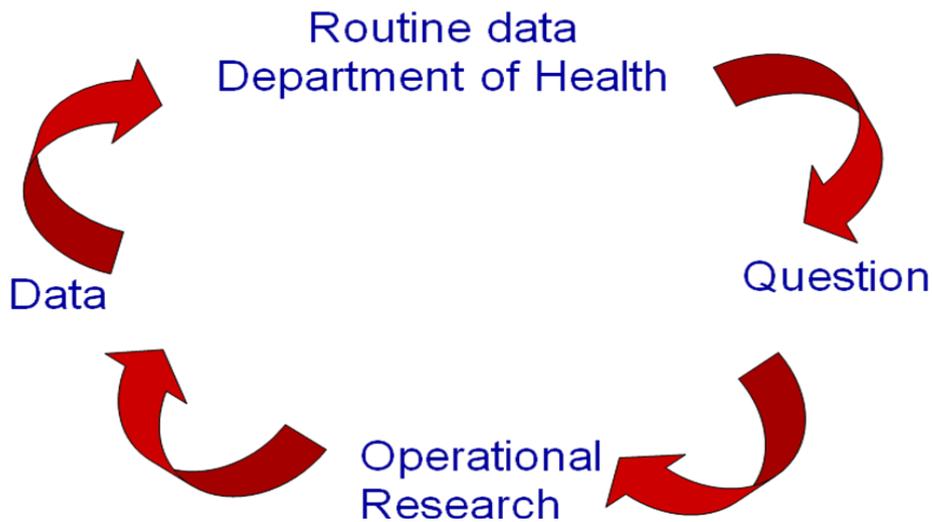
“Research has been called good business, a necessity, a gamble, a game. It is none of these – it’s a state of mind.”

(Martin H Fisher)

To get started in operational research, a challenge in the health system must be identified – this challenge is most often identified using routinely collected data. Of the many challenges that face health care workers every day, most of these can become valid research questions and be studied. The research should focus on and address challenges that are relevant to the locale. The most appropriate people to identify real and relevant challenges are those implementing activities in the health care services arena. It is advisable to think about the challenges and analyse whether, once the operational research has been completed, the challenge will be solvable with the new knowledge created by such study.

Operational research thus usually starts with a challenge identified through routinely collected data which leads to the formulation of a scientific question; in turn, this leads to the collection of appropriate data for analysis and eventually the findings and recommendations are disseminated back to the health services (**Figure 2**).

# What is the process of Operational Research?



**Figure 2. The process of operational research**

Identifying the relevance of a problem is important but can also be difficult. Several questions help to define the need for research focused on a specific challenge identified through routine data capture:

- Does the knowledge already exist to solve the challenge/problem? If the knowledge already exists, then this specific challenge is not a topic for research; rather prompt action based on existing knowledge should be taken to address this challenge. Research should not be an excuse for not acting on existing knowledge.
- Does the research address a problem that represents an action blockade?
- Will it be possible to relieve this blockade and thus improve health services using the new knowledge created by the research?

In order to ensure that research leads to action to address the challenge/problem, it is important to involve all stakeholders from the very beginning. Stakeholders include all those who need the new knowledge that will be provided by the research. Those who undertake the research must work with those who will use the results of the research and with those to whom the services are provided.

Engaging all stakeholders from the beginning implies that they should not be approached only when results are ready to be disseminated. Involvement of stakeholders is best achieved by creating an 'advisory group' which includes researchers, policy makers and individuals from the community. However, it is not always easy to know whether the consultation process should start at the community and then move up to health services (local, then regional and then national) professionals or whether consultation should start at the national level and then move

to regional, local and finally the community level. Whichever direction is selected, it is important to involve all levels at all stages.

Another principle of operational research involves assurance that priority areas are investigated. Determining priorities can be quite challenging. Some considerations include the prevalence of the problem, the risk to vulnerable groups, whether the problem can be solved and whether policy makers and the community are willing to act to solve the problem

Finally it is essential to think through the impact of the research, from the results of the research to what effect the recommendations will have on health services and the health system. Every researcher should spend enough time before the research starts, to think through every possible outcome of the research and the impact of such results. One method is to reflect on the hypothesis and think about potential impact be if the hypothesis is confirmed versus potential impact if the hypothesis is refuted.

*Questions to ask about the impact of the research include:*

- What change do you think this research will bring to the delivery of health services?
- Is the research likely to be equally effective for all people in the community – men, women, adults, children, poor, wealthy?
- What do you hope this research will change for patients?
- What might patients find difficult or untenable about this research and the possible outcome of this research?
- What might this research change for health care workers?
- What concerns might health care workers have about this research and its outcome(s)?
- What changes to the health system might be required (e.g., staffing, training, facilities, equipment, maintenance)?
- Will this research mean any changes to costs for the patient or health care system?
- Do you anticipate any resistance from anyone or any institution to this research?
- From whom will you need support during the project (e.g., technical partners, other institutions, other sectors, political or administrative structures, community groups, etc.)?
- If you conclude that the research is successful, which policy, guidelines, and practice would you propose changing?
- Who would need to validate any policy changes that you propose?
- If there is a policy change, how will the change be rolled out and be translated into practice?
- If there is a policy change, would such a change affect other programmes?
- Who are the designated individuals/focal persons responsible for impact assessment and tracking policy change?
- What impact, if any, is your research likely to have on international guidelines?

### 3. What is needed to get started?

Many academic colleagues consider that OR is not really 'proper' science. They sometimes suggest that the research is not as rigorous as with other types of research, and often think that the results of the research are 'soft' (not really convincing). None of this is true. Operational research, like every other type of research on human subjects must be disciplined, rigorous and precise. OR is not just a narrative describing an event or process but should be 'hypothesis driven' (answer a specific research question). Sloppiness in thinking or carrying out the research is no more acceptable in OR than it is in any other type of research.

#### 3.1 The research question

New knowledge (the intermediate goal of research) starts with a question (problem statement). A clear answer to the research question results in development of the action needed to improve health, the end goal of the research. The research question arises from problems in the routine provision of services within health systems.

Example:

One of the targets of tuberculosis treatment is to obtain results showing sputum to be smear-negative after two months of treatment in a high proportion of patients who were sputum smear-positive patients at diagnosis. This is important because such a result indicates a need to change case management (the patient moves from taking four drugs daily to only two drugs daily). We observe in the report 'Two-month sputum smear conversion' that the proportion of those reported as having converted is lower than the target in some area facilities.

There may be a number of reasons why this may have occurred. Perhaps patients present to the health services very late and consequently their sputum smears take longer to convert. It is also possible that a high proportion of these patients do not have sputum smear examination performed at the end of two months of treatment. Lastly, our patients include a large proportion of cigarette smokers who are more likely to show conversion at a later point in time than non-smokers.

In order to develop a research proposal to determine why this is occurring, we need to select one of the causes that we wish to study and develop a 'research question' about this. For instance:

**Is the sputum smear conversion rate low because the sputum smears are not being examined in a high proportion of patients?**

Posing this question allows us to develop research that can provide an answer and create new knowledge that may lead to action to solve the problem being studied.

This example takes a problem identified within the routine reports from the health information system and explores possible reasons for its occurrence. The final selection of the question focuses, among various possibilities, upon something that can be addressed within the health services system itself. This illustrates the role of OR in addressing problems in the health system that compromise quality or efficiency of services with an emphasis on those things that can be improved or rectified.

In order to conduct the research correctly, it is necessary to transform the question to a hypothesis and then to a 'null' hypothesis. A hypothesis is a positive statement of the content of the question and the null hypothesis is the opposite of this positive statement. This may seem rather unnecessary. However, the reason to do this is in order to construct the research in the way in which it is to be analysed. The statistical tests used to analyse a research hypothesis give a level of confidence within which one can say that two numbers are not the same. For example, a given statistical test gives a result that indicates with a 95% probability that two numbers are not the same.

To illustrate this procedure, we take our previous example and take it through the further steps:

Example:  
*The research question is as follows:*  
**Do facilities have low smear conversion rates low because they have a high proportion of patients with sputum not being examined?**

*Transforming this to a hypothesis, we get:*  
**Facilities with a low sputum smear conversion rate are those in which a high proportion of patients are not having their sputum smear examined after two months of treatment.**

*We take this one step further and convert it to the null hypothesis as follows:*  
**Facilities with a low sputum smear conversion rate are not those with a high proportion of patients who do not undergo sputum smears after two months of treatment.**

This process gives us two elements (variables): sputum smear conversion rates and the proportion of patients having sputum smear examination after two months of treatment.

The outcome is the problem to be addressed, which is a low sputum smear conversion rates. The explanation proposed by the research question/hypothesis is that a high proportion of patients are not having sputum smear examinations after two months of treatment. This is the 'key determinant' in the hypothesis. The relation of these two elements can be illustrated in a two-by-two table as follows (**Figure 3**):

		<u>Facilities with low conversion rate</u>	
		Yes	No
<u>Facilities with high proportion of patients who do not undergo sputum smear after two months of treatment</u>	Yes		
	No		

**Figure 3. Two-by-two table**

This two-by-two table summarizes what the research question is and what the study is about. Moreover, it provides the framework for statistical analysis that addresses

the research question/hypothesis; statistical analysis tells us what confidence we have that the two elements (variables) are unrelated to one another. By convention, if we cannot be at least 95% certain that the two variables are unrelated, we cannot 'reject the null hypothesis' (that is to say, we are at least 95% certain – one chance in 20 – that low conversion rates are not related to high proportion of patients without smear examination after two months of treatment).

Again, by convention, we place the outcome variable (the problem we wish to study) at the top of the table and the explanatory variable (the key determinant) at the left hand side of the table. Finally, we construct the table such that the worst event (low conversion rate and high proportion without examination) are in the upper left hand box and the best event (higher conversion rate and lower proportion without follow-up examination) in the lower right hand box.

Identifying the problem (outcome) and possible explanation (key determinant) and constructing the research question from the outcomes and determinants are the most important steps in putting a research proposal together. Moreover, placing them in the two-by-two table as illustrated clarifies the research question immeasurably and makes the development, implementation, analysis and interpretation of the research much easier.

### **3.2 Proposal outline**

Science is, in many senses, simply disciplined thinking. Discipline requires order and a framework system. This is how research operates - it requires a framework that provides the format and elements of the research proposal. A typical outline (and the one on which the material of the Guide is based) is given below.

Proposal Title

Investigators

Institutions  
Roles

Abstract

Introduction

Background  
Problem statement

Research question

Hypothesis  
Aims and objectives

Study Methods

Study design  
Study area/site  
Population and sample  
Sample size/power  
Sources of data  
Selection of study units  
Variables and definitions  
Measurements  
Instruments

Data Management & Analysis

Data management  
Quality assurance  
Data analysis plan

Ethical considerations

Strengths and limitations

Significance

Impact assessment  
Policy transfer analysis  
Dissemination

Project management

Budget

References

Appendices

When all the elements of a proposal have been completed, the research project can then be carried out in such a way that those for whom the new knowledge has been created can see how it was performed and those who wish to carry out similar research to confirm the findings are able to do so.

### 3.3 Using Routine Data for Research

#### *TB Data Collection Tools*

The National TB Control Programme is fortunate to possess a wealth of standardised information collection tools that produce data that can be used in OR. The list below indicates the data sources, content and reports available through the programme.

1. The “**Case Identification and Follow Up Register**” reports on
  - case detection
  - treatment initiation (started treatment, died before treatment started, initial defaulters)
  - sputum smear conversion timeframes

2. **Clinic card:** The standard structured clinical record used in all TB reporting units contains demographic information on the client, disease classification, treatment regimen, monitoring of treatment and response to treatment, clinical care and treatment outcomes. The clinic card is used as the source document to complete the paper-based TB register but contains more detailed information than is reflected in the register. A summarised version of this information is contained in a patient held card that tracks daily treatment.
3. **Tuberculosis Register:** The paper-based register is used to summarise key information on each registered client (demographics, disease classification, treatment regimen, monitoring and outcomes, HIV testing, CD4+ count, cotrimoxazole and antiretroviral treatment (ART)).
4. **Electronic TB register (ETR.net):** Information from the TB register is collated electronically at sub-district level and forms the basis for analysis of monitoring and evaluation. It contains line data on all registered TB clients on a client name basis. Data is aggregated and sent to provincial and national level health systems staff. The following reports are generated:
  - case findings
  - sputum conversions
  - treatment outcomes
  - HIV reports
  - DOT report
5. **Laboratory request form for Sputum Examination:** Duplicate copies of the form indicate whether it is: a new or retreatment suspect or case; at which point in treatment the specimen is taken; the type of specimen; and the tests requested. This data as well as information on results is available electronically from the National Health Laboratory Services.
6. **Transfer form:** This form reports client information from the register for clients transferred from one district to another.

The data elements available from Case Identification and Follow Up Register and Tuberculosis Register are listed in detail in **Table 3**, as these are the richest sources of information for OR purposes.

**Table 3: TB Data Elements**

Case Identification and Follow Up Register	Tuberculosis Register
<ul style="list-style-type: none"> <li>• date sputum taken</li> <li>• TB suspect number / PHC folder number</li> <li>• specimen code (barcode)</li> <li>• name and surname</li> <li>• age, sex</li> <li>• physical address</li> <li>• date result received</li> <li>• sputum results (pos, neg, grading)</li> <li>• date treatment started</li> <li>• died before treatment started</li> <li>• lost to follow-up</li> </ul>	<ul style="list-style-type: none"> <li>• registration number</li> <li>• registration date</li> <li>• registration type (new, transferred from outside sub-district, moved within SD)</li> <li>• surname, first name</li> <li>• age, sex</li> <li>• physical address</li> <li>• patient category (new or re-treatment after relapse, failure, default or other)</li> <li>• site of disease (ICD10-code) – PTB, E-PTB, Both</li> <li>• treatment start date</li> <li>• treatment regimen</li> <li>• sputum smear date and results - pre-treatment, end of intensive phase (2 or 3-months), end of treatment</li> <li>• TB culture date, results, DST performed, resistance</li> <li>• HIV status, Cotrimoxazole, CD4+, on ART pre-TB, Started ART</li> </ul>

Definition of Terms

The TB Control Programme uses well-established, defined terminology (international definitions in many instances). Before you start the process of data analysis, make sure that your understanding of these definitions is correct.

As a quick test, define the following terms:

- bacteriological coverage
- TB initial default rate (TB primary default rate)
- adult

*Bacteriological coverage* is the number of pulmonary TB cases diagnosed by bacteriological tests (smear and or culture) divided by the total number of pulmonary TB cases reported, excluding children 0–7 years with no smear. When analysing data, people who are not clear about the definition often make the incorrect assumption that low bacteriological coverage is partly due to large numbers of children being diagnosed.

*TB initial (primary) default rate* is the percentage of confirmed TB patients who fail to commence treatment within a specified time. Initial default rates would vary enormously depending on whether the specified period of time was 2 weeks, 1 or 3 months.

There is no standard definition of an *adult*. The South African Children's Act of 2005 defines a child as a person under the age of 18 years (unless married or emancipated by order of court) and an adult any person 18 years of age and older. From a TB Control Programme perspective, however, the World Health Organisation

defines children as being in the 0-14 year age group, whilst in South Africa they are defined as being in the 0-7 year age group.

The failure to use the correct definition can produce a seriously flawed analysis of the issues.

### Identifying the issue to be addressed

Operational research uses routine data to identify issues within health programmes that need to be investigated and analysed using formal research techniques; results can then be used to drive programme improvements. A structured approach to analysis process is essential for successful operational research.

#### 1 Problem Identification

Problem identification starts with the analysis of local data, usually from the electronic TB register (ETR.net).

- Before analysing the data, consider which cases are reflected in the data and which are excluded.
  - The example below is for new smear-positive pulmonary TB cases for 2010. Retreatment smear-positive cases, smear-negative cases, cases with no smears taken and extrapulmonary TB cases are therefore not included.

**Table 4: New Smear Positive Conversion, Sub-district X, 2010**

	Facility	Facility A	Facility B	Facility C	Facility D	Facility E
	<b>Total</b>	40	101	93	62	120
<b>Converted to smear-negative</b>	Number	25	73	65	40	83
	%	63%	72%	70%	65%	69%
<b>Defaulted from treatment</b>	Number	1	1	0	0	2
	%	3%	1%	0%	0%	2%
<b>Died during treatment</b>	Number	2	2	4	3	4
	%	5%	2%	4%	5%	3%
<b>Non-conversion</b>	Number	10	24	21	14	26
	%	25%	24%	23%	23%	22%
<b>Smear results not available</b>	Number	1	0	0	1	2
	%	3%	0%	0%	2%	2%
<b>Transferred</b>	Number	1	1	3	4	3
	%	3%	1%	3%	6%	3%

Sums for proportions in Facilities D and E exceed 100% due to 'rounding'

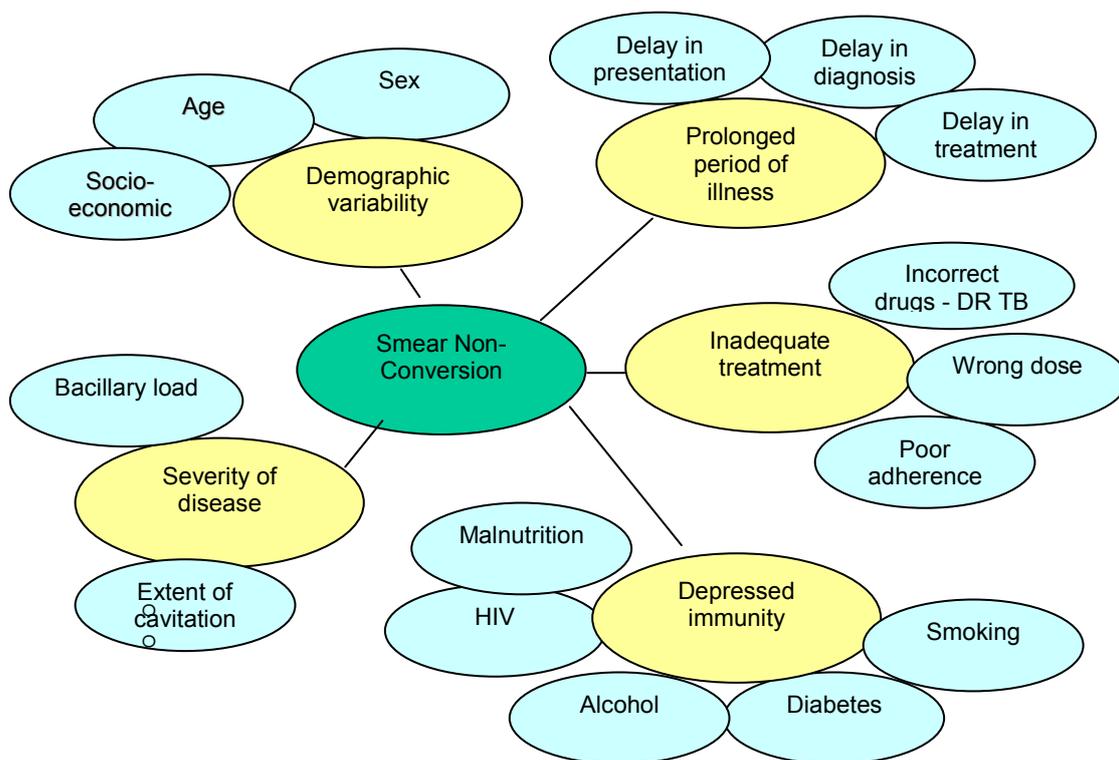
- For new PTB patients, smear conversion is assessed through sputum smear microscopy after 11 weeks of treatment. **Table 4** shows smear conversion data for 6 primary health care facilities in a sub-district.
  - results are reported as follows:
    - number/% with negative smears
    - number/% not taken due to default, death, or transfer out from the facility
    - number/% still smear positive (non-conversion)
    - number/% not available due to smears not being taken or results not being available in ETR.net for patients still on treatment.

- Analysis: What does the data show?
  - smear conversion is low (levels of 80% or more are desirable)
  - the biggest contributing factor is non-conversion (range of 22-25%)
  - death, default and transfers contribute minimally to the problem of low smear conversion.

## 2 Develop an understanding of the problem

Identifying the factors associated with the problem of smear non-conversion is best accomplished through a brainstorming exercise with people who have some understanding of the issues. Thorough consideration of the issues is an extremely important step in helping ensure that your project scope is appropriate and that potential limitations of your research are identified early on.

A creative approach to this is to use Tony Buzan's mind mapping methodology (<http://www.tonybuzan.com/about/mind-mapping/>). Use coloured pieces of paper to note down factors associated with smear non-conversion (colour stimulates right cortical activity). State these factors as single words or short phrases. Group and organise the information according to hierarchy and the association between factors as shown in the example below.



**Figure 4: Mind-map showing possible factors associated with smear non-conversion**

## 3 Benefits of defining the problem area in this way

Defining the problem area in this manner helps to identify what falls within the scope of OR. For example, is this something that is under the control of managers to change? This type of analysis can guide the formulation of the research question (is there an association between smear non-conversion and smoking or between smear non-conversion and bacillary load), as well as help determine the primary and secondary objectives.

Most importantly, it helps to identify some limitations of the research and the proposed intervention in addressing the problem. For example research into the association between smear non-conversion and diabetes with a proposed intervention to provide screening and treatment for those with diabetes, is unlikely to have a significant impact on smear non-conversion due to the relatively small number of cases that are likely to have diabetes. There are many other factors that contribute more significantly to the problem of smear non-conversion.

4 *Why is it important to address this issue?*

Make a case for why you would consider this an important issue to address. In this example, you may want to include aspects such as the public health impact of prolonged infectiousness; the possibility that smear non-conversion at 2 months is associated with poorer treatment outcomes (this could be something to assess as part of the research); the ability to do something to address the situation (early access to HIV testing and care; smoking cessation interventions; improved screening for DR-TB; etc.)

5 *Identify what routine data is available to help answer the research question*

Use the sources of TB data listed as well as your knowledge of additional data sources that may be locally available to identify data elements which will help to answer your research question.

- What is available in electronic records to help answer the question?
  - Data that is already available electronically will reduce the amount of work required to do the research.
- What is routinely collected in clinical records but not collated?
  - This would require capturing data from primary data sources and presents a significant and not-to-be-underestimated challenge.
- What is not available but important to assess?
  - A word of caution here, as the collection of primary data is beyond the scope of many OR projects.

## 4. What are the basic components?

### 4.1 Study design

Scientific research on health-related matters in human communities usually follows one of a limited number of 'study designs'. These have been developed for hypothesis testing primarily in epidemiological research but are equally relevant for health systems/ services (operational) research.

The study designs provide a framework for systematically carrying out the research to address the two essential elements of hypothesis testing – the key determinant and the outcome. This, of course, brings us back to the importance of being precise and to the usefulness of the two-by-two table at the core of the research question/hypothesis.

Within the study, the 'individuals' being studied usually consist of individual units (such as health facilities) within the health system, rather than individual patients or members of the community, as is usually the case in epidemiological research. The focus of operational research is the 'sick (poorly-functioning) facility' rather than the 'sick individual'. As noted above, in selecting the research question, the 'sickness' in the facility is usually a malfunction (low case detection, inadequate sputum smear conversion rate, high defaulter rate, high death rate) that is evident from the routine information reported from the facility/facilities.

The 'individuals' (facilities) within the population to be studied can be classified into four groups (as in the two-by-two table):

- facilities with the outcome of interest (malfunction) and with the determinant
- facilities without the outcome of interest (malfunction) and without the determinant
- facilities with the outcome of interest (malfunction) and with the determinant
- facilities without the outcome of interest (malfunction) and without the determinant

This is illustrated using the example of low sputum conversion rate in the two-by-two table as follows:

		Outcome of interest (low sputum smear conversion rate)	
		Present	Absent
Key Determinant (high rate of non-examination of sputum at two months of treatment)	Present		
	Absent		

**Figure 5. Two-by-two table**

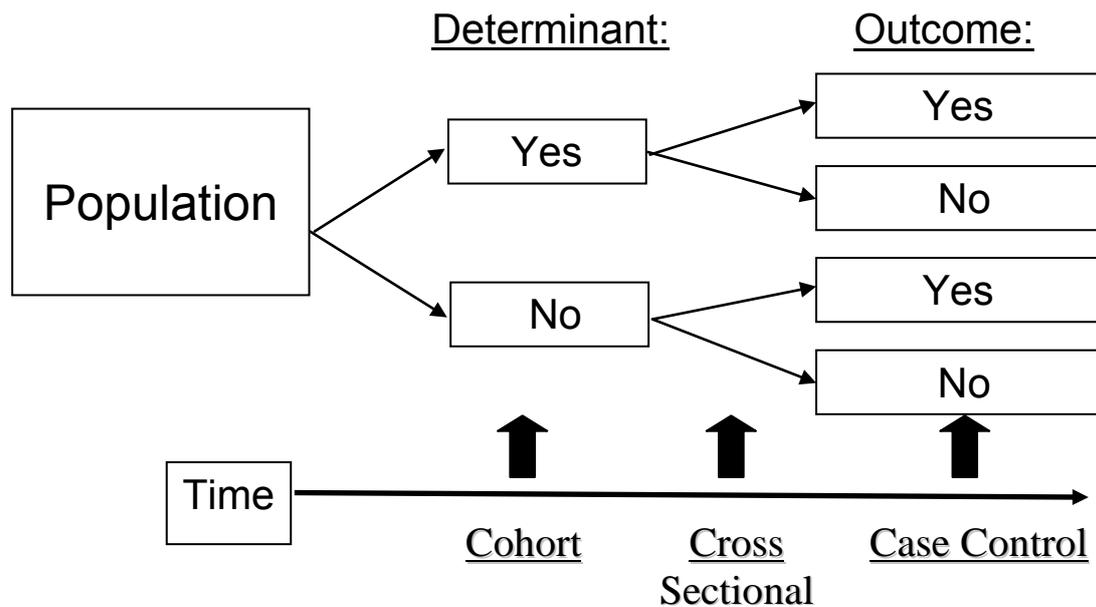
You can construct this table, using the example, as follows. The information about each of the individual facilities is entered into a table with the classification of the key determinant (high rate of non-examination of sputum smears at two months of treatment) and the classification of outcome of interest (low sputum smear conversion rate) as shown in **Figure 5**.

<i>Example:</i>			
Facility	High rate of smears not examined	Low rate of conversion rate	Location in the table
1	Yes	No	Right upper
2	Yes	Yes	Left upper
3	No	No	Right lower
4	No	Yes	Left lower
5	No	Yes	Left lower
6	Yes	Yes	Left upper
etc			

These numbers are added and the sums then entered into the boxes on the two-by-two table.

There are three standard types of study design: cross-sectional, cohort and case control design. There can be a great deal of confusion and discussion around the appropriate study design for a particular study proposal even among highly qualified experts. For the purposes of this exercise in proposal development, it is not really necessary to go into great detail and discussion about this matter but rather to describe exactly what the procedures will be used in the conduct of a study and then to choose one of the three designs and indicate how the study is performed. This can be done relatively simply, even though the theory behind study design may be complex and controversial. These complexities and controversies are outside the scope of this text.

The selection of study design can be made simply from the two-by-two table that summarizes the study. The 'architecture' of the study includes a 'population' that is being studied within which the individual units can be classified (as in the preceding table) by the presence or absence of the key determinant and of the outcome of interest. This is illustrated as follows:



**Figure 6 – Point of Departure for the Study**

This figure illustrates the population, the key determinant and the outcome of interest. On the right side of the figure is the final classification of the population into the four categories as illustrated in the preceding table and summarized in the two-by-two table. The four categories and their position in the two-by-two table, once again, are

- those with the determinant and the outcome of interest (left upper corner)
- those with the determinant and without the outcome (right upper corner)
- those without the determinant and with the outcome (left lower corner)
- those without the determinant and without the outcome (right lower corner).

The figure also incorporates time as a factor. Time is a key component in the ‘chain of causation’. That is to say, if we wish to conclude that something **causes** something else, the cause must have been present prior to the thing that it causes.

In the continuing *example*, the research hypothesis is:

**Facilities with a low sputum smear conversion rate are those in which a high proportion of patients are not having their sputum smear examined following two months of treatment.**

In this statement, we imply that the low sputum smear conversion rate is present because the services are not comprehensively carrying out the sputum smear examination at the end of the initial intensive phase of treatment. In other words, the cause of the low sputum smear conversion rate is the inefficiency of the health care providers in ensuring sputum smear re-examination.

For this to be the true cause, the inefficiency of carrying out sputum smear examination must have happened before the occurrence of the low sputum smear conversion rate was observed, meaning inefficiency in carrying out sputum smear examination **led to** the low sputum smear conversion rate.

Alternative methods for performing this study include:

- purposely introduce measures to improve services (sputum smear examination on all clients following the initial two month treatment). In this approach, we select the facilities where we will introduce the measures to improve the services (the **intervention** study). This is the strongest type of study in providing evidence not only of the cause of the problem but how to solve it. Step-wise change in policy and practice works best public health services and this provides the framework for the stepped-wedge design for an intervention study that allows comparing the outcomes in facilities before and after the intervention and in those facilities with and without the intervention at a given point in time. In order to carry this out scientifically, it is necessary to assign the choice of the time for introducing the intervention in a random order and to include a sufficient number of facilities (a minimum of eleven) and period of time;
- classify the facilities (the population) according to whether or not the site has a high proportion of patients not having sputum smear re-examination (the key determinant in our study) and then search the clinic records to find out whether there is a low sputum smear conversion rate (this is a **cohort** study design);
- classify the facilities according to whether or not they have a low sputum smear conversion rate and then search the clinic records to ascertain whether there is a high proportion of patients without sputum smear re-examination (this is a **case control** study design);
- collect all information about sputum smear conversion and sputum smear re-examination from a single source (for example, an existing database or register) and classify facilities by whether they have the key determinant (high proportion of patients without sputum smear re-examination) and whether the site has the outcome of interest (low sputum smear conversion rate). This is a **cross-sectional** study design.

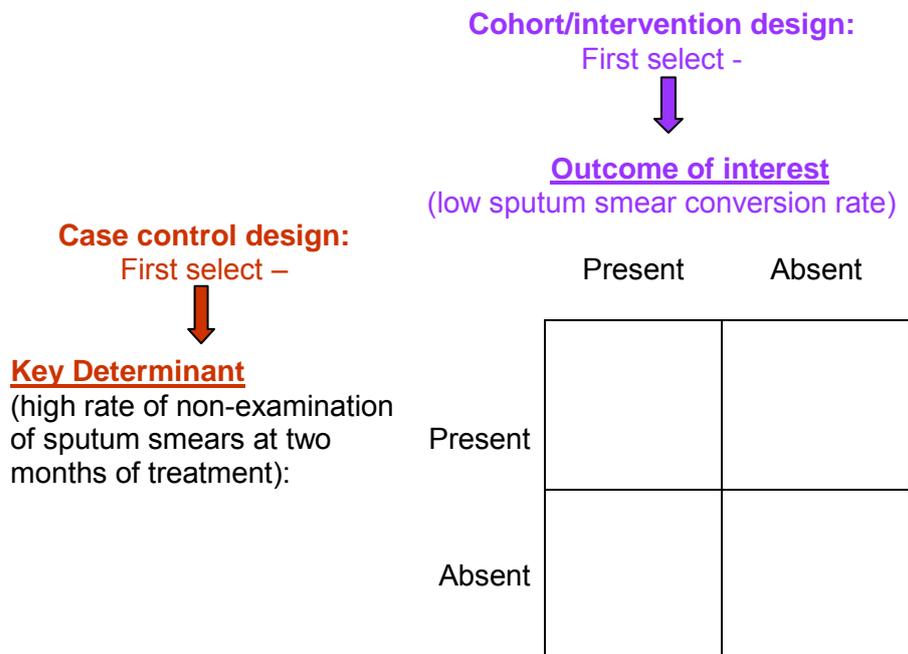
The study design chosen from among the four types of design is determined by which action was taken first. This may involve: 1) collection of information about the key determinant first and then ascertaining the outcome (the **cohort** and **intervention** study designs); 2) collecting information on the outcome and then searching for the key determinant (the **case control** study design); or 3) collecting all information from a single source and then classifying them by determinant and outcome (the **cross-sectional** study design).

The process of deciding which study design to use can be simplified by reference to the two-by-two table you constructed when you first decided on the research question.

In constructing this two-by-two table, it is important to take special care to always put the key determinant on the left hand side of the table (the rows) and the outcome of interest at the top of the table (the columns). Moreover, take care to place in the left upper corner the worst case (the facilities with the outcome of interest which is the 'disease', 'dysfunction' or 'inefficiency' and the key determinant which is the problem that you suspect caused the outcome of interest). Alternatively, you will place the best case (the facilities without the negative outcome or the negative key determinant) in the lower right hand box.

With a two-by-two table constructed in this fashion, you can then decide which action you will take first: choose facilities according to whether or not they have the key

determinant (enter through the 'left' hand of the table or according to whether or not they have the outcome of interest (enter through the 'top' of the table). In the former case, your study will be a **cohort** study design; in the latter, it is the **case control** study design.



**Figure 7 Two-by-two table for cohort/intervention design**

**Cross-sectional design:** collect all information from the same source. Cross-sectional design is used in most operational research studies as typically all the information is obtained from a single source at the same point in time (such as an electronic database or paper register).

The **cross-sectional** design has a number of advantages, the first being efficiency in that the information is gathered from a single source; data is often easily extracted from an existing electronic database. The design uses existing information within the services that likely reflects the actual function of the services, without the effect on routine practice of knowing that a research study is being carried out. Disadvantages of cross-sectional design include the fact that the sequences of events may be difficult or impossible to determine. In addition, one can only study the information that is already present in the records from which you obtain the information.

The **intervention** design provides the most powerful new knowledge. If the stepped-wedge approach is used, it can be undertaken in parallel with the scaling up of any new policy. Disadvantages include the facts that intervention studies are more expensive and require sufficient expertise to ensure proper conduct of the study. It is, however, the preferred design for evaluating any intervention (in our example, strengthening capacity through training or through increasing numbers of personnel).

The **cohort** design has a number of advantages. For instance, the sequence of events can be accurately determined and incidence can then be calculated. A number of determinants can be studied simultaneously allowing an evaluation of complex environments (such as health services) in which a number of factors may lead to a specific outcome. Careful and standardized measurements can be applied to improve the quality of the information obtained. Disadvantages include the fact that often, a large population must be studied. This is particularly the case when the

outcome is uncommon. Such studies often require a study timeframe and consequently are expensive. Due to the prolonged timeframe, there may be losses of participants within the study (for example, if followed up prospectively, some facilities may be merged, closed, or burnt down). This design is used infrequently used in operational research.

The **case control** studies are usually much cheaper and easier to undertake than cohort design. It is often relatively easy to identify the presence or absence of the outcome of interest and to take a sample of those with and without the outcome. This is the only practical method for the study of rare events. This design also has disadvantages. Case control design cannot be used to study sequences of events and therefore cannot conclude whether or not a determinant is a cause, nor can incidence be determined. Case control study becomes difficult to analyse if the methods of assessing an outcome of interest are not standardized. This design may call on participants to recall events or conditions in the past possibly introducing bias. Bias may also occur if cases and controls are drawn from the same source.

Example:

In the continuing example, (study of low sputum smear conversion rate), study design will be considered as follows:

**If we find all the information in a single database (for example, an electronic register):**

We gather all the information we need for our study from the database and create a research database containing only this information. Within this database, we classify each of the facilities by sputum examination rate (our key determinant, either high or not high) and then classify the conversion rate (our outcome, either low or not low). We then post put the numbers of facilities into the table according to the four categories of classification. This is a **cross-sectional** study design.

**If we find only the sputum smear conversion rate in the database:**

We gather this data and enter into the database, as above. We then must find the information we need on sputum examination rates for each of the facilities. If this exists in another database (for example, the laboratory database), we can obtain the information from this source. If it is not in a database, we may have to review the files at each site to collect this information. This (first collect the information on the outcome and then collect the information on the key determinant) is a **case control** study design.

## 4.2 Population

We have already encountered the term 'population' in this text. A population, for epidemiological research purposes, is defined as "all the inhabitants of a given area considered together". In this definition, 'inhabitants' are the 'units of observation' in scientific terms. Unlike epidemiological research where the units of observation are individual people who are either residents or patients, in health systems / services (operational) research, the units of observation are health facilities. For our purposes, the 'sick' health facility is being studied rather than the 'sick' individual. Thus, for OR purposes, the definition of population is "all facilities of a given area considered together in one study". The population, as we have previously noted, consists of facilities with and without the adverse outcome (low rate of sputum smear

re-examination) we wish to study and with and without the key determinant (low smear conversion rate), thus fitting the previously described two-by-two table.

In undertaking research, we propose a research question that aims to discover 'truth in the universe'; however, within our OR study, we are only able to uncover 'truth in the study'. If we conduct our study carefully with due attention to scientific principles, the 'Truth in the study' should reliably reflect r 'Truth in the universe' so that we not only discover new knowledge needed improve the functioning of our own health services; this new knowledge should be sufficiently reliable so it provides a basis for decision-making by others in various locations who may addressing similar health services problems.

As we move from 'truth in the universe' to 'truth in our study', we go through three steps:

- We define the **target population**, which in OR is usually (all) health services / facilities. This is the population for which we hope the results of our study will have relevance;
- In order to carry out the study, we must then define an **accessible population**. It is impossible to study all health facilities in the universe so, normally; we choose to study those that are accessible to us to study (for example, in our country or in our province or, for example, in the public sector). The accessible population should be broadly representative of the target population and is defined by the period selected for the study and the geographical location where the study is to be undertaken;
- Finally, we identify the **study population**. This is the sample of the accessible population that is actually part of the study. This sample must be, as much as possible, representative of the accessible population; for this reason, a standard sampling strategy needs to be defined and used. Finally, the selected sample must be of sufficient size to address the problem but not so large as to make the study too costly and difficult to conduct.

In defining populations, great care must be taken to carefully identify the target population, taking into account the action to undertake and the location where the action needs to be implemented to address the problem. If our aim is to improve health services for the poor and vulnerable (as is usually the case when we study tuberculosis), the 'population' usually consists of those services where the poor and vulnerable seek care. In many locations, these are the public health services (as opposed to the private health services). This naturally varies by location so the choice of target population must be made taking into account the local situation. When the target population is selected, it is crucial to ensure, as much as possible, that the population studied is representative of this target population.

In moving from the target population to the accessible population, we need to ensure, as far as possible, that the accessible population is representative of the target population. What does this mean?

Example:

In our example, the **target population** might be all public health facilities that provide case management for tuberculosis.

The **accessible population** in this example will probably be either all public health facilities in the country or specific province.

The **study population** will consist of the determined number of facilities necessary to answer the study question. This group should be selected from the accessible population in such a way that is representative of the accessible population.

If the study is conducted in one province and someone reads about it in another, the reader needs to understand to what extent the population in province studied is similar to or different from that of the reader. Consequently, when describing the accessible population (the public health facilities in our province), enough information must be provided such that the reader in another location can understand the setting. For example, the following should be described:

- structure of the health system (i.e., primarily public, private, a mix)
- type of clientele served by the system (i.e., mainly men, women, or children)
- nature of services (i.e., free-of-charge or fees paid by clients)
- distribution of the services (i.e., widely distributed or centralized)

A section of the proposal must describe the setting and the nature of the 'population' (health facilities) that is being studied and to what extent this is representative of all health services in the community, province, and/or country.

In going through this process, it is very important to understand how the situation changes as we move from target to accessible and finally to the study population. The ability to approach 'truth in the universe' from 'truth in the study' is highly dependent on ensuring that the target and study populations are truly comparable.

Example:

The **target population** consists of all health facilities that provide case management for tuberculosis.

The **accessible population** would be all public health facilities.

The **study population** consists of facilities that must be studied in order to answer the question.

In this example, the accessible population may differ substantially from the target population if a high proportion of patients seek care in the private sector; it is important to consider such population characteristics to ensure that study conclusions are correct.

### 4.3 Selecting a Sample

Occasionally, it is possible to study the entire accessible population (for example, when there is a national database containing the information we wish to use for the study). However, even when this is the case, typically the information is available or standardized only for a limited period of time. Occasionally it is available for only a part of the population (for example, certain provinces), or is incomplete (for example

some vulnerable groups such as prisoners are not included). In most cases it is wise (for the sake of efficiency) and perhaps necessary to select a sample from the accessible population, which in turn becomes the **study population**.

A sample is drawn from what we call a 'sampling frame'. That is to say, the study population is drawn from a list of all possible study subjects (in this case, facilities or services). The original list is the sampling frame. Ordinarily, this will be drawn from an official register of facilities in a government office, from a list of licensed facilities in a regulatory office or from a 'census' (a process of counting all facilities in a defined area) that is performed by the researchers. It is essential to describe the sampling frame in detail as part of the study proposal.

The process of sample selection performed in such a way to maximize the possibility that it is truly representative of (is similar in every possible way to) the accessible population. Thus, one must ensure that every possible study subject (facility) on the list has an equal opportunity to be selected from the list. This usually entails use of a random selection process, in which the subjects on the list are numbered and a set of random numbers are selected (drawn from a hat, taken from a list of random numbers). The subjects corresponding to the random numbers then become the study population. Other possibilities for maximizing representativeness include systematic sampling and cluster sampling.

The number of 'units of observation' (study subjects) to be selected is determined through a process called **sample size (or power) calculation**.

Sample size calculation is undertaken when there are a limitless number of possible subjects that can be selected for study. In these cases, it is necessary to decide in advance on some assumptions. This can be a challenging task for a novice investigator. Frequently, new researchers pose the question 'if I know the answer to these assumptions, why do I need to do the study?'

Example:

Continuing with our study of low conversion rate, a case control study, we classify the facilities by conversion rate, based on previously reported health information. We then divide the data into those with a low conversion rate and those without a low conversion rate. Next, we obtain the proportions of patients having sputum smear examination at the end of the initial intensive phase of treatment.

We go to [www.openepi.com](http://www.openepi.com) and open sample size for unmatched CC (case control). We choose two 'controls' for each 'case', assuming that for 'control' facilities (those without low conversion rates), seven per cent will be classified as having a high proportion of sputum smears not examined at the end of the initial intensive phase (this figure was obtained from routine reports of supervisor visits to facilities).

An odds ratio of 10 is chosen for the size of effect we wish to detect and to then calculate the sample size needed. The resultant sample size of 15 clinics with low conversion rates and 30 in which conversion rates are not low is needed to detect a significant difference with an odds ratio of 10. We can repeat the exercise using various odds ratio levels.

The first required pieces of information are estimates of the frequency of the key determinant and the outcome in the population that is to be studied. More

specifically, the frequency in the group without the outcome of interest that is needed. Usually the precise level is not known. In most cases, the overall frequency of at least one of the variables is known and the other must be estimated. For estimations, one of the first things to do is to review other studies that have looked at the problem and use the frequencies that have been reported in those studies as the base from which to select your own estimates. Where there are no previous studies and no real indications of what the frequency might be, it is advisable to conduct a small pilot study to obtain this figure in, perhaps, one location; then one can use this figure for the sample size calculation. It is important that the results obtained in the pilot study not be included in the final results of the study.

The second piece of required information for sample size calculation is the size of difference (odds ratio, rate ratio, hazard ratio) that you wish to detect. The best approach is to consider what level of problem would justify the expenditure and effort to fix it. This is a decision that is best taken in consultation with health service managers as it is they who must take the action to correct the problem and they also know what other problems exist; therefore, they are able to set priorities when it comes to dealing with problems in health services. When these two pieces of information are known, it is possible to undertake a sample size calculation as in the example above.

When the possible number of study sites is fixed (there are only so many sites in your area or you have limited resources to undertake the study), instead of a sample size calculation, you need to do a **power calculation**. This calculation tells you, given a certain size of the population you are able to study, what size of a difference you will be able to detect that is statistically significant.

#### 4.4 Variables and Definitions

Variables are the pieces of information that we collect in our research study in order to address the research question. The definition of a variable is 'an element, feature or factor that is liable to vary or change'. The variables in a research proposal refer to all the terms listed in the proposal that refer to any information to be collected.

*Example:*

In the continuing example study of low conversion rate, we propose to study health facilities in Cape Town and to collect information on conversion rate, and on re-examination rate. In addition, we may propose to collect information on facility case loads, health facility locations and types as well as case types and treatment outcomes for TB patients at the facility. The list of variables therefore includes:

- Cape Town
- health facility
- conversion rate
- sputum smear examination
- case load
- location
- type of facility
- case type
- treatment outcome

There are generally two types of variables. **Discrete variables** are those that include values that are categorical or unique occurrences. The values of such variables do not overlap in any way. Thus, an individual is either male or female; there is no 'intermediate' category. Such variables can often be recorded as 'yes' or

'no'. Such variables often reflect 'states' or 'diseases'. For example, one is either dead or alive - there are only these two possibilities, or are there? What about the person who is in a vegetative state with no cognitive function but whose vital functions still operate? It might be possible to classify such a person as either dead or alive and how this is done must be precisely spelled out in the research proposal.

**Continuous variables**, on the other hand, may have a value anywhere along a continuum. So, for example, someone's height may be measured in centimetres. The true height, if it could be measured, would be in centimetres with multiple decimal points. The point is illustrated similarly, and more complicated, with age. The age could be said to be 52 years but is truly 52.546348..... And by the time this number is written, the age has already moved on several decimal points. For this reason, we always 'round' the number at a certain level.

It is important to indicate the type of variable is because the statistical analysis of the information contained in the variable is different for the two types of variables. We will discuss this point further when we come to the section on data analysis.

Information may be collected either as numbers or as text. For efficient management of the information collected, it is the usual practice to transform text into numbers. For example, when we speak of sex, it is either male or female. By convention, males are 'coded' as 0 and females as '1'. This not only provides more efficient handling of the information but also sets up the information in a way that it can be easily managed for statistical analysis.

When all the variables have been identified and listed, they must be defined.

Describing the variables includes specifying the:

- type of variables (continuous or discrete)
- category of variable (number or text)
- range of possibilities for the values of the variable
- definition (providing the precise meaning of the term as it is used in the study).

It might seem obvious that the name of the variable is, in fact, its definition, but this is not the case. Let us take, for example, the simple term 'tuberculosis'. What is its precise meaning in the context of the study? It might mean any of the following (as used in various scientific publications) patients:

- treated with more than one medication
- bacteriologically confirmed
- sputum smear positive
- reported as having died from TB

Each of these definitions can be found in one or other scientific publication and each is technically acceptable. However, they have quite different connotations. For example, if the study includes only those patients bacteriologically confirmed, it will, by and large, exclude small children. If the study includes only sputum smear positive patients, it will exclude all those with extra pulmonary disease and a substantial proportion of those with sputum smear negative pulmonary tuberculosis (a selective exclusion of patients living with HIV when not on treatment for it). If we include all patients given more than one medication for treatment of the disease, this will exclude all those diagnosed but never treated ('initial defaulters') and include a certain number of patients with serious conditions affecting the lung but who are incorrectly treated for tuberculosis (more frequent in patients living with HIV). The precise definition of the terms used in scientific research is crucial to the quality of the research being undertaken. Each of the listed variables must be defined precisely to

enable others to understand exactly what the study refers to and to be able to replicate the study if they so desire.

#### 4.5 Measurement

For operational research as for all forms of research on human subjects, the information from which the new knowledge is created comes from measurements. Each piece of information (variable) is drawn from measurements. For each measurement, the definition of the variables collected needs to be precisely recorded. The methods by which the measurement is taken must be described in detail. The criteria for definitions and data collection must be precisely adhered to. Failure to follow this scrupulously will lead to errors and undermine the usefulness of the information in creating new knowledge.

For the type of research described here, the information to be collected may be routinely recorded or it may need to be collected specifically for the study. The data source for routinely collected health system information usually consists of clinical charts, standardized records or routine reports. These may be stored in hard copy or may already be entered into electronic format.

For information that is collected specifically for the study, the data collection may be accomplished using a questionnaire, a standardized data collection form or through physical tests.

Example:

Research on tuberculosis has a number of advantages. Definitions used in tuberculosis case management are highly standardized and internationally comparable;

There are standard forms to record information related to tuberculosis case management that vary little from one country to another. The list of variables therefore includes:

- treatment card
- laboratory register
- treatment register
- quarterly reports

Each of these has standard definitions and accompanying training materials.

Some general principles underlie the development of forms that work well, whether they are forms for routine use in health services or for use in research.

- Most important, the forms must be useful for case management. If the busy health care workers do not see the advantage of using specific forms, they will not be completed well and become useless both for case management and for health information;
- The forms must be clearly laid out so they are easy complete and so that each pieces of information can be easily seen
- Forms must contain the minimum amount of information. This information should be clearly useful for case management as well as for monitoring
- Forms must be piloted before being introduced for general use. Very often, simple errors (missing essential information, unclear layout) are only evident after the form has been pilot-tested;
- Once the forms have been introduced, their use must be carefully monitored to improve accuracy and ensure completeness.

The highest priority in measurements for research is to ensure that measurements are comparable throughout any study. If measurements are more complete/precise in some cases than in others, incorrect wrong conclusions may result. This will be discussed further in the next section. Data collection and recording must be systematic so that the comparisons are made correctly.

First, it is important to precisely define the terms used for the variables. It is best to use terms (and definitions) that have been universally recommended. These can be found in other publications on the subject and often have been published in scientific articles specifying international usages for research or monitoring purposes. Second, it is important to select measurement techniques that are standardized, validated and universally recommended.

In addition, it is important as well for all measurements to be as precise as possible. Finally, it is essential to monitor the collection of the information to ensure that the forms are as complete as possible and that the information is recorded precisely as it appears in the records or from the measurements.

Example:

Tests used for diagnosis of tuberculosis are well described, widely used and standardized. Those used to define the variables in a research study should follow these recommendations:

- sputum smear examinations should be carried out using the Ziehl-Neelsen or fluorescence microscopy method; most important, the precise method used in the study facilities must be specified;
  - sputum cultures use a limited number and specified recommended and standardized methods
  - drug susceptibility testing may be performed via several different standardized methodologies and thus must be specified in the proposal
- Routine testing methods may change over time. If this occurs during the study period, it needs to be specified and it may even be necessary to separately analyse the periods using different methods

There are special considerations to consider for measurements in OR. Operational research usually makes use of routine information recorded in the health system. Additional review of facility records and occasionally interviews with clients, patients or staff are sometimes utilised. Information most useful for OR purposes is that which is routinely collected on a standardized case management form that has been developed as described above. If information is recorded without the use of such a form, it is much less useful for research purposes. Even where a standard case management form is used, the information may vary due to the fact that multiple individuals have entered the information and recording may not be as consistent as would have been desired.

In many instances, routine health information is collated into electronic databases, thus facilitating access to the information. However, the accuracy of such data is much more difficult to verify. One method of verification involves taking a sample of original files (if they are accessible) and determining the rate of transcription errors.

In any case, the methods should include a careful and precise description of the systems and standard operating procedures for data recording and reporting. Whenever there is a change in definitions or recording procedures, these must be described in detail. In addition, the methods should describe any procedures to ensure consistency and comprehensiveness of data recording.

In handling any health facility information from such records, strict procedures are required to ensure confidentiality. This includes ensuring that no personal identifiers are included in any data file nor used for analysis. In some instances, personal identifiers (names, dates of birth, address, etc.) may be required to perform record linkage. If this is necessary, record linkage procedures must be described along with processes to ensure that confidentiality is maintained after linkage.

In the methods section of the proposal, it is necessary to describe the source of the information, how it is accessed, how the information is transcribed to the research file and what procedures are used to assess and assure data quality.

## 5. What are the building blocks?

### 5.1 Data management<sup>7 8 9</sup>

When thinking about data collection and the database used for a particular study, one needs to begin with the end in mind. In other words at the time the proposal is being written, think about how the data will be analysed and therefore in what format the data must be collected. Focus on the research question and the hypothesis to ensure that these can be answered using the chosen data collection methods.

Think about the following:

- What is your hypothesis?
- What exactly do you want to analyse?
- What data form/format is needed for data analysis?

Good data collection and management are needed to ensure that researchers can manage and track data, samples (if samples are collected), follow-up visits if needed, missing results (which always occurs) and the flow of data and samples. Suitable data management ensures the reliability and accuracy of the data.

Data management thus starts during initial proposal development, meaning that the proposal must contain data requirements and data management sections. Both data requirements and data management must be thought through carefully as there are budget implications including needs for the following:

- equipment (e.g., computer, printer)
- computer supplies (e.g., toner, printer cartridges)
- software or computer programs
- data collection tools whether paper based or electronic (e.g., personal digital assistants (PDAs))
- stationery (e.g., paper (various colours), envelopes, labels, barcodes (for sample storage), pens, glue, tape, clipboards, scissors, files)
- staff (data manager/developer, data clerks)
- data storage and backups (hardware/software, locked cabinets)

When thinking about which database to use, one should obtain advice from an expert as there are many different types of databases. The principle is to use the most robust but also the most user-friendly and least complicated database. Flat file databases like EpiData and EpiInfo are fine when the database will be fairly small and when a one-to-one relationship of data is required. Often the questions asked in OR are not overly complex and a flat file database will be sufficient. Microsoft Excel is a spreadsheet rather than a database and is not appropriate for research purposes. However sometimes this is the only software available and will be used for small studies or data export from e.g., an electronic TB register. When using Microsoft Excel, the researchers should be extremely careful to ensure that columns or rows do not move. When more complex data are used and a one-to-many relationship is required, a relational database like Microsoft Access, Microsoft SQL Server or Oracle must be used. These relational databases are complex and their use will typically require expert assistance.

---

<sup>7</sup> Hulley SB, Cummings SR, Browner WS, Grady D, Hearst N, Neuman TB. Designing Clinical Research, 2<sup>nd</sup> Edition

<sup>8</sup> Enarson DA, Kennedy SM, Miller DL, Bakke P. Research Methods for Promotion of Lung Health. Paris: The Union 2001.

<sup>9</sup> EpiData and EpiAnalysis, Jens Lauritsen

If electronic data (for example an electronic TB register) is used, it is often not necessary to set up a separate database – the specific electronic data needed for the study can be exported directly into a statistics programme for analysis.

Whichever database/spreadsheet is used, always ensure that a final database (once complete and checked for accuracy and completeness) is locked and stored in a separate folder and that copies are used to perform calculations and analysis. Never use the final database as the working document– if a mistake is made and the final database/spreadsheet becomes corrupt, the whole project will be in jeopardy.

Important principles:

- lock final database
- never use final locked database/spreadsheet for calculations/analysis
- store original locked database in separate folder
- always make a copy of locked database to us for calculations/analysis

Data management includes the entire process of collecting, capturing, storing and preparing the data for analysis. All data should be handled and managed according to Good Clinical Practice (GCP) requirements and ethical standards. For good data management the following are needed:

- carefully planned data forms (e.g., case report forms (CRFs) or questionnaires)
- even if routinely collected data are used (e.g., TB register data) a form should be set up to clearly indicate which variables from the routine data will be used
- data and sample flow algorithms and logistics
- data management plan
- data dictionary
- standard operating procedures (SOP) for collecting and storing data

A data dictionary is essential element for data documentation, as it forces the researcher to think logically about data structure and format. The data dictionary should contain at least the following for each data variable:

- variable name
- variable description
- format/type (character, integer, date or memo)
- length
- value/format/range (permitted values)
- logic checks (e.g. root vs. nested question, sex may not be unknown)
- missing values (e.g., 0=No; 1=Yes; -5=Unknown) – make sure that the symbol used for unknown value is not a value that can occur in the actual data for that variable

Following is an example of a data dictionary (**Table 5**) with all variables correct, but sex (male, female) as characters”

<b>Variable name</b>	<b>Variable description</b>	<b>Type</b>	<b>Length</b>	<b>Values / format / range</b>	<b>Logic checks</b>
PID	unique identifier	integer	10	may not be null	no duplicates
Q01_INTDT	date of interview	date (D)	10	<=today (if not captured real time) unk=01/01/1800	if unknown – raise query – return to field
Q01_DOB	date of birth	date (D)	10	<=today...; unk=01/01/1800	if unknown – enter age
Q01_Age	age	int (I)	3	>0; (if working only with adults>=15)	if DOB entered – default value=999
Q03_Sex	sex	char	1	M=male; F=female	may not be null

It is best to enter data as numerical values (integers). For example decide and record in data dictionary that male will be 0, female will be 1; rather than characters (male = M and female = F).

<i>Example of data dictionary with sex (male and female) as integers rather than as characters</i>					
Q03_Sex	sex	int (I)	1	0=male; 1=female	formatted for analysis

The standard operating procedures (SOPs) for data management must include step-by-step instructions on how data are to be collected in the field. SOPs also include details of how to handle missing data or incorrect entries from routinely collected data. The SOP must clearly state that one of the most important aspects of data management is consistency – data must always be collected and captured using exactly the same methods. Instructions in the SOP must be explicit so that anyone following these instructions is able to repeat the study – in other words the study and the data collection and management should be reproducible by other groups.

It is essential and an integral part of GCP and ethical principles to always maintain confidentiality and not to use patient names when collecting and analysing data. However, sometimes the only way to access data (e.g. when doing a folder search or matching files) is to use names of clients. The principle is still to use unique study numbers and not to have the name and results in the same document or the same piece of paper or on the same electronic spreadsheet or database. Thus there should never be names of clients on any paper, document or electronic database used for data collection, management or analysis. The only location that should contain a name is on the consent form or on an enumeration form, which may contain at the most demographic data (age and sex). Often the enumeration form is used as a

management form to keep track of progress (e.g., which cases have been interviewed).

Occasionally it will be necessary to use names during data collection – this happens when one collects data from various sources and has to match data to the same individual.

Example:  
 A research study aims to determine the association between drug resistance and infection in HIV:

- information on drug resistance is reported in a standard register;
- information on HIV status is not recorded in the register and must be sought in the clinic folders;
- in order to locate the folders, it is necessary to have the names of the study subjects.

In such a case, the investigators may ask the ethics committee for a waiver of written informed consent to access data from the folders. The exact manner in which data will be accessed will be reported to ethics committee and only when approval has been granted, can the study proceed. One method for accessing data without ever having the name of a client on the same form as the results to be collected, is to develop (or ask a senior data specialist to develop) and compile 3 lists:

1. A list with subject name, surname and unique subject ID
2. Another list with unique subject ID and unique study code (no name, no results). In other words, this list contains two codes but no name or results.
3. A different list with unique study code (no name, no unique subject ID), columns for results.

Example of a list with subject name, surname and unique subject ID

Name/Surname	Unique Subject ID

Example of list with unique subject ID and unique study code (no name, no result).

Unique Subject ID	Unique study code

Example of a list with unique study code (no name, no unique subject ID), and columns for results.

Unique study code	Results

These 3 lists should be maintained and stored separately. This system (or a similar system) is acceptable (if permission is granted by the ethics committee) for data collection from different data sources (e.g. patient folders and registers) and to later link the data together.

Once data collection has started, data entry should start. It is important to perform dual data entry – in other words two people capture the same data separately in identical databases – each person has his/her own copy of the database. The reason

for this is that it is only human to make errors (often about 10%) during the data capture process. However, when 2 people capture the data, the two data sets can be compared and corrected by verifying against the original source document, leading to a dramatically decreased error rate.

Example of validation process using dual entry.

Dataset 1 is captured by data capturer 1 and dataset 2 is captured independently by data capturer 2. For data validation, dataset 1 is compared with dataset 2 and all discrepant answers are listed. In this instance gender is captured in dataset 1 as 1 (male) and in dataset 2 as 2 (female).

Dataset 1		Dataset 2	
Unique ID	1224	Unique ID	1224
Sex	0	Sex	1

The next steps are to check the source document (e.g. patient folder) and mark the correct item on the validation document. The last step is then to establish which dataset (1 or 2) has the least errors and to make all the corrections for the final database on this dataset. After data have been captured, dual entries corrected and validated and all queries resolved, the database should be locked and a copy of the locked database stored safely. There should never be names in any locked database.

Principles for locked database:

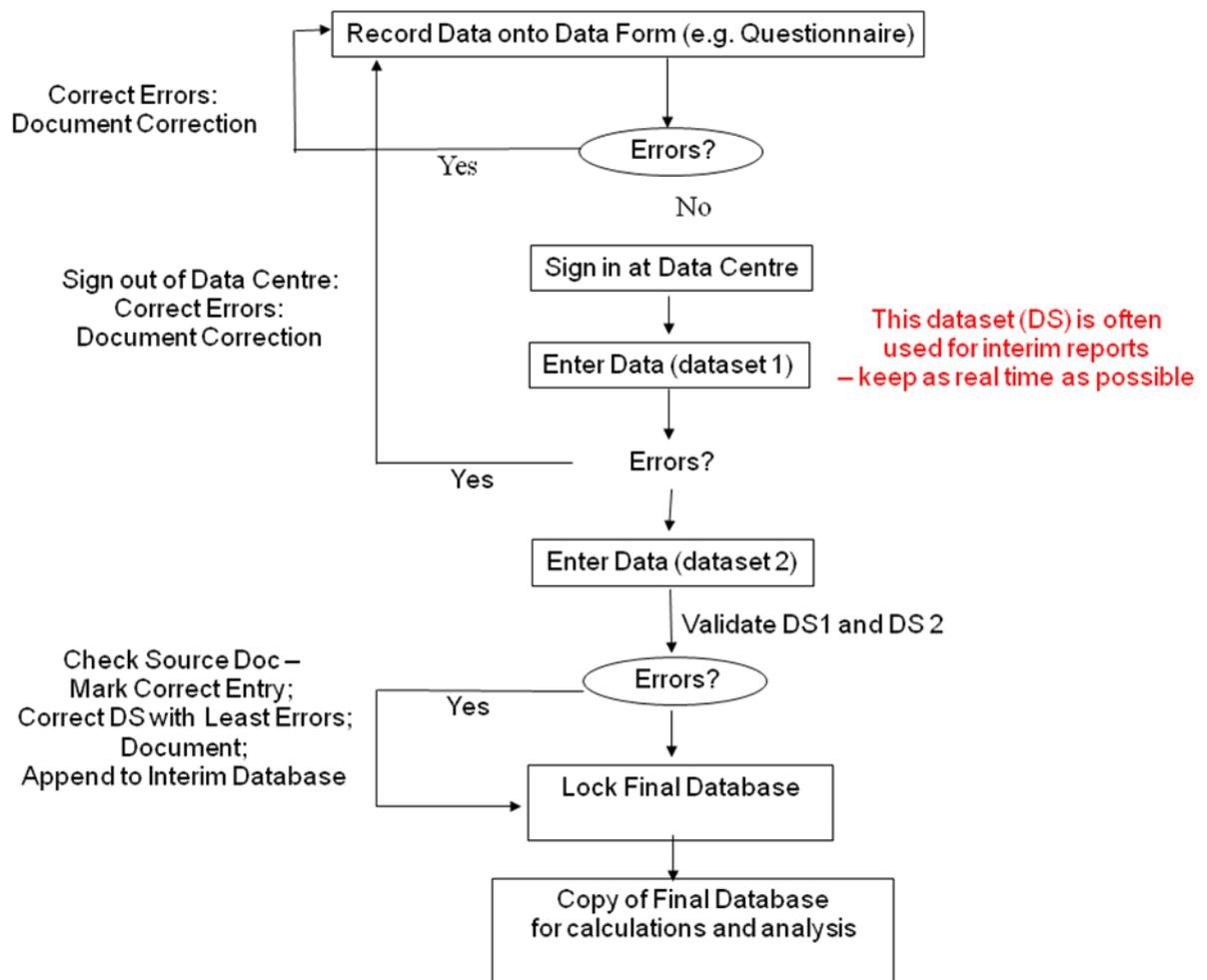
- Locked database should never contain patient names
- Locked database should be stored safely
- Original locked database should never be used for calculations/analysis – always make a working copy

Data storage integrity is essential; all data must be stored for a minimum of 5 years (some studies require data to be stored for up to 15 years.). Data can be stored either in electronic format, or in hard copies, but preferably in both.

For electronic databases regular backups are essential. A suggested schedule consists of the following backups:

- Daily– keep the most current backup off-site
- Weekly Backups (keep for at least a month)
- Monthly Backups (keep for 6 months)
- Quarterly Backups (keep for year)
- 6 Monthly backups (keep for 5 years)

All paper documents must be kept for a minimum of 5 years – some studies require all data to be kept for 15 years. For this long term storage one needs a safe locked facility that is preferably safe from natural disasters like floods, fire and other destructive elements for example rats and moths. As a further precautionary mechanism against water from burst water pipes, never store data directly on the floor, always store it on shelves. Data should be stored in a logical format, e.g., by community or by date or by unique identifier. Consent forms and linking lists that contain names should always be stored in a separate locked filing cabinet.



**Figure 8. Data Storage, Sketch of logical and easy data flow and management of data** (by kind permission of Ms Kathy Lawrence)

## 5.2 Comparison (analysis plan)

To answer a research question and to address a hypothesis, it is necessary to make comparisons. This is what differentiates research from routine reports.

### Example:

For the research hypothesis we state that:

**Facilities with a low sputum smear conversion rate are those in which a high proportion of patients do not undergo sputum smear re-examination.**

Therefore:

- this hypothesis implies an outcome 'smear conversion rate'
- it also identifies a determinant of interest 'sputum smear re-examination'
- for the outcome, there two classifications – 'low' and 'not low' rates;
- for the determinant, we also have two classifications – 'high' and 'not high' rates.

This formulation allows us to prepare our two-by-two table that places the group having both the outcome of interest and the key determinant in the left upper corner of the table (the facilities with low conversion rate and high proportion not having examination). Once the table is constructed, the comparison is easily made.

In making the comparison of the four categories, we must take account of several issues:

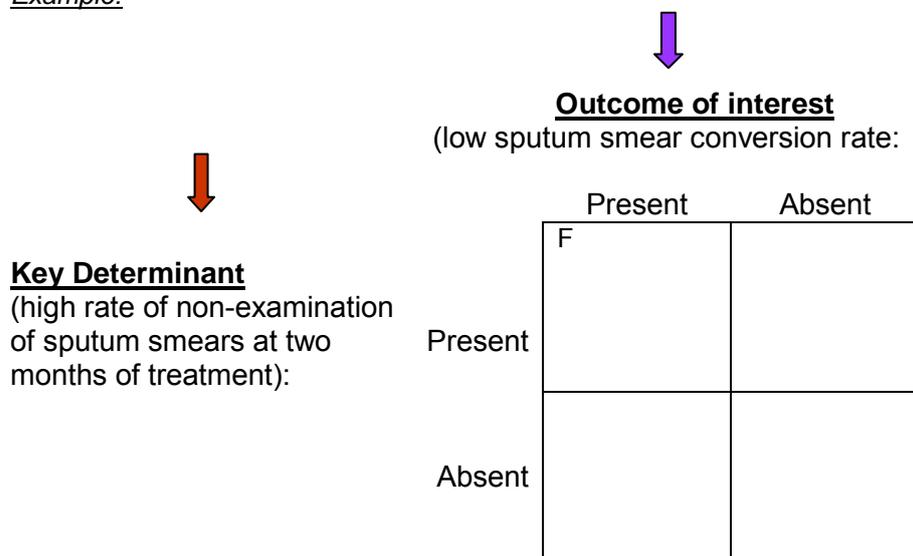
- the type of data;
- the size of the difference among the groups;
- errors in measurement of the variables;
- estimates of probability.

As noted above, there are two main types of data – categorical and continuous. The usual method of showing and analysing categorical variables is to prepare a contingency table (in its simplest form, the two-by-two table) and then performing a chi-square analysis. Most programmes that make this calculation automatically provide an odds ratio and 95% confidence interval. When the confidence interval does not include 1.0, we have 95% confidence that the two values are not from the same underlying group. In this case, with 95% confidence, we can 'reject the null hypothesis' that there is no difference between the two groups (they are not two measurements that came from the same group).

For analysis of continuous variables, the usual method of comparison is either regression analysis (if both the determinant and the outcome are continuous variables) or variance analysis (if the determinant is categorical).

In our example, we select categorical as it is the clearest means to illustrate the comparisons made. Even when data are continuous, they may be 'converted' into categorical variables. The usual method by which this is done is to divide the range of distribution of the results into five equal parts, called 'quintiles'. This identifies a 'central' group and two groups to either side and allows easy examination of the relationship between the variables.

Example:



**Figure 9.**

When we performed the measurements and entered all data into a database, we can then count the frequency of occurrence of each of the four possibilities in the two-by-two table (**Figure 9**) and enter the frequency into each of the boxes. We can then calculate the percentage of the numbers for each of the rows (per cent of those with / without the outcome according to the presence/absence of the determinant). This allows us to easily compare the frequencies.

Example:

High non-examination rate	Low conversion rate				Total	
	Yes		No		Number	Per cent
	Number	Per cent	Number	Per cent	Number	Per cent
Yes	36	30%	84	70%	120	100%
No	10	3%	330	97%	340	100%
All	46	10%	414	90%	460	100%

From this table, we see that the 'worst case' is in the upper left box, indicating 30% of those with a high rate of non-examination have a low conversion rate. This compares with only three per cent among those where the rate of non-examination is not high. Without statistical analysis, we immediately see that the difference between the two figures is ten-fold. This is called the 'rate ratio'.

In a case control study(used most frequently when the rates being examined are very low), the correct comparison is the 'odds ratio' rather than a direct comparison of rate ratio. The odds ratio is calculated as follows:

$$\frac{\text{Upper left} \times \text{lower right}}{\text{Upper right} \times \text{lower left}}$$

In our example, the odds ratio is 14.1 which compares with a rate ratio of 10.0. This illustrates the point that the odds ratio approximates the rate ratio only when the event being studied is rare. The two ratios are measurements of the size of the difference in the table. In this example, the rate ratio is a more accurate reflection of the true difference between the two groups.

### 5.3 Error and bias

All measurements have a certain amount of error. This error can be either random or systematic. Random error can obscure a real difference between the groups being compared. If the error is random and not large, the likelihood to find a true difference between groups increases with the size of the population studied.

Systematic error, on the other hand, can lead to bias. This may result in finding of a difference between groups when one does not actually exist. Bias is far and away the more serious problem and careful attention must be taken to minimize its possibility in scientific studies.

Random error is intimately associated with, and the result of, actions taken during measurement. As discussed in the section on measurement, error can result from a number of factors:

- use of inappropriate tools for measurement
- non-standardized measurements
- inadequately trained researcher personnel
- lack of systematic evaluation
- lack of carefully monitored measurements

Systematic error, which can lead to bias, is of several types: selection bias, information bias and confounding.

Selection bias may occur if: an inappropriate population is being studied, there is inadequate participation of the eligible population, the classification of the

determinant changes over the study period; the population consists of the most accessible groups or volunteers.

Example:

In our example, we propose to study the sputum smear conversion rate. If we study all cases, the sputum smear conversion rate will be falsely low; therefore, in order to 'convert' the sputum smear, it must have been positive to begin with. Thus, for this study, we must choose only those who are sputum smear positive. A population made up of 'all cases' will include cases that are sputum smear negative (much more frequent in children) as well as extra pulmonary cases (often more frequent in referral facilities). These cases clearly cannot have a possibility of sputum smear conversion from positive to negative. If we select all cases to study and some units have higher proportions of children or extra pulmonary cases, they will naturally have a lower sputum smear conversion rate than facilities with a high proportion of adults and which treat primarily pulmonary tuberculosis.

The selection of an appropriate population is very important. If we wish to study a determinant, we must select a population that has the possibility of having the determinant.

The effects of participation in creating bias occur when participation in the study is selective in relation to the determinant and/or the outcome. This is why it is essential to report on the total eligible population and to determine what proportion actually participated in the study. If the proportion is very high (for example, over 80%), the possibility that bias may have occurred due to selective participation is much diminished.

Example:

In our continuing example, we again propose to study the sputum smear conversion rate. Sites at which patients are less likely to be adherent to taking smear examination are less likely to provide information. If we miss this group in our study, we will not include those units with a high rate of sputum re-examination. We will then fail to find a difference between the two groups.

A change or variation in definition of outcome can lead to bias. Consequently, it is vital that the definition be the same in all groups under study.

Example:

In one group of units, the definition is 'those not reported as positive'. In another group of units, the definition is 'those reported as negative'. The two groups are inherently different, simply by their varying definition. The possible outcomes of sputum smear examination include negative, positive, not performed and missing. The inclusion of the last two categories in one set of units and their exclusion from the other misclassifies the units in terms of the outcome being studied.

The participation of an easily accessible or volunteer group for participation can also lead to systematic error. Volunteers or easily accessible populations may have more or less of the problem being studied and may also have a different level of the determinant, as compared with the total eligible population.

Example:

A group of units that volunteers for the study where the sputum smear is more likely to be performed tends to have patients who are more likely to be adherent to treatment and, consequently are more likely to show sputum smear conversion. If a group of units is selected for study because of low sputum smear conversion and compared with a group of units that are volunteers, there may be a false difference between the two groups.

Bias is a very serious problem in scientific investigation and every effort must be made to minimize it. Specific care must be taken in study design:

- The population to be studied must be appropriate to the question;
- Every effort must be made to ensure a high participation rate;
- Comparison should always be presented between those who did and did not participate (for example, age, sex, residence).

Efforts to address bias may also be undertaken during analysis:

- Any part of the population excluded from the numerator should also be excluded from the denominator in calculating rates;
- Analysis using 'person time at risk' may be used;
- Estimates can be made of 'worst' and 'best' case scenarios.

Systematic error may also occur due to information bias. Information bias may result from subject variation, observer variation, deficiency of measurement tools or technical errors in measurement.

Subject variation occurs when the same facility has a different outcome from one point of observation to another. For example, the unit may have a marked change in staffing during different times of the year, with associated variation in completeness of examination and/ or recording. The unit may vary in terms of its efficiency with seasons of the year, if the unit serves an agricultural community or is subject to extremes of climate. The unit might also vary in efficiency if it is aware that it is being evaluated.

Example:

A facility, during one period, has a high proportion of multi-drug resistant-(MDR-)TB patients under treatment because of an outbreak. These patients will show slower sputum smear conversion than drug-susceptible patients.

Observer variation may also contribute to information bias. This variation may occur between several observers (inter-observer error) or in the same observer at different points in time (intra-observer error).

Example:

One observer visits the facility at lunch hour so is unable to find the laboratory register while the other visits mid-morning, so is able to record the results of microscopy. The same observer visits the facility at various times, so records different results for each visit.

Technical areas of measurement may also lead to information bias. This is more typical of biological measurements but may also occur in OR. Unless measurements are carefully standardized, conscientiously undertaken and systematically recorded, errors may occur.

Example:

One research technician searches the laboratory register, the treatment register and the patient treatment card while another searches only the treatment register. The estimates of sputum smear conversion rate may differ greatly between the two technicians (or between two studies).

In order to minimize information bias, it is critical to specify criteria and procedures in advance, to analyse according to the pre-set criteria (and not based on a *post hoc* classification); to reduce the number of observers, to monitor the performance of the observers and to use standardized tools for measurement.

A special kind of bias is due to confounding. A 'confounder' is a factor that is associated with both the determinant and the outcome and consequently leads to a false association between determinant and outcome.

*Example:*

For instance, we may observe an association between smear conversion rate and whether or not the sputum was re-examined. However, those facilities that do not have access to sputum smear microscopy may well be the facilities where the sputum is not examined and are the cause of the low smear conversion rate. The real explanation in this case for low smear conversion rate (and the factor on which action is required) is access to a smear microscopy laboratory rather than negligence on the part of the facility staff in obtaining the examination.

For a factor to act as a confounder, it must be independently associated with both the determinant and the outcome. In investigating the possibility of confounding, it is important first of all to test for an association with the determinant and the outcome. If the factor meets the criteria for confounding, an analysis of the association between determinant and outcome needs to be undertaken, stratifying for the presence or absence of the confounder. If the association between determinant and outcome persists after stratification for the potential confounder, the association can be accepted and the possibility of confounding rejected.

## 6. What is the justification?

### 6.1 Literature review and referencing

What is a literature review? It is a systematic and thorough search of the literature in order to identify as many relevant items as possible related to the subject being studied. We should do a literature review not only at the start of the study, i.e. at a fixed point in time, but continuously throughout the proposal writing process and the study itself, and during data analysis and writing up the manuscript. Each time we do a literature review, we look for different information. So we need to think about the different stages/time points of the study and adjust our literature review strategy accordingly. This is of course not to say that we will not use the previous strategy, which may include keywords, specific authors or geographical settings, or MeSH terms (**M**edical **S**ubject **H**eadings), which we will address later in this document.

So before you start, you need to have a clear idea about what you are searching for and what the extent of the search will be. Remember, there is an extraordinary amount of material available! Therefore, to start with, you need a *search question*. This question may be the same as your research question, but sometimes it differs in the sense that the background information you need in order to formulate your research question may include a broader scope. For the purpose of this section, we will use the following question as an example: *Is isoniazid prophylactic therapy (IPT) effective for the prevention of tuberculosis in children?*

There are different kinds of literature which you can search and before you start with the search, you need to think about these different kinds and which will be most applicable to your question. The different kinds of literature include: published literature, grey literature and unpublished literature. You also need to define the period of time of your search, i.e. the dates when publications became available. Another key issue is the language(s) which you are going to include in your search. You have to think carefully about the implications of omitting some languages from the search and remember that although some articles are not written in English (or a language you are familiar with), the abstract may be. If the abstract seems important for your literature review, think about having the article translated.

When looking at published data, you have to think about searching databases, doing manual searches or contacting researchers in the field who may be involved with the type of research you are interested in. There are different types of literature databases<sup>10</sup>, i.e. databases are organised differently and depending on your search, you may choose either the one or the other type. These include:

- a database organised using a structured thesaurus with MeSH terms (e.g. Medline), where a keyword, abstract or specific author is used to identify an article of interest according to the category (or MeSH) it is ordered to; this type of database may however not include the latest concepts, especially in a rapidly developing science such as medicine.
- a database without a in-built structure (e.g. Google Scholar) - and greater thought and preparation is needed to identify keywords which would incorporate all possible references.
- a citation database (e.g. Scisearch), where articles which have been cited in the field are identified.

Which search strategy you are going to use depends on the resources you have available and on how familiar you are with your search topic. Published data could

---

<sup>10</sup> Eyers, John. How to do (or not to do)... Searching bibliographic databases effectively. Health Policy and Planning; 13(3): 339-342

also be searched manually by looking at indexing journals, abstracting journals, reference lists or library holdings at a medical library. This depends, again, on the resources available. You could also contact researchers working in the field on the topic that you are interested in. Names of researchers working in the field are usually available on published articles or guidelines, where the first author is often the corresponding author and an email address is supplied in the article.

There are databases for unpublished data as well. In some instances, reports and conference proceedings contain good accounts of unpublished data<sup>11</sup>. One may also consult the websites of organisations, for instance the International Union against Tuberculosis and Lung Disease website to access the annual conference abstracts. Another possibility is to access different conference websites<sup>12</sup>, which include many of the global health conferences. Grey literature could be searched manually, but this is mostly done based utilizing library resources and a subscription is needed. A manual search is suggested even if an electronic search is complete (including reports, conference proceedings and theses) and references of grey literature can be found in other sources as well. Contacts or experts could also be asked to contribute, for example with their conference abstracts. When experts are contacted directly, make sure that you have a list of specific questions, i.e., a questionnaire which you can go through with every researcher that you contact. It is also best to contact the researcher beforehand and send the questionnaire, to ensure that s/he has the information at hand when you conduct the follow-up.

Now we will go through an example of a Medline search, but please keep in mind that each database differs and you will have to study each database separately to know how to use it efficiently and effectively.

The website where Medline is accessed, is [www.ncbi.nlm.nih.gov/pubmed/](http://www.ncbi.nlm.nih.gov/pubmed/). You will need an active internet account to access this website, unless you have access to it through a university. The basic search strategy is as follows:

STEP 1: Divide the search question into concepts

To refresh your memory, here is the search question again: *Is isoniazid prophylactic therapy (IPT) effective for the prevention of tuberculosis in children?*

The concepts in this search question could be:

- Isoniazid prophylactic therapy
- Prevention of tuberculosis
- Childhood tuberculosis

STEP 2: Compile a composite term to represent each concept

You can either use free-text (exact words from the title or abstract to search a database such as Google Scholar) or MeSH (when using Medline or a controlled thesaurus based database).

MeSH terms could be: 'tuberculosis' or 'prevention', and could be extended according to the MeSH structure on Medline.

---

<sup>11</sup> <http://opensigle.inist.fr/> or <http://www.ntis.gov>

<sup>12</sup> <http://www.webofconferences.org>

STEP 3: Combine the individual concepts with Boolean operators AND, OR, or NOT

Remember that the order of the Boolean operators used may make a difference to the result you get when searching. So keep careful notes of which concepts you combined with which Boolean operators and in which order you have combined and used them.

The positive predictive value of your search will increase by using more specific thesaurus terms, specifying major topics, using subheadings, using additional thesaurus or free-text terms or using the limit options of the database itself. The sensitivity of your search could be increased by adding additional MeSH terms or free-text terms, by looking for key authors, searching for 'related articles', using additional data sources and searching manual sources. These strategies all contribute to a more complete and accurate literature review.

Once your search is complete, it is crucial that you record your search strategy in order to be able to replicate it in future. The following details should be recorded: date of search, data sources selected, search terms used (MeSH, and how related), any limits applied, the results of the search, which abstracts were read, which articles were read, and which were evaluated and included in the review. Remember that materials (for example, full text articles or abstracts) can be obtained on the Internet (Pubmed Central) or from other sources such as a medical library (with interlibrary loans) and/or reprints. Colleagues can also assist with copies of literature material.

A last point on literature reviews is about referencing. Different systems of referencing can be used, and depending on the academic institution you are affiliated to, you will either use the Harvard system or the Vancouver system. Make sure that you know which system to use and how to use it. These systems have been described in detail and the descriptions are available at libraries or on the internet. There are also computer software programs available which can assist with referencing, such as Endnote (which must be purchased) or Zotero (which is downloaded free of charge).

## **6.2 Strengths and limitations**

All operational research has its strengths (the qualities that deliver results) and limitations (conditions that constrain or restrict). These exist on a continuum and relate to issues such as the research topic, scope of research, research methods (including the use of routinely collected data) and the extent to which research is action-oriented and aimed at yielding practical results and/or at developing solutions. An honest discussion of the practical or methodological issues that could influence findings and subsequent action is important if one aims to influence policy or practice.

Issues to consider along such a continuum include:

- the extent to which there is participation from researchers, service providers and policy makers as well as their influence
- the cost of the research (think also about costs that will not necessarily be borne by your project, for example, the cost of facility staff time where they are required to participate and the use of facility space or other resources as this is a contribution from the health services)
- the geographic scope and limits of the research
- the timeliness of results
- the acceptability of likely changes based on research findings
- the ethical acceptability of the research

- the extent to which results will be disseminated (locally and internationally and the means through which this will be done, for example, feedback through different fora, presentation at conferences and papers published)
- the ability to influence policy, improve practice and ultimately lead to better health outcomes for the population served.

The use of routinely collected data for OR has several strengths: it is relatively simple as standardised information is readily available that is relevant to the problems experienced in addressing the TB epidemic. The information available has local relevance and is comparable across different sites. Routinely collected data is less expensive to use compared to collecting new information.

There are also substantial limitations to using routinely collected data. The quality and completeness of records tends to be lower than the usual research standards. The reliability of the information may be questionable. For example, to what extent does the information in the electronic TB register (ETR) correlate with that in clinical records? This will need to be assessed if the ETR is used as the source for information. There is also the possibility of introducing bias because of categories of clients whose folders are not available or not entered into the TB register. Initial defaulters for example are often not recorded in the TB register.

Data may also not be comparable over time and between countries. The recent change in the international definition of a smear-positive case (from at least two smears positive to at least one smear positive) is one such example. In the past, South Africa did not include transfers out in the denominator when calculating treatment outcomes, resulting in better treatment outcomes than one would find using the current definitions. It is important to document relevant issues for your research.

### **6.3 Significance and impact**

Significance reflects the importance of the research to the health system:

- Does the project address an important “barrier” to the delivery of health services?
- What is the magnitude and severity of problem?
- Is this a local or national priority research area?
- What is the socioeconomic relevance? Who is affected? How will the research address issues of equity and reduce inequities in health services.

The impact is the effect of the research (and perhaps the subsequent proposed intervention) on the health system and particularly on people’s health and wellbeing. This includes both positive and negative effects, whether intended or unintended. The Impact Assessment Framework developed by the Liverpool School of Tropical Medicine, provides an approach to considering impact. Although developed in the context of assessing the impact of new diagnostic tools, it is useful in the context of OR.

1. Effectiveness	How likely is the proposed intervention to achieve the desired outcome?
2. Equity	Who needs the intervention most? Who is most likely to benefit? (Socioeconomic; sex; age, patient groups)
3. Health System	What are the: human resource, infrastructure, operating procedures, procurement, monitoring/evaluation implications of the proposed intervention?
4. Scale-Up	What will scale-up entail (costs, outcome)
5. Horizon Scanning	What other similar changes to policy or practice are being considered/available? How do these compare to the proposed intervention?

Whilst a thorough impact assessment may be beyond the scope of most OR projects, an impact assessment may increase the likelihood of the research influencing policy and practice.

---

<sup>13</sup> Adapted From: Mann G et al. *Beyond accuracy: creating a comprehensive evidence base for TB diagnostic tools. Int J Tuberc Lung Dis* 2010;14:1518-24. Adapted to also reflect a change in terminology (pers comm. Dr S.B. Squire).

## 7. What is the context?

### 7.1 Ethical issues and completing ethics forms

Ethics are rules or principles that govern appropriate and correct conduct, but this definition the question, what is appropriate and correct conduct? Is 'appropriate and correct conduct' the same for science, for the subjects of research and for society at large? It is important to think carefully about this question as ethics based on social benefit and scientific merit are potentially in conflict with ethics protecting the rights of research subjects. Thus, when thinking about ethics, all different aspects of a complex problem should be thought through and addressed, and usually the answer to the dilemma consists of a compromise between the various ethical principles and roles.

Medical ethics as we know it has a relatively short history. This history started after the Second World War at the Nuremberg Trials when the Nuremberg Code was compiled. This included the first guidance on informed consent in research.

Subsequently the Declaration of Helsinki was developed and adopted by the World Medical Association in 1964, with 2008 as the most recent revision. Further ethics guidelines were developed at the Council for International Organizations in Medical Sciences (CIOMS) in 2002. In these guidelines, the emphasis is mainly on the protection of the rights of individuals. These rights include (according to the United Nations Universal Declaration of Basic Human Rights):

- Articles 1 and 3: the right of freedom to decide to participate in research
- Articles 3 and 5: the right of freedom from harm during the course of experimentation
- Article 12: the right of personal privacy

In terms of research, the implications of individual rights are analysed in terms of three basic principles, namely:

- autonomy,
- non-maleficence and beneficence (which some groups count separately) and
- justice. Usually compromises are needed between the three principles in order to find a solution to an ethical dilemma.

What does this mean in practice? It means that one has to think about ethical principles when writing a research study protocol and that ethical issues must be addressed in the protocol. These issues must be reviewed by peers and by an ethics committee, which must be properly constituted according to specific guidelines. It also means that study findings must be published in order to realise the 'greater good'. In developing countries, the conduct of research may also entail wider public health responsibilities, for instance healthcare of research participants in specific contexts.

#### Free and informed consent

If an individual is approached to take part in a research study, there must not be unreasonable pressure on the individual to take part in the study. This means that incentives must be evaluated within the possible participants' context. However, the contradiction between autonomy and scientific validity must also be taken into account. A high response rate is essential for a study to be scientifically valid and especially in minimum risk research, incentives may be a good strategy to maximise response rates.

Sometimes incomplete information is given to participants as part of the informed consent process because of the study design, for instance, in the case of a case-control study. If participants are informed about the specific risk factor that the

hypothesis is based on, this may consciously or subconsciously influence their response and may bias the study results. Therefore, the information leaflet could rather include a general statement about risk factors being investigated, rather than focusing on the specific hypothesis.

#### Privacy and confidentiality

Any participant has the right to decide whether and how any information regarding his/herself is used. This does not however, apply to information which is publicly available, but it means that informed consent should be received before the start of the study. Such a process may reduce participation rates and validity.

A participant's anonymity should also be ensured by making personal data non-identifiable. There are also caveats here, because some socially important questions can only be answered by not making the data non-identifiable. Sometimes the researcher has to identify the possible participant before tracing is possible, for instance.

#### Physical or psychological damage

The sampling of blood or tissue may be physically damaging and this aspect must be included in the informed consent document. When a participant has to recall embarrassing events or feels anxious about confidentiality, it may cause psychological damage, which must also be addressed in the information leaflet. If there is any possibility of stigma within the group or community, community consent must be discussed.

#### Ethical practice in research

It is important to realise that the responsibility of good ethical practice remains with the researcher, even though the protocol has been approved by an ethics committee. Informed consent must be received from each participant, although sometimes it is waived by the ethics committee for minimally invasive research. Information must however still be provided to all participants and this process must be included in the protocol. Sometimes the informed consent process takes place in two steps, particularly when genetics research is suggested. In community-based intervention trials, community consent is crucial. When doing research in children, assent and consent is needed. In other vulnerable communities, such as individuals with impaired mental capacity or unsophisticated population groups, special informed consent processes may be needed to ensure understanding and insight.

#### *Access to personally identifiable data sources*

The use of medical records or samples obtained for clinical purposes can only be justified for research when:

- access to the samples/records is essential to achieve the objectives of the research
- there are no alternative sources
- informed consent is logically or economically impracticable OR could prejudice scientific value
- there is consent of the custodian of the samples/records
- the data are protected against those not involved in the research
- the data are not used for new research (unless the researcher re-applies for approval)
- the research is approved by an ethics committee

#### *Confidentiality of personally identifiable data*

A written code of practice for each study protocol must include signed declarations from researchers to keep data confidential. Records must be kept separately

(physically under lock and key and logically) but could be linked by barcode. Records must only be kept as long as it is needed (which may depend on the ethics committee and the funding source). Records must be disposed of securely and publications may never include material which could identify subjects. Be aware that in community-based randomised trials for instance, the community name must also not be published if it could lead to stigma.

#### *Permission to approach subjects*

When medical records will be used in a research study, it is common courtesy to approach the custodian of the records (or a nominee) to ask permission to use the records. The custodian of medical records is usually the Department of Health. Usually the custodian will be sensitive to contextual issues and know if the proposed research may cause harm. If this consent is not obtained *a priori*, access to medical records would probably be denied for future projects.

#### Communication of results

For any outcome of a research study, there is an obligation for the researcher to inform the participant (especially when a risk factor or disease is discovered). This process should include an explanation of the significance of the finding, a recommendation of appropriate action and ensuring that the recommendation is being followed (if the participant has consented). The researcher has to provide information on study outcomes in general if the risk is real, but be sensitive not to cause anxiety by reporting the findings out-of-proportion and context. The outcomes of genetics research should be discussed with participants who gave consent. If any study findings include conditions or risk factors, which may be hazardous to others, the results must be communicated taking into account the individual's right to privacy and not reporting any findings out-of-proportion.

Lastly, in an article by Ezekiel Emmanuel<sup>14</sup>, the author addressed the question "What makes research in developing countries ethical?" which can be summarised in the following points:

- development of collaborative partnerships in which responsibilities are shared, the community is respected and local capacity is developed while the communities benefit from the research
- address the social value of the research, i.e., who are the beneficiaries of the research, what is the value of the research and how will the knowledge be disseminated
- ensure that the scientific validity of the study is sound, i.e., that the design realises social value and the scientific objectives, and that it is feasible (socially, politically and culturally)
- the study population must be selected in a fair manner, taking into account scientific validity, minimal risk strategies and vulnerable populations
- there must be a favourable risk-benefit ratio when comparing risks and benefits from the collaborative partnership, the social value and the respect for the study population
- when independently reviewed, there must be evidence of public accountability, transparency, independence of the researchers and competence.
- the informed consent process must involve the community and disclose information in culturally and linguistically acceptable formats; community members must have the freedom to withdraw or refuse.

---

<sup>14</sup> Emanuel, Ezekiel. What Makes Clinical Research In Developing Countries Ethical? The Benchmarks of Ethical Research. *J Infect Dis* 2004;189 (1 March).

- the participants and communities must be respected by ensuring confidentiality, withdrawal of a participant voluntarily without penalty, dissemination of information during and after the study and providing interventions for medical conditions, at least as good as local healthcare would provide.

Ethical dilemmas are always challenging, but by keeping these basic principles in mind, you will have a framework on which to base your ethical decisions.

## **7.2 Regulatory aspects of research**

In order to submit your proposal for ethics review, you may need to be working with an academic or other institution with a recognised ethics committee. Each institution has its own guidelines and procedures regarding the ethics review process. It is important that you contact the institution you are affiliated or working with in advance to ensure that you complete the correct forms and hand in it before the next meeting of the ethics committee. In some instances, ethics committees only meet every six months, which may mean that your project could only start after the next meeting. If you are collaborating with someone from a different institution, you may be advised to submit your proposal to both these institutions, especially if the institutions are in different countries.

Important aspects to take into account when submitting your proposal to an ethics committee are:

- is a waiver of consent is needed (especially when the study has been discussed with the Department of Health who are the custodians of the data and is a retrospective analysis with anonymised data)?
- if it is a retrospective study, what is the responsibility towards the participants?
- ensure that the motivation/justification and impact of the study is clearly stated in the application.

Only when the ethics committee has given its approval can you start the study. If your study will continue for longer than a year, an annual report will have to be submitted to the ethics committee. The study can only continue if the committee, after review, gives its approval again. The committee can at any time request to audit the study in order to ensure good ethical practice. This includes good scientific practice, which underlies any ethics review.

## 8. What does it take to move ahead?

Research projects have three phases. The first phase, as you are discovering in this text, is *proposal development* in which the methods and plan of research must be precisely developed and described before research begins.

The next phase is *project execution*, meaning the steps you must take to ensure that the project is carried out on time and on budget, and that the information you collect is of high quality, ethically collected and will allow you to meet your objectives.

The final phase consists of interpreting and reporting the results.

In order to attract support for your research proposal it is important that the *project execution* phase is described in detail in your proposal, so that there is a clear indication on how the project will be executed once approvals have been obtained.

### 8.1 Work Plan and Timelines

A written work plan with timelines will help to ensure that the study is carried out within the timeframe specified and within the allocated budget.

The work plan should begin by listing all the tasks to be undertaken in the study and is usually presented as a narrative but it is also useful to present the activities with the timelines in a Gantt chart, either as part of the proposal or as an appendix. Typically, the work plan, with timelines, would include when the following important “milestones” are expected to be reached:

- finalisation of research protocol
- submission of documents for ethics approval
- receipt of final approvals (e.g. health departments) and allocation of funding
- signed of a contract/service level agreement with the donor/funding source (if applicable)
- implementation of research:
  - recruiting study personnel
  - training study personnel
  - procurements
    - stationery
    - printing forms and questionnaires
    - furniture and equipment
  - piloting techniques and procedures (if necessary)
  - recruiting participants/sourcing data
  - collecting, checking and collating the data
  - analysing the results
- preparing progress reports to mentors and donors/funding source (if applicable)
- preparing scientific reports and preparation of paper/abstract
- presentation of the results to relevant role players (e.g., donors, funders, health departments, community involved, the academic community)
- publication of paper/abstract to scientific journal/conference organisers

<u>Example:</u>												
Timelines												
Activity: Year 1	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Finalization of research protocol												
Submit documents for Ethics approval												
Final adjudication; approval and allocation of funding. Obtain provincial permission Finalise donor contract												
Extraction/cleaning of routinely collected data												
Qualitative tool development												
Activity: Year 2	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Qualitative data gathering												
Data processing and preliminary analysis												
Analysis												
Report/article writing and feedback												

## 8.2 Roles and responsibilities – who, what, how and when?

Once all the activities have been listed, responsibility for carrying out each of the tasks should then be assigned to an individual member of the research team.

The responsibilities of the principal investigator (PI) include:

- Project management
- Ethics
- Quality assurance
- Public relations
- Analysis and reporting of the results of the study to the participants, authorities and the scientific community.

### Project management

The PI will identify the *personnel* required to carry out the research and define their tasks. The PI will justify the personnel proposed in terms of the tasks, the level of training necessary to carry out the tasks and the amount of time required. When the number and type of study team member has been determined, the responsibilities of

each member of the research team will be defined and recorded. The advertisement, recruitment and employment procedures of the funding source must be followed.

The PI will set out the *duration of the project and the timelines* anticipated for each phase of the study and prepare the work plan. The PI will take responsibility for ensuring that the work plan and timelines are followed.

The PI will be responsible for the *budget*, but should also share budget information with all staff. This need not include staff salary information, but should include the duration of time each staff member is assigned to the project

If the PI delegates functions regarding the schedule and budget, the PI should hold periodic progress meetings or obtain reports from appropriate team members.

Most funding sources require a contract to be signed and the PI would take responsibility for completing and submitting the relevant documentation.

### Ethics

The ethical considerations, principles and ethics review procedures are the responsibility of the PI (see previous chapter). Health authorities may also require permission to be granted for access to facilities, patients and data.

### Quality assurance

The PI must ensure that the procedures of research *precisely follow* those outlined in the protocol. The precise *indicators* for evaluating the quality of information and measurements need to be specified and the technique for recording and interpreting them defined.

A number of steps may be taken to ensure that the quality of the research is maintained at a very high level. This includes pilot testing of untried methods. Procedures must be tested in advance to ensure that they will work as planned. For example, if your protocol depends on recruiting 10 new patients every day, you should check (either through existing records or with a small pilot test) that this would be feasible. Similarly, all new equipment must be tested and questionnaires pilot tested to ensure that respondents are able to complete them as you expect. Many investigations start with a funded pilot phase, before the main study protocol is developed.

There must be procedure manuals for each step in collecting the data and for keeping track of the information after it has been collected. Manuals of procedures must be developed that describe exactly how techniques are to be carried out. In some instances, such instructions already exist; in others, they must be developed for the specific study. In addition, a careful plan must be developed and recorded for the *management of data*. Special attention must be paid to data tracking, ensuring quality control of data, checking for errors and ensuring secure storage of the results

Staff who will be collecting the information must be trained for the study. Even for staff with specific clinical expertise (e.g. clinicians, respiratory technicians, radiologists etc.) the PI needs to ensure that these people understand the research objectives. For example, the procedures for eliciting a clinical history are different from those for completing a research questionnaire; technicians who are used to dealing with very ill patients in hospital may not realise the need for different research procedures when dealing with a mostly healthy study population.

When more than one person carries out the measurements or when information is being collected at more than one centre, it is essential to make comparisons of the results obtained by the various technicians and by the same technician on different occasions. This ensures that there is close correlation of the results (inter and intra-observer comparisons), or at least that you have a measure of the differences across centres or technicians.

#### Public relations

The PI should provide regular feedback to the relevant team members regarding progress of the research and should be able to address any concerns, especially from the community. Should there be any untoward events related to the research, the PI should make a public statement to allay fears and answer questions.

#### Analysis and reporting of the results of the study to the participants, authorities and the scientific community.

Follow-up procedures for study participants should be specified where appropriate (e.g. will each study participant be informed of the study results? How will you respond if you uncover clinically relevant disease in a participant who is not being treated?)

A special session should be planned and organised by the PI where the results of the study are made known to the community and authorities.

### **8.3 Budget**

When drawing up a research budget, each expenditure item expected to be needed in the conduct of the study must be specified, even if the cost is covered by routine operations of the health service or by other sources outside the study itself. A budget is usually presented as a spreadsheet, in local currency and in the currency of the potential donor/funding source. A written budget justification should be included to explain the various expenditures in further detail. Expenditures should, as much as possible, be given in units, hours, trips, kilometres etc. For example, salaries can be calculated against full time equivalents or per hour according to qualifications.

A budget is divided into categories, such as: personnel; travel and transportation; accommodation; equipment ; materials; and other costs (e.g., communications, rentals, honorariums, contracts with service providers)

#### Personnel

Relevant personnel may need to be appointed (e.g., research nurses, research assistants, data capturers etc.) on a part-time or full time basis. A human resources officer should be consulted regarding job descriptions, level of experience and salary levels to ensure that roles, responsibilities, and accountabilities are defined and clearly understood, and that deliverables are achieved. It is important that employment contracts and other conditions of employment are well defined and in place.

#### Travel and Accommodation

The geographic realities associated with research project activities must be taken into account as well as the mode of transport that research personnel will use (e.g., own vehicle, bus, taxi, train, etc.). The cost effectiveness of each travel day should be measured and trip plans designed accordingly. Justification for overnight accommodation should need to be given. Projected costs for travel and accommodation must be presented within the budget and a narrative on

reimbursement for travel will be helpful in the budget justification. *Per diem* rates should be decided upon and budgeted accordingly.

### Equipment

The budget should include all equipment for the project on-site (e.g., scale, laptop & desktop computers, printers, desks, chairs, and filing cabinets). Some donors require that distinctions are made between minor assets (not capital) and between assets of various values (e.g., assets valued <R5000 and those >R5000).

### Materials

Basic supplies are required for successful implementation and coordination of project activities including routine items such as paper, copies, faxes, printer ink cartridges, folders, pens, binders, flip charts, dry erase boards, training books, and funds for printing of select project documents.

### Other Costs

*Office rental*-Consideration must be given to where appointed staff will be accommodated and office rentals included in the budget if necessary.

*Telecommunications*-Funds should be requested to defray telecommunications costs by project personnel as project activities are implemented. This includes additional costs associated with internet charges, mailing, faxes, and telephone for project personnel.

*Training*-Training of project personnel may be required for successful implementation and coordination of project activities. Training should be arranged either through a registered training service provider or the researcher may offer in-house training to appointed personnel. If a venue, materials or refreshments are required for the training, it should also be budgeted.

*Printing*-Depending on the project, it may be necessary to budget for document printing pertaining to meetings, workshops and other interactions and/or bulk copying of documents (e.g., questionnaires and case report forms). The project may also require development and printing of a variety of posters, pamphlets, and educational materials in order to support key activities and messaging.

*Translation*-If consent forms and questionnaires are to be used in the research project, these research tools must be translated into at least three local languages. Costs for translation should be calculated and included.

*Ethics review*-Research proposals will require ethics review from a registered/recognised Health Ethics Committee. These committees usually have a fixed fee, which should be added to the budget.

*Honoraria*-Small honoraria may be required to recognize the contributions of key collaborators who play important roles in the successful implementation of project activities, particularly with regard to databases and statistics within the approved projects.

*Consultants*-A consultant may be employed on a short term hourly or daily contract. For example, a consultant may be contracted to build a database for the project, assist with questionnaires, etc.

**Table 7: Example of a Budget Spreadsheet**

OPERATIONAL RESEARCH PROJECT BUDGET (Beginning date) to (End date)						
Study Title:						
Categories	Item	No. of Units	Unit Cost	Amount in ZAR	Amount in US \$ (R6.80/\$1 US)	
<b>Personnel</b>						
	Study Nurse	FTE		0.00	0.00	
	Research Assistant	FTE		0.00	0.00	
	Data Capturer	FTE		0.00	0.00	
	Clinical Assistant	FTE		0.00	0.00	
<b>TOTAL PERSONNEL COSTS</b>				<b>0.00</b>	<b>0.00</b>	
<b>Travel &amp; Transportation</b>						
	Provincial Travel	trips		0.00	0.00	
	District Travel	km		0.00	0.00	
	Car hire for travel to remote areas	days		0.00	0.00	
	Accommodations	days		0.00	0.00	
<b>TOTAL TRAVEL &amp; TRANSPORTATION COSTS</b>				<b>0.00</b>	<b>0.00</b>	
<b>Equipment</b>						
	Scale	unit		0.00	0.00	
	Computers, printers, external hard drives	unit		0.00	0.00	
	Office furniture, filing cabinet, desks, chairs	unit		0.00	0.00	
<b>TOTAL EQUIPMENT COSTS</b>				<b>0.00</b>	<b>0.00</b>	
<b>Materials</b>						
	Stationery	unit		0.00	0.00	
<b>TOTAL MATERIALS COSTS</b>				<b>0.00</b>	<b>0.00</b>	
<b>Other Costs</b>						
	Honorarium, Statistician	per hour		0.00	0.00	
	Honorarium, Data management	per hour		0.00	0.00	
	Printing	lump sum		0.00	0.00	
	Telephone & IT Cost	lump sum		0.00	0.00	
	Catering for training	lump sum		0.00	0.00	
	Venue use for training	lump sum		0.00	0.00	
<b>TOTAL OTHER COSTS</b>				<b>0.00</b>	<b>0.00</b>	
<b>TOTAL BUDGETED EXPENSES</b>				<b>0.00</b>	<b>0.00</b>	

**Table 8. Example of a completed Budget Spreadsheet**

<b>BUDGET: OPERATIONAL RESEARCH ASSISTANCE PROJECT: FIRST WAVE</b>						
<b>1 October 2010 to 30 September 2011</b>						
<b>Study Title: Association between delayed Tuberculosis treatment after pre-treatment sputum diagnosis and two-month sputum smear non-conversion</b>						
<b>Categories</b>	<b>Item</b>	<b>No of Units</b>	<b>Unit Cost</b>	<b>Amount in ZAR</b>	<b>Amount in US \$ (R6.80/1 USD)</b>	
<b>Personnel</b>						
	Principle Investigator	days	15	2,040.00	30,600.00	4,500.00
	Junior Researcher	days	30	1,020.00	30,600.00	4,500.00
	Data Manager	day	20	1,020.00	20,400.00	3,000.00
	Data Capturer	days	30	680.00	20,400.00	3,000.00
	Transcriber	days	10	680.00	6,800.00	1,000.00
<b>TOTAL PERSONNEL COSTS</b>					<b>108,800.00</b>	<b>16,000.00</b>
<b>Travel &amp; Transportation</b>						
	Provincial Travel to remote areas	car hire	12	388.00	4,656.00	684.71
	Provincial Travel to remote areas	km	1000	3.00	3,000.00	441.18
	Accommodation	days	12	550.00	6,600.00	970.59
	Follow up Workshop in Cape Town	days	2	4,000.00	8,000.00	1,176.47
<b>TOTAL TRAVEL &amp; TRANSPORTATION COSTS</b>					<b>22,256.00</b>	<b>3,272.94</b>
<b>Equipment</b>						
	Computer	unit	1	12,000.00	12,000.00	1,764.71
	External hard drives	unit	1	2,000.00	2,000.00	294.12
	Modem	unit	1	1,750.00	1,750.00	257.35
<b>TOTAL OF EQUIPMENT COSTS</b>					<b>15,750.00</b>	<b>2,316.18</b>
<b>Materials</b>						
	Stationery	lump sum	1	4,000.00	4,000.00	588.24
	Audio Tapes	unit	40	20.00	800.00	117.65
<b>TOTAL MATERIALS COSTS</b>					<b>4,000.00</b>	<b>588.24</b>
<b>Other Costs</b>						
	Honorarium - Statistician	per hour	120	300.00	36,000.00	5,294.12
	Honorarium - Data management	per hour	120	250.00	30,000.00	4,411.76
	Printing	lump sum	1	5,000.00	5,000.00	735.29
	Telephone & IT Costs	months	12	1,225.00	14,700.00	2,161.76
	Catering- feedback to role players	unit	150	100.00	15,000.00	2,205.88
	Ethical review	lump sum	1	3,000.00	3,000.00	441.18
<b>TOTAL OF OTHER COSTS</b>					<b>103,700.00</b>	<b>15,250.00</b>
<b>TOTAL BUDGETED EXPENSES</b>					<b>254,506.00</b>	<b>37,427.35</b>

## **8.4 Getting support for research**

Gro Harlem Brundtland, former Director General of the World Health Organization, has stated, "Developing countries must build up their own basis for research. Only they will be able to establish the diagnosis and implement the cure. The international community must assist the process".

### **8.4.1 Local sources**

If one considers all the resources (both human and material) expended on research in low-income countries, unquestionably the countries themselves contribute the lions' share. Relevant research that is also cost-effective almost always originates at the bedside of the "patient" (or, in the case of public health activities, the programme). Research and practice must always go hand in hand. This is the reason that the Commission on Health Research for Development has proposed (and many funding agencies have accepted) that a portion (2-5%) of the budget of all public health programmes must be set aside for research. This is the most important source of support for research because it is more likely to be targeted to research relevant to practice in the community.

A second source of financial support for research at the local level is central government. This support is usually directed through the official channels such as universities and research institutions. To access such support, investigators usually must have collaboration with such institutions or other groups. For clinicians and programme managers, this should be a fruitful path to research. Should there be support available from both the programs themselves and special research funds, the two can sometimes be joined leading to highly productive research which takes its hypothesis from the field but harnesses research expertise such as epidemiology or statistics to carry it out.

A third source of financial support for research at the local level is the pharmaceutical industry. If the topic is of particular interest to the firm, research support may very well be made available. Much of the recent research in low-income countries into the distribution and determinants of asthma has been funded in this way.

A final source of local funding is the development agencies, both governmental (through embassies) and non-governmental. Frequently the development departments of local embassies have limited (but often sufficient) funds for support of humanitarian and community action. Service organisations, such as the Lions or Rotary Clubs, may provide support to research at the local level. If a good rapport can be developed with the officials of such programmes, and if the research can be shown to have clear practical relevance to the situation in the country, funds might be obtainable.

### **8.4.2 International sources**

International agencies that supports research in low-income countries, are few in number but fund a substantial amount of such research. The agencies may be multilateral (the United Nations system and the European Union) or bilateral (Cupertino agreements between industrialised and low-income countries).

#### Multilateral agencies

The United Nations system is the most active agency in this regard. Most assistance for health research from the United Nations is provided through the World Health Organization (WHO). This budget increased greatly during the 1980s to over 50

million dollars annually. Most of this budget (over two-thirds) has been spent on disease prevention and control and has been managed mainly through two agencies of the WHO, the Human Reproduction Programme and the Tropical Diseases Research Programme. The latter programme has spent a large amount of money on research into malaria, schistosomiasis, filariasis, trypanosomiasis, leishmaniasis and leprosy. Interestingly, although lung diseases are the single most frequent cause of death in the world in small children and young adults, relatively little money has been committed to research in this area. Future priorities for activities in WHO will include tobacco (prevention and control) and this may be a fruitful area for obtaining research support in the near future.

The European Union has a very large budget for health research and has specific structures to promote research collaboration with low-income countries. This research, however, almost invariably takes the form of international collaborative research and usually requires the collaboration of several research centres in Europe in order to be eligible for funding.

#### Bilateral agencies

Traditionally, there have been only two agencies specifically mandated to provide support for health research in low-income countries. These were SAREC (Department of Research Cooperation), a Swedish Government agency (which now works in conjunction with the European Union) and IDRC (International Development Research Centre), a Canadian Government agency whose specific mandate is research on development. Traditionally, the IDRC has supported a great deal of health research but more recently has less of a focus on specific health research. The current emphasis in this agency is on health services and systems research.

Other bilateral co-operation agencies fund research, although that is not their specific mandate. The Japan International Cooperation Agency funds a number of projects in collaboration with their citizens in a variety of low-income countries. Other agencies (for example, in the USA and UK) frequently channel their support through national research institutions in their own countries.

#### Foundations

A number of international foundations have provided research support. The main foci of research in these foundations have been infectious and tropical diseases (one third) and epidemiology, policy and management (one-third). Major donors to health research in low-income countries have been:

- The Aga Khan Foundation, Switzerland (primary health care management),
- Carnegie Corporation, USA (human resources development),
- Edna McConnell Clark Foundation, USA (tropical diseases research),
- Ford Foundation, USA (reproductive health),
- MacArthur Foundation, USA (tropical diseases, women's health),
- Pew Charitable Trusts, USA (health policy and management),
- Rockefeller Foundation, USA (population, neglected diseases),
- Sasakawa Memorial Health Foundation, Japan (leprosy),
- Wellcome Trust, UK (medical and veterinary research).

## **About The Union**

Founded in 1920, the International Union Against Tuberculosis and Lung Disease (The Union) is dedicated to bringing innovation, expertise, solutions, and support to address health challenges in low- and middle-income populations. With nearly 10,000 members and subscribers from over 150 countries, The Union has its headquarters in Paris and offices serving the Africa, Asia Pacific, Europe, Latin America, Middle East, North America, and South-East Asia regions. Its scientific departments focus on tuberculosis, HIV, lung health, and non-communicable diseases, tobacco control and research. Each department engages in research, provides technical assistance and offers training and other capacity-building activities leading to health solutions for the poor.

For more information about The Union, please visit [www.theunion.org](http://www.theunion.org)

## **About the Desmond Tutu TB Centre**

The **Desmond Tutu TB Centre** is an academic research centre of the Department of Paediatrics and Child Health, Faculty of Health Sciences. It has its main offices on the Tygerberg Campus and satellite offices in various communities affected by TB and poor health. It has as its mission the improvement of the health of vulnerable groups through influencing policy based on new knowledge created by research focusing on health, mainly TB and HIV. To achieve this, the Centre works closely with the Department of Health and the local communities. It provides training to academic and health services staff, builds capacity in the University and the Department of Health, provides service to communities and advocates for TB and health. Himself a former TB sufferer, Archbishop Desmond Tutu champions tuberculosis research and care. He is also the patron of the on-campus Tygerberg Children's Hospital.

For more information, please visit [www.sun.ac.za/tb](http://www.sun.ac.za/tb)