Dying to count: mortality surveillance methods in resource-poor settings

Edward F. Fottrell
The statistics in this work are more than just numbers – they represent individuals, families and whole communities.

I dedicate this work to them.
ABSTRACT

Background Mortality data are critical to understanding and monitoring changes in population health status over time. Nevertheless, the majority of people living in the world’s poorest countries, where the burden of disease is highest, remain outside any kind of systematic health surveillance. This lack of routine registration of vital events, such as births and deaths, constitutes a major and longstanding constraint on the understanding of patterns of health and disease and the effectiveness of interventions. Localised sentinel demographic and health surveillance strategies are a useful surrogate for more widespread surveillance in such settings, but rigorous, evidence-based methodologies for sample-based surveillance are weak and by no means standardised. This thesis aims to describe, evaluate and refine methodological approaches to mortality measurement in resource-poor settings.

Methods Through close collaboration with existing community surveillance operations in a range of settings, this work uses existing data from demographic surveillance sites and community-based surveys using various innovative approaches in order to evaluate and refine methodological approaches to mortality measurement and cause-of-death determination. In doing so, this work explores the application of innovative techniques and procedures for mortality surveillance in relation to the differing needs of those who use mortality data, ranging from global health organisations to local health planners.

Results Empirical modelling of sampling procedures in community-based surveys in rural Africa and of random errors in longitudinal data collection sheds light on the effects of various data-capture and quality-control procedures and demonstrates the representativeness and robustness of population surveillance datasets. The development, application and refinement of a probabilistic approach to determining causes of death at the population level in developing countries has shown promise in overcoming the longstanding limitations and issues of standardisation of existing methods. Further adaptation and application of this approach to measure maternal deaths has also been successful. Application of international guidelines on humanitarian crisis detection to mortality surveillance in Ethiopia demonstrates that simple procedures can and, from an ethical perspective, should be applied to sentinel surveillance methods for the prospective detection of important mortality changes in vulnerable populations.

Conclusion Mortality surveillance in sentinel surveillance systems in resource-poor settings is a valuable and worthwhile task. This work contributes to the understanding of the effects of different methods of surveillance and demonstrates that, ultimately, the choice of methods for collecting data, assuring data quality and determining causes of death depends on the specific needs and requirements of end users. Surveillance systems have the potential to contribute substantially to developing health care systems in resource-poor countries and should not only be considered as research-oriented enterprises.

Key words: mortality; surveillance; verbal autopsy; survey methods
ORIGINIAL PAPERS

The thesis is based on the following papers:


The original papers are reproduced here with permission from the respective publishers.
ABBREVIATIONS

BRHP – Butajira Rural Health Programme
CDC – Centres for Disease Control and Prevention
CRSN – Centre de Recherche en Santé de Nouna
CSMF – Cause Specific Mortality Fraction
DHS – Demographic and Health Survey
DSA – Demographic Surveillance Area
DSS – Demographic Surveillance Site
EBM – Evidence Based Medicine
EPI – Expanded Programme on Immunisation
GPS – Global Positioning System
HIV/AIDS - Human Immunodeficiency Virus/Acquired Immunodeficiency Syndrome
HRS – Household Registration System
ICD – International Classification of Diseases
ICD-10 – International Classification of Diseases, version 10
Immpact – Initiative for Maternal Mortality Programme Assessment
INDEPTH – International Network of Field Sites with continuous Demographic Evaluation of Populations and Their Health
MDG – Millennium Development Goal
PDA – Personal Digital Assistant (handheld computer)
PPES – Probability Proportional to Estimated Size
PPS – Probability Proportional to Size
PPV – Positive Predictive Value
PR – Physician Review
PSU – Primary Sampling Unit
P-Y – Person-years
SP – Symptom Pattern method
SPSS – Statistical Software for the Social Sciences
STEPS – WHO STEPwise approach to chronic disease risk factor surveillance
TB – Pulmonary Tuberculosis
UNICEF – United Nations Children’s Fund
USAID – United States Agency for International Development
WHO – World Health Organisation
VA – Verbal Autopsy
MATERIALS AND METHODS ........................................................................................................................................ 24

Empirical Investigations Using Existing Data ............................................................................................................ 24

Data Sources .............................................................................................................................................................. 24

Data Acquisition .......................................................................................................................................................... 29

Sampling ........................................................................................................................................................................ 30

Data Quality .................................................................................................................................................................. 32

Cause-of-death Determination ..................................................................................................................................... 34

Data for Action ........................................................................................................................................................... 35

Ethical Considerations ............................................................................................................................................... 35

RESULTS ...................................................................................................................................................................... 37

Sampling ....................................................................................................................................................................... 37

Data Quality ................................................................................................................................................................. 39

Cause-of-death Determination ..................................................................................................................................... 40

Data for Action ........................................................................................................................................................... 43

DISCUSSION ............................................................................................................................................................... 44

Further Investigations ............................................................................................................................................... 55

CONCLUSION ............................................................................................................................................................... 56

ACKNOWLEDGEMENTS ............................................................................................................................................. 58

REFERENCES ............................................................................................................................................................... 60
INTRODUCTION

A Brief History of Mortality Surveillance

“To make people count we first need to be able to count people”


The concept of using mortality data as a basis for public health action arose in Europe some 600 years ago with the emergence of scientific thought during the Renaissance. Records of vital-events were preserved in numerous European towns beginning with the first London Bills of Mortality being prepared in 1532, although they were not used use for health and scientific purposes for some time. One of the earliest examples of surveillance for public health action was witnessed one hundred years later when London parish clerks made weekly reports of the number of burials, with cause of death, to the Parish Clerks’ Company who, in turn, were responsible for compiling the statistics of deaths for the City of London area and then interpreting them to provide information on the extent of plague. This information was disseminated in a weekly ‘Bill of Mortality’ to those who required it so action could be taken. Detailed analyses of the weekly Bills of Mortality were made by John Graunt (1662), who was the first to estimate the population of London and to count the number who died from specific causes. He was also the first to conceptualise and quantify patterns of disease and to understand that numerical data on a population could be used to study the cause of disease [1]. Meanwhile, the Swedish ecclesiastical registry was being initiated by cannon law, instructing all parish clergymen to keep records on vital events, such as births, deaths, migration, marriages and baptisms. By the mid-eighteenth century (1749), this system was implemented nationwide and the later establishment of a Registrar General's Office enabled national statistics to be compiled and statistical tables with mortality information by geographical areas were published. These data could then be used to monitor trends in public health which, in turn, influenced public health action [2, 3]. These early systems illustrate the main principles of surveillance which are still applicable today – data collection and analysis, interpretation to provide information, and dissemination of the information for action [1].

By the nineteenth century, the collection and interpretation of health-related data for the purpose of identifying appropriate actions was becoming fully established in many European countries. It was through surveillance that Sir Edwin Chadwick, Secretary of the Poor Law Commission in England, was able to confidently identify the close relationship between poverty and disease. Similarly, Lemuel Shattuck’s Report of the Massachusetts Sanitary Commission (1850) related living conditions to rates of infant and maternal mortality and morbidity, recommending a decennial census, standardisation of nomenclature for diseases and causes of death, and the collection of health data by age, sex, occupation, socioeconomic level and locality. By the middle of the nineteenth century, the General Register Office was established in the United Kingdom (UK) and Sweden’s Registrar General’s Office became the Central Bureau of Statistics, both with a view to meeting the demands for more accurate and complete mortality data, and, subsequently, universal death registration and medical certification of death by cause, which was instituted in the UK in 1837 [1-3].
Surveillance

As demonstrated by John Snow’s archetypal investigation of cholera in the nineteenth century, epidemiology traditionally focuses on the three related issues of person, place and time as the route to describing patterns of health and disease [4]. In general public health terms, surveillance is the ongoing systematic collection, analysis, interpretation and dissemination of data. The concept of who people are, where they live and when they are born, die or migrate is also key. It is not surprising, therefore, that some epidemiologists define surveillance as synonymous with epidemiology in its broadest sense, including investigation and research. However, it may be argued that surveillance has developed into a complete discipline quite distinct from epidemiology [1].

The general purpose of surveillance activities is to generate data for planning public health programmes, developing regional, national and global policies, and implementing and evaluating public health action [5]. Surveillance can relate to several specific aims, including quantitative estimates of the magnitude of a specific health problem, detecting epidemics, documenting the distribution and spread of disease and monitoring changes in infectious agents or risk factors. As such, surveillance activities may be long- or short-term in nature and can vary in character.

There are many sources of data that can be used for public health surveillance, which vary from country to country depending on the stage of development and sophistication of public health, medical and laboratory services, as well as information technology infrastructure. Death registration, epidemic, laboratory and morbidity case reporting, individual case reports, field investigations, surveys, animal reservoir and vector distribution studies and environmental data are all valuable sources of surveillance data. Additional secondary sources of data, such as hospital statistics or attendance records at work or school, are often created for unrelated purposes and may be used in supplementing routine surveillance data. One or a combination of data-collection methods may be used in a surveillance system.

Analysis of data is a dynamic, expert and intellectual process of interpretation that results in the production of important information on which to base action. The analysis process demands expertise and a broad awareness of existing relevant knowledge in the subject area, as well as skills in analytical techniques. Generally, analysis begins by addressing the key components of epidemiology – person, place and time.

Dissemination of information must be timely and requires communication skills and experience. In developing-country settings, where communication infrastructures are weak and literacy may be low, communication strategies must often seek innovative methods of dissemination. Surveillance reports serve two primary purposes of information and motivation, which are key to the longevity of surveillance in situations where vital-event reporting is not obligatory under law.

It is somewhat artificial to consider the key components of surveillance as discrete activities. Rather, they are interrelated and rely upon each other in a data process that must be carefully designed. Good design of the data process can yield quality data and information whilst poor systems can be a waste of resources and can be misleading.
The Need for Mortality Data

Mortality registration is the oldest form of disease surveillance and its importance for epidemiology and public health planning is perhaps obvious. Two of the Millennium Development Goals and one out of three essential elements of the Human Development Index are specific mortality measures [6, 7], enhancing the need for valid and representative data on mortality and its risk factors to track progress, evaluate disease-control programmes and monitor major global health initiatives.

In addition to the practical utility of mortality data, however, is the concept that one’s death must be recorded for their life to count. Failure in the registration of births and deaths has been described as ‘the single most critical failure of development over the past thirty years’ and the lack of any record of the lives of millions as a ‘scandal of invisibility’ [8]. Tracing the imprint of a person’s existence not only confirms their citizenship, but also represents the first step in securing their right to life, freedom and protection [9]. Having one’s death recorded can contribute to an invaluable legacy of greater knowledge and potential for the prevention of suffering and premature death for future generations – no death can then be in vain.

There are different levels at which mortality patterns are needed, ranging from the local to the global. Health managers require mortality data to effectively plan health services based on local patterns of disease. At this level, it is essential to be able to monitor major causes of death simply and cheaply. The breakdown of cause-of-death categories based on a few hundred cases is needed and very detailed causes of death would be superfluous. Epidemiologists, health service researchers and assessors of specific health interventions, such as safe motherhood interventions, need a consistent assessment of cause-specific mortality to determine trends in causes of death that enable evaluations of the effectiveness of interventions across time and regions. For such users, a reduction in rates of a specific cause of death is usually an important endpoint – for example, detailed sub-causes of maternal mortality. National and global authorities concerned with building respective pictures of health patterns require consistent and reliable cause-specific mortality data from a wide range of settings [10, 11].

Where There Are No Data

“It is not because countries are poor that they cannot afford good health information: it is because they are poor that they cannot afford to be without it”

Health Metrics Network, World Health Organisation, 2005

The Global Burden of Disease (2000) estimates that 63% of healthy life lost in 2000 resulted from premature deaths [12]. Information on deaths by cause is key to tackling this. However, cause-of-death recording methods among United Nations member states who have supplied data to the WHO for 1990 or later is only of a ‘high’ quality for countries with 12% of the world’s population, ‘medium’ for 17% and ‘poor’ for 5%; no data supplied for 66% of the world’s population [13, 14]. The chance of a death being registered and the cause of death documented strongly depends on the socioeconomic status of the community and nation in which it occurs [11]. Despite its well established utility, therefore, there has been little or no progress in civil registration systems
in the last fifty years, and between two-thirds and three-quarters of the world’s population remain outside any kind of systematic health surveillance [4, 14-17]. As such, we now know more about epidemiology and public health within certain eighteenth century European populations than we do about the current situation of many developing countries. Five hundred years after the early surveillance systems in London, a significant proportion of global births and deaths pass unrecorded. For epidemiologists working in low- and some middle-income settings this means vital registration often has to be implemented from first principles. Not only can this place demand on resources at the local level, it also precludes epidemiologic considerations at a regional or national level, and inevitably makes monitoring of health and population dynamics difficult. Estimates of key health and development indicators are frequently varied and wide ranging, with official estimates often being speculative and occasionally mystifying in their origin. For example, maternal mortality rate estimates derived by different methods in Ethiopia range from 230 to 1400 per 100,000 live births [18].

**Why There Are No Data**

The lack of progress towards complete civil registration systems cannot be attributed to neglect by international agencies. As early as 1953, the United Nations Statistical Office issued *Principles for a Vital Statistics System*, the first of several handbooks and reviews on the subject. More recently, the Health Metrics Network has been established to facilitate better health information at country, regional and global levels [16, 19]. Rather, the root problem is easy to identify: active systems of vital-event searching and recording in poor countries with under-resourced health systems are impossibly expensive to sustain at a national level [4, 16]. However, effective health information systems are indispensable for good management and planning in any kind of health service, and perhaps even more so when resources are particularly scarce.

Weakness in death registration is inextricably associated with weakness in birth registration. If someone dies without their birth being registered, it is less likely that their death will be registered. This is especially so if the deceased is a child. Even in settings where systems are in place, under-reporting or the concealment of events is a problem. In any data-capture system there are two key actors: the informers and the registrars or data collectors. In relation to reporting deaths, a number of factors may interfere with frequency and quality of reporting in less-developed settings, such as cultural values and perceptions associated with different categories of death, the social status of the deceased and the perceived benefits or penalties of reporting a death. For example, Buddhist and minority groups living in the FilaBavi surveillance area of Ha Tay province of Vietnam may tend to under-report infant and newborn deaths as such young individuals are not perceived to be fully human until their first birthday and have no significant relationships with other community members. Stigma and shame are strongly associated with certain diseases and causes of death, and thus this may also act as a disincentive for death reporting. Traditional beliefs and cultures relating to a human soul, death and spirit are considered to further influence perceptions, and thus, the reporting of deaths [20]. Add to this a lack of appreciation of the value of mortality reporting, legal regulations (if any) and processes involved in the reporting of deaths, as well as weak enforcement of legal sanctions and poor understanding of the consequences of death reporting on government benefits, a reluctance or delay in death reporting is not surprising.
Sentinel Surveillance and Demographic Surveillance Sites

Overview of DSS Methods

Over the past ten years, increasing emphasis has been placed on localised demographic surveillance systems (DSS) in developing countries as a substitute for a lack of more widespread health information systems. In simple terms, DSS is concerned with tracking the presence, demographics and health indicators of individuals in a defined study area [21, 22]. Individuals can enter and leave the surveillance population in a small set of well-defined ways, thus DSSs may be considered to be dynamic cohorts (Figure 1).

Whilst not direct substitutes for more widespread systems, it is believed that these initiatives may characterise the local situation, which may be more or less representative of the wider situation. Several DSSs have accumulated a substantial database over an extended period, whilst others have been established more recently and further systems are being planned and set up [23, 24].

The starting point in the design of DSS strategies is the selection of a demographic surveillance area (DSA). Surveillance systems are often set up around specific intervention studies and later converted to standing DSSs, thus the selection of DSAs will have been determined already. There are many examples of DSSs being established for demographic and health surveillance as the primary purpose, however (e.g. FilaBavi, Vietnam; Butajira Rural Health Programme (BRHP), Ethiopia). In such instances, the selection of DSAs is likely to be determined by logistical factors, such as the distance from managing and academic institutions, as well as scientific factors, such as trying to select a DSA that represents wider local or national diversity and population distributions. BRHP in Ethiopia is a prime example of a DSS established for the primary purpose of demographic and health surveillance. The selection of the Butajira area was determined to some extent by the fact that the area incorporates both highland and lowland, as well as urban and rural areas, and comprises of a mixture of ethnic and religious groups [22, 25, 26]. Whatever the genesis of the DSS, the DSA must be well defined.
Following selection of the surveillance area and, where necessary, selection of the surveillance population, the on-going task is then to maintain regular surveillance of vital events within the identified population, which is likely to require periodic household visits. The frequency of these update rounds depends on several factors and varies from monthly to annual in different DSSs. The longitudinal process of update rounds forms the core of a DSS, which, once implemented, can then be used as a platform on which to build other health-related surveillance and research activities [24].

The precise reasoning and justification for various aspects of DSS methodology, such as the frequency of update rounds, does not appear to have a strong evidence base and methods may often be determined by financial and logistical factors. The way in which differing surveillance approaches affect validity and comparability of data between sites is a concern and an area in which methods are not well established.

An individual’s name, date of birth, relationship to the head of the household, as well as location of the household are key parameters collected in the DSS. The vital events of births, deaths and migrations typically constitute the core events of interest in any DSS, but are often supplemented with social and economic correlates of population and health dynamics [27]. All DSS systems attempt to register all deaths within the surveillance population and the majority apply verbal autopsy (VA) methods to determine likely cause-specific mortality fractions (CSMFs). VA is the process of eliciting information about the circumstances of a death from family or friends of the recently deceased person in cases where medical certification is incomplete or absent [28, 29]. The method is based on the assumption that most causes of death can be distinguished by their signs and symptoms and that these can be recognised, recalled and reported accurately by lay respondents [28, 30]. The potential advantage of multiple visits at regular intervals to the same households is obvious; enumerators armed with a list of household members from the previous visit are well placed to detect additions (births/in-migration) and deletions (deaths/out-migrations) [16].

Data processing in DSSs is complex and various data models have been proposed [21, 31]. In general, it is desirable that the core data process is common between field sites to better facilitate cross-site collaborations and data sharing. The Household Registration System (HRS) is a software system that implements the demographic core, maintains a consistent record of significant demographic events that occur in a population under surveillance, generates registration tools to be used by field workers, and computes basic demographic rates [32]. Adaptations and variations of the HRS underpin data-process and database systems in most DSSs as well as a wide range of other population surveillance and research activities [21].
The INDEPTH Network

Following a workshop in Dar es Salaam, Tanzania, in 1998, an international network called INDEPTH (International Network of field sites with continuous Demographic Evaluation of Populations and Their Health in developing countries) was created. INDEPTH’s vision is to ‘be an international platform of sentinel demographic sites that provides health and demographic data and research to enable developing countries to set health priorities and policies based on longitudinal evidence’ [33]. Specifically, data generated in INDEPTH sites is intended to guide the cost-effective use of tools, interventions and systems to ensure and monitor progress towards national goals and to ‘harness the collective potential of the world’s community-based longitudinal demographic surveillance initiatives in resource-constrained countries to provide a better empirical understanding of health and social issues and to apply this understanding to alleviate the most severe health and social challenges’. The number of DSSs constituting the INDEPTH Network has risen from 17 sites in 13 different countries at its inception to a total of 38 separate sites in 19 different countries in 2007 [23]. Data from three INDEPTH member sites, FilaBavi (Vietnam), Butajira (Ethiopia) and Nouna (Burkina Faso), form the basis of this research.

Figure 2: Location of INDEPTH Demographic Surveillance Sites.
Source: www.indepth-network.org
Rationale

Approaches to conducting censuses and surveys have changed profoundly during the last fifty years [16]. As a result, surveillance methodologies vary widely between surveillance settings and seemingly ad hoc approaches and lack of standardisation make meaningful comparisons of data between sites and over time difficult. There exists widespread acceptance that more refined methods are needed if the outputs of sentinel sites are to be extrapolated and aggregated with realistic margins of error and validity. Researchers and surveillance personnel already overwhelmed with implementing and running surveillance systems should not be expected to carry out such methodological development on their own. Instead, an international collaborative approach is considered important in order to obtain an overview of pressing methodological issues in surveillance [4, 34].

Aims and Objectives

The growing recognition of the value of surveillance and its expansion into a separate scientific area within public health has not been accompanied by parallel growth in the literature about its principles and methods [1]. This thesis aims to address some of these methodological issues in an attempt to inform practice and improve the evidence base for mortality surveillance in resource-poor settings. Specifically, the objectives of this work are to:

• explore the extent to which choice of sampling method affects representativeness of 1% population sample data in relation to various demographic and health parameters important to measuring mortality (Paper I);
• empirically investigate the extent to which missed events and measurement errors affect overall health and demographic profiles in large, longitudinal surveillance datasets (Paper II);
• build on and develop a Bayesian probabilistic approach to determining cause of death from VA data and evaluate this method (Paper III);
• adapt the probabilistic approach to VA interpretation to the specific needs of maternal mortality measurement and evaluate this method (Paper IV);
• explore the possibility of using international mortality threshold levels in population surveillance settings for the prospective detection of humanitarian crises and discuss the ethical imperatives for utilising surveillance data for public health action (Paper V).

Through the evaluation and development of standardised, comparable methods and simple tools for cause-specific mortality estimates, it is hoped that this thesis will provide an evidence base for fit-for-purpose surveillance activities. Furthermore, it is hoped that this work will better inform the public health functioning and extrapolation of sentinel mortality data for local communities as well as for international bodies monitoring progress towards international health benchmarks, such as the Millennium Development Goals. Ultimately, it is hoped that this thesis will highlight the benefits of recording deaths and their causes with a view to making every death count.
BACKGROUND

Sampling

Experiences from the Sample Registration System in India and the Disease Surveillance Point System in China have shown that continuous mortality surveillance in a nationally representative sample of the population is feasible for monitoring mortality trends over time and differentials between subgroups [17, 35-37]. Ideally, a sample should be representative of the population from which it is drawn. In fact, it has been suggested that the resource-intensive active follow-up of DSS systems can only be justified if the population and results can be extrapolated meaningfully into the surrounding 100-fold population [4]. However, there remain no ‘best-practice’ guidelines as to which survey sampling methods give the most representative samples in relation to various demographic and health parameters in rural, developing-country settings in general, and in DSSs in particular. A wide variety of sampling procedures exist, not least with regards to their complexity [38-41].

The most commonly used survey sampling methods in population surveillance in less-developed countries include simple random sampling, proportionate to population size sampling, stratified sampling and multi-stage sampling. The most basic method is simple random sampling whereby a random selection of sampling units is taken until the target sample size is reached. This approach is ideal from a statistical perspective and gives every eligible sampling unit the same chance of being selected. However, true randomness is often difficult to achieve in practice and, in relation to household surveys, may be expensive to conduct as those sampled may be scattered over a wide area. Systematic random sampling is an alternative to simple random methods whereby a starting point is chosen at random from the list of all eligible units, and units for inclusion in the sample are then selected at regular intervals (i.e. every n<sup>th</sup> case). Strictly speaking, this alternative does not give every eligible unit an equal chance of being selected, however, in practice, systematic random sampling is usually accepted as being random. One must be wary of sampling bias when using systematic methods if, for example, every n<sup>th</sup> unit coincides with some hidden pattern within the population (e.g. hypothetically, every tenth household may always belong to the village chief and thus may have characteristics different to other households).

A more complex procedure of sampling with probability proportional to size (PPS) increases the probability of sampling more populous units, in an attempt to make any individual’s chance of being included in the sample similar, irrespective of the size of the unit in which they live [42]. PPS methods are often chosen over simple random methods when sampling frames of all eligible individuals are unavailable [39]. Furthermore, PPS methods are thought, from a logistical viewpoint, to be more efficient in large geographical areas. In many settings the size of sampling unit is unlikely to be known with great precision, thus sampling using probability proportional to estimated size (PPES) may be necessary [42].

Stratified sampling is used to ensure the fair representation of major groupings within an overall population, for example, urban and rural areas [42]. With this approach, the geographic area of interest is divided into mutually exclusive and exhaustive strata. If approximately 10% of the total population live in urban areas, for example, it may be desirable to ensure that approximately 10% of the sample drawn from the total population comes from urban areas through proportional
stratified sampling. It is important to note that, depending on the measurement objectives of the sample, true representation of the population does not always imply proportionate sampling between strata, however.

In DSS situations, a sample tends to be drawn at a local rather than national level, and multi-stage sampling is employed. Selecting samples in stages permits the sampler to isolate, in successive steps, the geographic locations where the survey operations (listing houses and administering interviews) will take place. Once the DSA has been selected (Stage 1), which is often referred to as the ‘cluster’ or ‘enumeration area’, it is necessary to create a sampling frame from which to select primary sampling units (PSUs) for inclusion in the survey (Stage 2). The sampling frame consists of a list of units of the population from which the sample will be drawn. A lack of existing sampling frames in areas where DSSs are likely to be established means that the sampling process must begin with substantial fieldwork in census and mapping activities to identify PSUs, which are typically households in DSSs but may be individuals or villages. The size of the area to be listed may be limited in resource-poor settings with obsolete samplings frames [42]. Clear definitions of what comprises sampling units are not always straightforward in less-developed countries. Typical houses in many rural African and Asian settings are temporary structures, and so attaching the label of ‘household’ to the structure rather than the individuals may be problematic (Images 1a and 1b). Furthermore, households may consist of a number of separate house structures that are built close to each other and are occupied by members of the same family. Some DSSs find it useful to define households as the group of people who eat together [43].

The most well-known examples of the use of multi-stage sampling on a large scale are the Multiple Indicator Cluster Surveys conducted by UNICEF [40] and USAID’s Demographic and Health Surveys (DHS) [44]. Most DHS samples use two-stage stratified designs involving the selection of area units in a single stage, normally with PPS, followed by listing and selection of households to yield around 30 female respondents per area unit. Typically, the surveys are based on large numbers of area units (about 300 units). Stratification of the area is usually explicit in terms of type of area (urban-rural) and location (region). Within each explicit stratum, the area units are arranged geographically and selected systematically [45]. The analysis of data from such cluster sampling techniques requires consideration of design effects whereby individuals within clusters may be more likely to share certain characteristics or disease patterns and so may not be representative of other clusters or the wider population [46].

Sampling issues are not only important for the establishment of the DSS, but also for nested surveys within the DSS setting. Sampling for such surveys may be more straightforward than for establishing the DSS itself, as a key benefit of DSS is the provision of a good sampling frame. Nevertheless, research activities, including clinical trials, are conducted within the platform of DSSs and thus knowledge about how the choice of sampling method can influence the generalisability of such work is of great importance. In relation to research and programme evaluation in the field of maternal health and safe motherhood, for example, several innovative sampling approaches and adaptations of traditional approaches, such as non-probability and convenience sampling, have been utilised [47-50].
For a truly representative sample, every variable of interest should have the same distribution within the sample as in the population from which the sample is drawn. Selection of appropriate sampling methods must therefore be influenced by the parameters one is trying to measure. In addition to counting the number of deaths themselves, mortality measurement requires the reliable measurement of basic population parameters such as age and gender distribution, which are needed to develop deeper understanding of causal pathways and potential intervention strategies. Other parameters known or likely to be risk factors for mortality are also important, such as socioeconomic indicators. Any such parameters are likely to have a wide distribution among any population, which itself should influence the choice of sampling method.

Sample size is a further important consideration in population surveillance and survey design and has important influences on the precision and cost of such operations. The average population size of INDEPTH member sites is between approximately 67,000 and 87,000, with a range from just 5,000 (Ouagadougou DSS, Burkina Faso) to 225,000 (Matlab, Bangladesh) [23]. There is a dearth of evidence, however, relating to the ideal size of a surveillance population and mortality surveillance systems to date have generally been determined by the size of a population within a given administrative area and by the specific objectives of the system. Such determining factors have been criticised for failing to take into account the number of deaths needed to yield sufficiently robust information on cause-specific mortality [17]. Mathematical formulae are available to calculate necessary sample sizes for acceptable degrees of precision [40, 42, 51], including methods for determining efficient sizes for sample-based mortality surveillance systems in situations where prior information on the cause composition of mortality is lacking [17, 52]. In relation to cause-specific mortality surveillance specifically, sufficient numbers of deaths for the rarest causes of interest are important [28]. Nevertheless, available resources and the specific data-collection approaches that will be utilised must also be taken into consideration when making a final decision on sample size.

It is clear therefore, that sampling necessarily increases uncertainty in survey and surveillance methods. As such, it is prudent to know a priori whether and to what extent this matters. It is likely that those faced with designing sample-based systems would benefit from systematic investigation and guidelines on these issues.
Image 1a: A typical urban house in Butajira, Ethiopia.

Image 1b: A typical rural house in Butajira, Ethiopia.
Data Quality

As in any measurement process, a certain amount of error is to be expected in population surveillance [53, 54]. Measurement errors may occur in a variety of ways: instrumental errors arising from imprecise instruments or questionnaire limitations; underlying variability leading to differences between replicate measurements taken at different time points; respondent errors arising through misunderstanding, faulty recall or reporting bias; observer error, including imprecision and mistakes; and data-processing errors during coding and data entry [55]. These errors can be broadly categorised as being systematic or random.

Systematic errors relate to the study design, methods and tools utilised and may be more common in certain demographic groups or regions. For example, systematically missing the deaths of infants due to excessively long periods between surveillance rounds combined with cultural reporting biases introduces systematic errors. Such differential bias (i.e. bias that does not affect everyone equally) is unacceptable if realistic and useful estimates of infant mortality rates are to be made. Random errors, on the other hand, occur independently of study design and methods used, and are unrelated to the value of other variables in the dataset. Key sources of random error are mismeasurement and human error. For example, incorrect coding of data, digit transposition or missed events are random errors if they are unrelated to the particular variable of interest and other characteristics of that particular case. Large degrees of errors in longitudinal surveillance are potentially a major problem that may invalidate the results of otherwise well-designed studies.

Measurement error often has both systematic and random components [56]. Problems that may result from errors occurring when measuring exposure or outcome variables generally relate to false population representation and univariate regression dilution bias, whereby estimated regression coefficients of single exposure-effect estimates may be biased towards the null value of no exposure effect, so that the magnitude of the association between the exposure and the outcome will tend to be underestimated – the association is said to be attenuated [53, 55-57]. Nevertheless, there are important exceptions where error can result in an increase in the regression coefficient, such as when there is error in the measurement of confounders resulting in residual confounding. The effect of data errors in multivariate analyses, with errors in a number of interrelated variables, can lead to more complex, serious distortions in the estimation of real exposure-outcome associations [54, 58]. All types of non-differential random measurement error reduce the chance of true significant associations being identified; in other words they reduce statistical power. However, the effects of random, but not systematic, errors are reduced with increased sample size [56].

A significant proportion of population surveillance operations and resources are dedicated to data quality-assurance mechanisms [59]. Key to quality control is a thorough appreciation of the data process and data management roles in assuring high quality data. The data process refers to the entire progression of data from collection in the field to data entry, storage and analysis. Providing researchers, DSS management personnel and field supervisors with at least a basic understanding of the database structure and functioning can help to ensure quality data [31]. A clear definition of what constitutes a vital event and the establishment of well-defined and unambiguous rules, such as which variables must remain constant over time (e.g. date of birth and sex), at the very outset of surveillance activities helps to prevent complex corrections at a later
stage. Such rules may be built into the database enabling automated monitoring of data quality, logic and consistency [32, 60]. It is important that all stakeholders are involved at every stage of designing the data process, as this can not only facilitate the process but also enhance feelings of ownership and understanding amongst stakeholders.

To avoid confusion at both the data-collection and data-entry stages, it is typical for each type of vital event being monitored by the surveillance system to have a separate, colour-coded form. Evidence suggests that the rate of errors in data collection and entry increases in relation to the length and complexity of the form [59]. Therefore, data-collection forms should be designed to make entries by the field worker as simple as possible with minimal writing and data entry required.

Recent developments in direct data capture using handheld computers or Personal Digital Assistants (PDAs) and Global Positioning Systems (GPS) present innovative approaches that may simplify data capture and enhance the quality of household and individual identification data. Although such approaches are relatively uncommon in routine field operations at present, there is growing interest in their use and several studies have demonstrated their usefulness for data capture, even in rural African settings with limited electricity supply and harsh environmental conditions [61-65]. Given that longitudinal data capture and many of the associated quality-control procedures require repeat visits to uniquely identified individuals in uniquely identified households, the importance of house, household and individual identification numbers cannot be overstated. Advances and increased affordability of technology such as PDAs with integrated GPS present exciting possibilities in this respect.

Drawing correct quantitative conclusions that can form the basis for public health intervention necessitates that the effects of measurement error are appreciated and accounted for [54]. Sensitivity of data to error, often termed ‘robustness’, is important in assessing the degree of uncertainty associated with surveillance outputs [53]. If measurement error must be adjusted for, the extent of error must be determined. If the extent of error is unknown, it may be estimated from a validation study in which the ‘true’ measurement is observed alongside the error-prone value, usually in a sub-sample of the population or from a replication study in which the error-prone measurement is made more than once in some or all subjects [58]. Rarely in epidemiology and population measurements, however, can a truly ‘gold standard’ measure be used for validation studies and the use of imperfect or ‘alloyed’ gold standards risks introducing more bias than they are correcting [66].

In DSSs, checks of completed surveillance tools for evident errors and omissions is a crucial aspect of quality control and measurement-error estimation, and is often performed at every level of field organisation with checks becoming more detailed as data progress through the system (Figure 3) [24]. Questionnaires with obvious errors or missing data that cannot be corrected by supervisors are returned to the field, while those that are free of errors proceed to data entry, which in some settings is performed twice to enhance data quality. Random duplicate household visits are often considered to be an additional important aspect of data-quality assurance. These activities are important for providing constructive feedback to fieldworkers with the aim of improving interview techniques, whilst taking repeated measurements enables estimation of error rates and thus correction factors. Of the 38 member sites of the INDEPTH network,
describe scheduled random re-visits of primary sampling units as a method of data quality control, with the percentage of households re-visited ranging from 2% (Agincourt DSS, South Africa) to between 5 and 10% (Nouna DSS, Burkina Faso). Several DSSs that perform re-visits do not specify the percentage of the total population revisited [23, 24]. Such variation between sites is perhaps a further reflection of the lack of standard approaches or best-practice guidelines relating to DSS methodologies in general.

Despite the best efforts of population surveillance operations, however, it is unlikely that measurement error can be completely eliminated. If the effect of measurement error is appreciable then it may be possible, though often difficult, to adjust for bias in the analysis. Various methods are available to correct measurement errors, the aim being to simulate true population profiles and exposure-outcome effects which would be observed if errors were eliminated [58, 67-69]. These techniques include regression calibration, linear imputation and Bayesian frameworks, which use results from validation or replication studies and are based on the assumption that errors are uncorrelated with true values [58, 67, 70]. Nevertheless, these methods are rarely used in epidemiological studies [71] and it is seldom possible to regain lost power using statistical fixes [56].

**Figure 3: Summary of DSS quality-control procedures**

- **Raw data**
- **5-20% rechecked**
- **Data entry into transaction files**
- **Corrected/Verified data**
- **Errors identified**
- **Data rejected from main database**
- **Back to field**
- **Database updating**
- **DATABASE**

---

19 describe scheduled random re-visits of primary sampling units as a method of data quality control, with the percentage of households re-visited ranging from 2% (Agincourt DSS, South Africa) to between 5 and 10% (Nouna DSS, Burkina Faso). Several DSSs that perform re-visits do not specify the percentage of the total population revisited [23, 24]. Such variation between sites is perhaps a further reflection of the lack of standard approaches or best-practice guidelines relating to DSS methodologies in general.

Despite the best efforts of population surveillance operations, however, it is unlikely that measurement error can be completely eliminated. If the effect of measurement error is appreciable then it may be possible, though often difficult, to adjust for bias in the analysis. Various methods are available to correct measurement errors, the aim being to simulate true population profiles and exposure-outcome effects which would be observed if errors were eliminated [58, 67-69]. These techniques include regression calibration, linear imputation and Bayesian frameworks, which use results from validation or replication studies and are based on the assumption that errors are uncorrelated with true values [58, 67, 70]. Nevertheless, these methods are rarely used in epidemiological studies [71] and it is seldom possible to regain lost power using statistical fixes [56].

**Figure 3: Summary of DSS quality-control procedures**

- Raw data
- 5-20% rechecked
- Data entry into transaction files
- Corrected/Verified data
- Errors identified
- Data rejected from main database
- Back to field
- Database updating
- DATABASE
Consistent and reliable cause-of-death data constitute a crucial and major resource for health planning and prioritisation [11]. The idea of assessing causes of death by retrospective interview is as old as medical statistics, with so-called ‘death searchers’ visiting the houses of people who had died in seventeenth-century London to enquire about the death. Modern systems of death registration have replaced such activities in Western countries, but in less-developed countries there is still a need for such investigations into causes of death [72]. Pioneer projects in Asia and Africa in the 1950s and ‘60s used systematic interviews by well-trained physicians to assess causes of death. Workers at the Narangwal project in India labelled this technique ‘verbal autopsy’ and the method has since spread and developed, particularly during the 1970s when the WHO suggested lay-reporting of information about health issues by people with no medical background [72, 73].

It is now rare for physicians to conduct the systematic VA interviews themselves, rather, trained fieldworkers interview bereaved relatives using a questionnaire to elicit information on symptoms experienced by the deceased before death. Questionnaire-based VA methods aim to record objectively all information that may later be used to derive probable cause of death. There has been a proliferation of interest, as well as research and development, in all aspects of the VA process by numerous institutions, including individual DSSs and the WHO. This work has included developments in data-collection systems, questionnaire format, application of VA to different age groups, analysis and interpretation of the resulting information and coding and tabulation of causes of death according to international standards [74]. However, being largely research driven, there has been little coordination between activities and apparently no overriding concern to ensure comparability of datasets between countries and over time. This has resulted in a failure to reach consensus on what to cover in the interview, how questions should be asked and how the data should be analysed and reported. In turn, this has led to the development of numerous different VA data-collection tools, with eighteen different VA questionnaires being used by various DSSs [75]. Subsequent concern over comprehensiveness, validity and reliability of different tools is not surprising. Nevertheless, efforts by the INDEPTH network, as well as the WHO, have outlined the key characteristics of suitable VA data-collection instruments [43, 76]. Typically these include both an open-ended section for recording a verbatim account of symptoms and circumstances leading to death, and a closed section with filter questions on symptoms and signs of disease. The open-ended questions require respondents to recall and volunteer specific information, whereas closed questions require recognition, with more information likely to be recognised than recalled [75]. In addition, it is common for questionnaires to have detailed questions regarding use of health services, lifestyle and disease risk factors [75, 77]. Important considerations in VA data-collection tools are cultural sensitivity and awareness of reporting bias due to concepts of health and illness that are often not transferable between cultures.

Gathered VA data need to be interpreted to derive probable causes of death. There are several options for this, which can be broadly categorised as those using and not using pre-defined diagnostic criteria. The most commonly used method is physician review (PR) of the information without pre-defined diagnostic criteria. This typically involves independent review of the data by physicians. Normally, two local physicians trained in VA coding review each case, trying to reach
consensus on a single cause of death. Where no consensus is reached, a third physician, aware of the opinions of the other physicians, may also review the data. Normally this results in a single probable cause of death even if a death is likely to have been due to multiple causes [75]. If all three physicians disagree, the cause is usually listed as ‘indeterminate’.

Physicians are generally considered to be better able to appreciate the nuances of individual cases, and to be a valid method of cause-of-death diagnoses in various VA studies [28, 30, 78-82]. Physician agreement on diagnoses from VA data have been shown to be good in some settings [29], although this may simply reflect physicians’ prior perceptions of local epidemiology. In other settings, inter-observer agreement has been shown to be lower. In Bangladesh, for example, one physician attributed 41% of all maternal deaths to direct obstetric causes, while another group determined the proportion as 51% [83]. Such discrepancies give rise to concerns about inter- and intra-observer reliability, which can be misleading and preclude comparisons of cause-specific mortality between regions and over time, where different physicians and their methods of interpreting evidence may differ [84]. This lack of standardisation has been tackled with efforts culminating in the development of various algorithmic approaches based on the concept of distilling the process of PR into standardised rules [85].

Algorithms map diagnostic criteria in order to provide a systematic means of deriving cause of death from VA [75]. There are two types of algorithmic approaches to VA interpretation: expert algorithms and data-derived algorithms. Expert algorithms are pre-defined diagnostic criteria agreed by a panel of physicians. Data-derived algorithms derive probable causes of death according to various analytical techniques, including linear and discriminate regression, probability-density estimation and decision-tree and rule-based methods.

Diagnostic algorithm-based cause-of-death determination may be less accurate than PR, but has the advantage of being transparent and repeatable. Nevertheless, algorithmic procedures make it impossible to consider parallel possibilities of causes of death along the lines of classic clinical differential diagnoses, and their consistency is dependant, not only on the consistency of diagnostic criteria, but also on the consistency of signs and symptoms reported by respondents. Most DSSs do not currently employ diagnostic criteria for deriving causes of death.

Emphasis is often placed on the importance of open-ended, free-text information collected using VA questionnaires [75, 86]. This frequently includes verbatim accounts from respondents, which probably fit better with physicians’ customary approaches to diagnosis than a series of closed questions. The omission of open-ended information from most algorithmic approaches to VA interpretation has been criticised, which has further hindered their acceptance and caused concern over validity.

Multiple causes of death for individuals are a further important consideration when dealing with cause-specific mortality data. When there are multiple causes, the one with the greatest public health significance may be lost. International Classification of Disease (ICD) coding and international rules for the hierarchy of importance when choosing immediate and underlying causes of death attempt to tackle this problem but may not be suitable for all surveillance purposes. Determining multiple, rather than single, causes of death for any particular case is likely to more accurately reflect the interaction of different diseases that lead to death and give a more complete representation of the burden of different diseases at the population level. Co-morbid
BACKGROUND

conditions, which are particularly common in children and the elderly, may contribute to one’s death equally and therefore to count only one condition would distort mortality estimates and underestimate potential gains from health interventions. Nevertheless, distinguishing underlying and contributory causes of death with VA, as required for ICD reporting of mortality statistics, is difficult.

The International Statistical Classification of Diseases and Related Health Problems, tenth revision (ICD-10), is the mandatory level of coding for international reporting to the WHO mortality database. ICD-10 comprises 21 chapters and 2,046 categories of disease, syndromes, external causes and consequences of external causes [87]. The idea is that every cause-of-death category can be diagnosed by clinical judgement and, where necessary, laboratory tests. Such precision is impossible using VA methods and so mortality surveillance systems using VA that wish to code according to ICD-10 usually use an abbreviated list of cause-of-death diagnoses. As with other aspects of VA methodologies, the structure of such cause-of-death lists vary substantially in the field [75].

Rigorous validation of VA procedures is needed in the settings in which they will be used in order to understand the operational characteristics of VA in the population under study and to identify misclassification patterns, which may then be corrected [88]. What is often termed ‘validation of VA’ includes multiple components (validity and standardisation of VA instruments and interview, validity of VA interpretation(s), validity of arbitration between various interpreters and multiple validity issues around candidate gold standards such as medical record assessments) and discussions of VA validity typically focus on sensitivity, specificity and positive predictive values (PPVs) derived by comparing VA diagnoses with those of a reference diagnosis. In general, two types of reference gold standards are used for validating VA tools: health-facility-based diagnoses and community-based PR diagnoses [28, 75, 89]. The limitations of PR have already been highlighted and, whilst facility-based validations enable comparison of VA findings with a comparatively highly accurate medical diagnosis of cause of death, such studies are subject to selection and information bias and do not represent the populations for whom VA is intended, most of whom die without medical attention.

True validation of VA methods is therefore difficult. Whilst the limitations of PR and hospital records as gold standards are widely acknowledged [75, 89] and attempts have been made to adjust for imperfect gold standards [79, 88, 90], their preferential use continues and there is apparent unwillingness to contemplate that there is in fact no gold standard for cause-of-death determination. This is most obviously true in settings where only a very small and unrepresentative proportion of deaths occur in the care of physicians. It seems inappropriate therefore to discuss such methods in terms of sensitivity, specificity and PPV as this assumes that the referent diagnosis gives the right answer. Evaluations of VA methods should instead be considered in terms of comparability, reliability and adequacy of purpose, avoiding reference to sensitivity, specificity or PPVs, which would imply inherent superiority of referent methods.
Methodological Developments in VA Interpretation

Bayesian probability models, originated in principle by Thomas Bayes in 1763, have been extensively explored in the context of decision support systems in the fields of medicine and law and have been shown in many instances to be relatively effective [91-93]. Application of Bayes’ theorem for VA interpretation was developed and evaluated using VA data from FilaBavi DSS, Vietnam [94].

The method, named InterVA (Interpreting Verbal Autopsy), seeks to define the probability of a cause (C) given the presence of a particular indicator, sign or symptom (I), represented as P(C|I), and calculated by:

\[
P(C|I) = \frac{P(I|C) \times P(C)}{P(I|C) \times P(C) + P(I|!C) \times P(!C)}
\]

where P(I|C) is the probability of the cause not being cause (C). Therefore, VA data, representing a set of indicators I₁…In, can be used to calculate the probability of a set of causes C₁…Cₘ. The method requires that each indicator (I) and each cause (C) has an associated probability of occurrence at the population level, i.e. the probability that each indicator and each cause occurring in any death at the population level, as well as a matrix (n x m) of probabilities. For example, if the approximate probability that fever occurs in any death at the population level is 20% (P(I) = 0.2), the probability of death from malaria at the population level is 5% (P(C) = 0.05), and finally the probability of someone who has died from malaria having suffered a fever is 80% (P(I|C) = 0.8), then it is possible to calculate the probability that anyone who suffered from a fever before death died of malaria (P(C|I)) i.e. 17%. Based on this principle, a set of P(I₁…In), P(C₁…Cₘ) and the matrix P((I₁…In)|(C₁…Cₘ)) can be used to determine the probability of a range of causes of death for a range of indicators.

A simple computer programme provides a user interface for InterVA into which indicators for a particular case can be entered, leading to an output of the most likely causes and associated probabilities. For each case, the three most likely causes are listed. The programme also estimates a certainty factor for each case [95].

Although deriving a set of realistic probabilities may seem a difficult task, work suggests that a high degree of precision is not necessary to build a workable model. Probabilities used in the preliminary InterVA model were estimates based on the researchers’ accumulated personal experience, without any attempt to validate or establish internal consistency between estimated values, and yet gave satisfactory results in comparison with local reviewing physicians in approximately 84% of cases [94]. This early evaluation demonstrates that such an approach has the potential to improve the performance of expert algorithms whilst overcoming limitations of reliability and comparability of PR, however, refinement and further evaluation of the method using data from a variety of sources was deemed necessary, the process of which forms a substantial part of this thesis.

King and Lu (2008) [96] have developed a sophisticated method for directly estimating cause-specific mortality fractions (CSMFs) without individual cause-of-death attribution. Their method resolves the problem of generalising VA analysis to the population based on test proper-
ties quantified in health facility validation studies. Combining this approach with the InterVA method, Murray et al. (2007) [85] propose and have attempted to validate a new approach called the Symptom Pattern (SP) method. This method uses two sources of VA data, one for which the ‘true’ cause of death is known, which need not be representative of the population of interest (e.g. hospital-based data), and one undiagnosed VA dataset that is representative of the population of interest. From the hospital-based data the probabilities of reporting each symptom given a known, true cause of death can be calculated. These ‘symptom properties’ then allow estimates of population-level CSMFs. In turn, the CSMFs can be used as an input in assigning a cause of death to each individual VA case, which can then feedback and refine the population-level CSMF estimates [85]. This approach is a useful development in moving away from PR towards more cost-effective and consistent approaches to VA interpretation. However, the SP approach requires large databases of global symptom-level sensitivities and thus only addresses gaps in cause-specific mortality measurement at the global level, with other gaps remaining at various other levels for different end users, each requiring their own particular solution [11].

Maternal Mortality

The writing of this thesis coincides with the twentieth anniversary of the Safe Motherhood movement and the midpoint towards Millennium Development Goal 5 (MDG-5), which calls for a 75% reduction in maternal deaths by 2015. These landmarks bring renewed attention to what is essentially only a problem for the world’s poor [6] and a shameful indicator of inequality between rich and poor countries. Of the estimated total of over 500,000 maternal deaths worldwide, 99% occur in developing regions of the world, and more than half occur in the sub-Saharan African region alone [97, 98]. A 1-in-7 lifetime risk of dying from a maternal cause in Niger compared to a 1-in-48,000 risk in the Republic of Ireland demonstrates that maternal deaths need not happen given sufficient knowledge, infrastructure and, ultimately, political will to prevent them.

The ICD-10 definition of a maternal death is ‘the death of a woman while pregnant or within 42 days of termination of pregnancy, irrespective of the duration and site of pregnancy, from any cause related to or aggravated by the pregnancy or its management but not from accidental or incidental causes’ [87]. Direct obstetric deaths are those resulting from complications specific to pregnancy and childbirth. Indirect obstetric deaths are those resulting from previous existing disease, or diseases that developed during pregnancy, and are not due to direct obstetric causes but are aggravated by the physiological effects of pregnancy. The concept of pregnancy-related death included in ICD-10 incorporates any death during pregnancy, childbirth or the postpartum period even if it is due to accidental or incidental causes [97].

Contradictory claims that there is enough information to act on and concurrent calls for more information in relation to maternal health heighten the need for fit-for-purpose data on maternal mortality [99]. Monitoring maternal mortality and progress towards MDG-5 with sufficient precision and reliability, however, is a notoriously difficult methodological challenge, particularly in high mortality settings with weak health information systems [100-102]. Where the majority or a vast number of deliveries occur at home without ever accessing health services, and where vital registration systems with cause-of-death attribution are inadequate, the accurate identification of maternal deaths is particularly challenging. Even if such a death is recorded, the woman’s
Image 2: Mother and child in rural Burkina Faso.
Photo credit: Axel Emmelin, 2005.
pregnancy status may not be known and the death may therefore not be reported as a maternal death. Even in more developed regions of the world, where routine registration of deaths is in place, maternal deaths may be under-reported, and identification of the true numbers of maternal deaths may require additional investigation into the cause of death [97, 103].

Maternal mortality measurement thus requires community-based identification of deaths among all women of reproductive age, followed by determination of cause using VA, in order to identify maternal deaths [104, 105]. Methodological limitations in interpreting VA data in general are heightened by specific limitations in relation to identifying pregnancy-related deaths. Misclassification of causes of reproductive-aged female deaths is common and VA may fail to correctly identify indirect causes of maternal deaths and certain groups of direct obstetric causes, such as those occurring early in pregnancy (e.g. ectopic or abortion-related deaths) [82, 83, 97]. Such limitations can preclude monitoring and rigorous evaluation of the effectiveness of safe-motherhood intervention strategies. In order to tackle the problem of maternal mortality, and move closer to achieving MDG-5, therefore, there is an urgent need for the development and evaluation of new approaches to measuring cause-specific mortality among women of reproductive age.

Data for Action

“Declare the past, diagnose the present, foretell the future; practice these acts.
As to diseases, make a habit of two things — to help, or at least to do no harm”

Hippocrates Epidemics, Bk. I, Sect. XI.

In common parlance, ‘surveillance’ is often considered to be the watch or guard kept over individuals or populations. This concept implies a preventative role in line with the key nature of public health as the science and art of prevention. Population surveillance sites in developing countries are unique in their role of gathering longitudinal data on key health and demographic indicators. The proliferation of DSSs in recent years and subsequent collection of prospective data has been of immeasurable value to health researchers. Mortality measurements in particular are the most specific indicators of the health status of populations that can be used to monitor trends in disease [27]. However, a major challenge facing population surveillance activities in general, and DSSs in particular, is the accumulation of unanalysed data. All too often the period from data capture in the field to analysis, publication and use for informing public health action is very long. Even when data are processed efficiently, they are rarely made widely available or have an immediate effect on the lives of the surveillance population. That the value of data lies in their use, not in their collection does not always seem to be appreciated by surveillance systems, often burdened with tight budgets that hinder rapid local analyses without the need for complex equipment or expertise [14, 27].

Indeed, as a branch of medicine and medicine’s duty to help, it may be argued that those responsible for health and demographic surveillance in settings lacking other means of population measurement have an ethical obligation to act as the eyes and ears of health authorities and disseminate the information for action. In other words, to justify the intrusion of surveillance, the collected information must have a demonstrated utility and public health professionals should act in a timely manner on the information they have and expeditiously make important informa-
tion available to the public [106]. Efforts are needed to stimulate the debate and development of simple procedures and clearly defined surveillance responsibilities so that detection of entirely preventable morbidity and loss of life leads to public health action.

In relation to humanitarian crises, preparing for the health problems experienced by large populations displaced by natural or man-made disasters is among the greatest challenges facing public health officials today [107]. Traditional epidemiologic and surveillance techniques have previously been applied to the early detection and crisis management of affected communities [108]. Early warning systems of humanitarian crises might help avert substantial increases in mortality and morbidity and may prevent major population movements, thereby allowing the provision of assistance in situ without disruption to traditional social structures and lifestyle patterns [107].

However, the longstanding lack of vital-event data, coupled with already scarce resources in the settings most at risk of humanitarian crises, has often resulted in the late detection and inadequate response to such events [109, 110]. Any available data in resource-poor settings are generally weak and outdated, often aggregated for relatively high levels of administration, methodologically ad hoc and frequently interpreted out of context by external analysts [108]. These factors preclude timely detection of important changes in key indicators such as mortality rate.

The Centres for Disease Control and Prevention’s (CDC) humanitarian crisis threshold is 1 death per 10,000 population per day [107]. Recent retrospective inspection of the BRHP data revealed mortality peaks in 1998 and 1999, well in excess of the normally observed year-to-year variation and exceeding the CDC crisis threshold. Further investigation and enquiry revealed that these peaks related to a measles epidemic in 1998 and unseasonal rainfall and consequent widespread food insecurity in 1999 [110, 111]. This crisis was missed by BRHP at the time and Butajira district did not convey the impression of a population in crisis, resulting in no substantial humanitarian relief efforts being triggered [111]. This raises important practical and ethical questions relating to data processing routines in population surveillance in developing countries that are explored by the final paper in this thesis.
MATERIALS AND METHODS

Empirical Investigations Using Existing Data

Empirical research methods are a class of method in which data are collected in order to answer particular questions. This thesis applies a priori theories relating to surveillance and survey methodologies to data collected for other purposes with a view to testing and refining methods. An alternative approach would be to test different methods using experimental approaches, however, such an approach is far more resource-intensive and disruptive to routine surveillance activities. Nevertheless, more formal experimental approaches should not be ruled out for future investigations into surveillance methods as certain aspects of surveillance, such as the appropriate time interval between surveillance rounds, may only be satisfactorily addressed using experimental methods in the field.

Formal statistical methods can only be used as a theoretical framework for methodological investigations where there is adequate prior knowledge of statistics at the population level. This is a problem for DSS and other population surveillance design in resource-poor settings since the inherent need for DSS and surveillance from first principles often implies that there is little existing data with which to demonstrate that a particular method is most appropriate or that any given demographic surveillance area could be representative of the wider population. Therefore, two of the papers included in this thesis (I and II) define referent data as correct, gold standard data with which data resulting from modelling approaches can be compared. Such an approach provides the luxury of not having to be overly concerned with the real quality of data values in the reference dataset, as the studies are not assessing these values but rather how values in general are affected by methodological aspects of their collection. Nevertheless, the data used in this thesis is considered to be of high quality. In support of this, the gold standard population pyramid in Paper II, Figure 1, gives a plausible representation of the BRHP study population as one with a high birth rate and at the early stages of a demographic transition. The notable shift in male:female ratio in the mid-twenties and mid-thirties age group also suggests a good representation of population composition as it reflects the social and political instability of Ethiopia during the 1980s and early ’90s [112].

Data Sources

This thesis utilises data from three demographic surveillance sites in three countries, as well as data from a large population census and health survey conducted in rural Burkina Faso. The DSS data come from BRHP (Ethiopia), FilaBavi (Vietnam) and Nouna (Burkina Faso).

BRHP DSS

The Butajira Rural Health Programme (BRHP) in Ethiopia has maintained a programme of epidemiological surveillance in the Butajira District, some 130 kilometres south of the capital, Addis Ababa, since 1987. Continuous community-based surveillance of an open cohort population sample is conducted through quarterly household surveys relating to births, deaths and migrations, as well as socioeconomics and living conditions [24, 25]. Initially, the sampled communities
were selected from the entire district using PPS methods and covered approximately 10% of the district [22]. In recent years, VA procedures have been introduced, giving more detail on the circumstances and causes of death [113]. Given its wealth of data, BRHP acts as a platform for more specific epidemiological and public health studies.

Figure 4: Location of BRHP, Ethiopia
Source: www.indepth-network.org

Image 3: Rural village in BRHP surveillance area, Ethiopia.
FilaBavi DSS

The Epidemiological Field Laboratory of Bavi (FilaBavi) is in Bavi district of Ha Tay province in northern Vietnam, approximately 60 kilometres west of the capital Hanoi. Bavi district was selected for the epidemiological field laboratory on account of the varied geographical, socioeconomic and health characteristics of the area, which are considered typical of northern Vietnam. A baseline household survey of a sample of villages within the district (using PPS methods) was conducted in 1999, and quarterly demographic and health surveillance of vital events and a complete re-census every two years characterises the surveillance activities of FilaBavi since then [24, 114].

Figure 5: Location of FilaBavi DSS, Vietnam
Source: As Figure 4.

Image 4: Rural agricultural life in Bavi district, Vietnam.
Nouna DSS

The Centre de Recherche en Santé de Nouna (CRSN) is located in the Nouna Health District in the north-west of Burkina Faso, 300 kilometres from the capital, Ouagadougou. CRSN (Nouna DSS) has conducted regular population censuses since 1992, and further comprises quarterly vital-event and demographic surveillance [24, 115, 116]. Routine VA interviews are conducted for all deaths.

Figure 6: Location of CRSN Nouna DSS, Burkina Faso
Source: As Figure 4.

Image 5a: Surveillance staff outside field office in Nouna, Burkina Faso.
Immpact Census and Health Survey, Burkina Faso

In 2006, the Immpact (Initiative for Maternal Mortality Programme Assessment) initiative conducted a thorough census and household survey in two districts of Burkina Faso as part of a wider evaluation of safe-motherhood interventions. The survey aimed to cover 100% of the population in Koulpélogo and Tapoa provinces, in south-eastern Burkina Faso. Ouargaye town, the provincial centre of Koulpélogo, is some 230 kilometres from Ouagadougou, and Diapaga town, the provincial centre of Tapoa, some 370 kilometres.

A household census was conducted throughout the two provinces, registering a total of 86,378 households and 515,298 people. The census also included a retrospective survey of adult and perinatal deaths occurring since January 2001. A technology-based strategy using PDAs was employed for the direct capture of data for the census and survey.

Figure 7: Satellite image of Immpact census and household survey areas, Burkina Faso

Data Acquisition

DSS Data
Existing data from each of the data sources were accessed in collaboration with local, on-site research teams and data managers. For example, original completed VA questionnaires from Nouna DSS were accessed in the field office in Nouna, and all relevant information pertaining to the circumstances, signs and symptoms relating to the deaths of adult females were extracted and entered into an electronic database over a two-week period.

Image 6: Verbal autopsy data archive in Nouna DSS, Burkina Faso.

Immpact Census and Health Survey
Data capture for the large Immpact census and household surveys required careful planning and execution of field survey methods. Data-collection tools were designed by researchers at Immpact and transferred onto PDAs for direct data capture. Fieldworkers collected data over a five-month period and the specific methods employed are described elsewhere [117]. Following data processing, the data needed for this thesis were accessed with permission from Immpact and the local research teams in Burkina Faso.

Direct data capture using PDAs in the census and survey enabled all data to be collected from in just four months and streamlined the data-management process to a great extent [61]. This approach represents a methodological development in its own right that required careful consideration at the planning stages. The lack of reliable mains electricity in the study setting, for example, required innovative thinking with regards to keeping the PDAs sufficiently charged. Nevertheless, the need to deliver a large dataset within a short timeframe and to avoid lengthy data entry after the survey meant that the risks of using of a direct-data capture approach were justified. The use of PDAs with integrated GPS capacity was a further advantage.
Sampling

An empirical approach to the evaluation of various survey sampling methods applied using the Immpact household census data from Burkina Faso. From the large number of parameters captured in the census, gender (proportion of males), age (proportion under 5 years), education (proportion of population who have completed secondary-level education or above), economics (proportion in the lowest wealth quintile), number of maternal deaths that occurred in the last 5 years, and number of adult female residents were selected.

Samples were drawn at two administrative levels by each of seven sampling methods that are commonly used in population surveys in rural Africa and other resource-poor settings. Using SPSS software, the samples were generated from the entire dataset as outlined in Table 1.

Modelling of the sampling strategies was carried out by drawing 20 repeated 1% population samples using each of the methods outlined in Table 1. Each of the parameter based on the resulting 280 was then compared with the gold standard of the complete census. The concept of accuracy within the samples, i.e. the extent to which a particular sample represents the whole population, was evaluated according to whether the mean of the 20 samples from each sampling approach lay within a particular tolerance of the unsampled value.
### Table 1: Sampling methods, modelling technique and example of use in field surveys

<table>
<thead>
<tr>
<th>Sampling Method</th>
<th>Technique</th>
<th>Example situation in field surveys</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple Random</td>
<td>Step 1: Assign a random number to each sampling unit</td>
<td>Cross-sectional surveys within DSS settings</td>
</tr>
<tr>
<td></td>
<td>Step 2: Sort sampling units by their random number</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Step 3: Select sampling units in ascending order of random numbers until target population is reached</td>
<td></td>
</tr>
<tr>
<td>Probability Proportional to Size</td>
<td>Step 1: Assign a random number to each sampling unit</td>
<td>Cross-sectional surveys within DSS settings</td>
</tr>
<tr>
<td></td>
<td>Step 2: Multiply the population of each sampling unit by the random number</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Step 3: Sort sampling units on the number generated in Step 2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Step 4: Select sampling units in descending order of number generated in until target population is reached</td>
<td></td>
</tr>
<tr>
<td>Proportional Stratified Sampling</td>
<td>Step 1: Determine the proportion of sampling units in each strata</td>
<td>Cross-sectional surveys within DSS settings or establishing a DSS</td>
</tr>
<tr>
<td></td>
<td>Step 2: Assign random to each sampling unit</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Step 3: Select sampling units using simple random methods from each strata until based on ratio determined in Step 1</td>
<td></td>
</tr>
<tr>
<td>Multi-stage Sampling</td>
<td>Step 1: Randomly select geographical area for sampling</td>
<td>Establishing a DSS</td>
</tr>
<tr>
<td>(Stage 1 random; Stage 2 random)</td>
<td>Step 2: Assign a random number to each sampling unit in the selected area</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Step 3: Sort sampling units by their random number</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Step 4: Select sampling units in ascending order of random number until target population is reached</td>
<td></td>
</tr>
<tr>
<td>Multi-stage Sampling</td>
<td>Step 1: Randomly select geographical area</td>
<td></td>
</tr>
<tr>
<td>(Stage 1 random; Stage 2 PPS)</td>
<td>Step 2: Assign a random number to each sampling unit in the selected area</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Step 3: Multiply the population of each sampling unit by the random number</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Step 4: Sort sampling units on the number generated in Step 3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Step 5: Select sampling units in descending order of number generated in Step 3 until target population is reached</td>
<td></td>
</tr>
</tbody>
</table>

*Continued on page 32*


**Table 1 contd.**

<table>
<thead>
<tr>
<th>Sampling Method</th>
<th>Technique</th>
<th>Example situation in field surveys</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geographically Dispersed</td>
<td>Step 1: Randomly select two geographical areas</td>
<td></td>
</tr>
<tr>
<td>(Stage 1 random; Stage 2 random; Stage 3 random)</td>
<td>Step 2: Assign a random number to each sampling unit in each of the selected areas</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Step 3: Sort sampling units by their random number</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Step 4: Select sampling units in ascending order of random number until 50% of target population is selected from each geographical area</td>
<td></td>
</tr>
<tr>
<td>Geographically Dispersed</td>
<td>Step 1: Randomly select two geographical areas</td>
<td></td>
</tr>
<tr>
<td>(Stage 1 random; Stage 2 random; Stage 3 PPS)</td>
<td>Step 2: Assign a random number to each sampling unit in each of the selected areas</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Step 3: Multiply the population of each sampling unit by the random number</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Step 4: Sort sampling units on the number generated in Step 3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Step 5: Select sampling units in descending order of number generated in Step 3 until 50% target population is selected from each geographical area</td>
<td></td>
</tr>
</tbody>
</table>

**Data Quality**

Defining data collected between 1st January 1987 and 31st December 1996 in BRHP, a 10-year dataset covering approximately 336,000 person-years [118], as a gold standard dataset, simple programmes were written using Microsoft Visual FoxPro software to simulate versions of the 10-year dataset containing random errors and missing data in key parameters. Table 2 describes the parameters investigated, reasons for their selection, errors introduced, an explanation of how the errors were simulated, and a summary of how the simulated errors relate to practical survey methods in the field.

There are no standard procedures for this type of investigation and therefore the parameter modification described in Table 2 is arbitrary. Nevertheless, the extent of parameter modification in this study was influenced by probable random error margins in routine DSS procedures, which are unlikely to exceed 10% in most instances, as well as by issues of presenting the results – in our experiment, parameter modification of less than 10% failed to show any substantial differences in population representation and mortality patterns. Whilst in practical terms missing 10% or 20% of deaths, for example, may be considered a systematic failure, these errors have been distributed randomly in the dataset – in other words, having missing or incorrect data is unrelated to any other factor. Age modification may have been achieved by altering age in absolute terms, e.g. by adding 10 years to the randomly selected cases. In reality, however, it is likely that the degree of age misreporting in population surveys is proportional to one’s age, i.e. older adults may misreport by a greater number of years than children, for example.
**Table 2: Parameter conceptualisation, error simulation and relation to practical field methods**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Reason for Selection</th>
<th>Error</th>
<th>Simulation of Error</th>
<th>Relation to Practical Field Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>Male or female</td>
<td>Important health and demographic parameter, necessary for age- and sex-specific mortality estimates</td>
<td>Randomisation in 10% and 20% of cases</td>
<td>A random selection of 10% or 20% of cases was randomly assigned data values for sex (male or female)</td>
<td>Such an error may occur during data entry and coding</td>
</tr>
<tr>
<td>Age</td>
<td>In years, calculated from date of birth</td>
<td>Increase by 10% and 20% of ‘true’ value in 10% of cases</td>
<td>A random selection of 10% or 20% of cases</td>
<td>Age value was increased by 10% or 20% in a randomly selected 10% of cases</td>
<td>Age misreporting is likely to cluster in certain portions of the population, such as the elderly</td>
</tr>
<tr>
<td>Death</td>
<td>Whether or not an individual in the surveillance population died during the 10-year period</td>
<td>Death data changed to missing value for 10% and 20% of cases</td>
<td>A randomly selected 10% or 20% of cases</td>
<td>A randomly selected 10% or 20% of cases recorded as having died had information about their death removed, thereby simulating missed events</td>
<td>Deaths are most likely to be missed for certain age groups, such as the very young</td>
</tr>
<tr>
<td>Literacy</td>
<td>Literate or illiterate</td>
<td>Measure of the number of deaths within the population during the specified time</td>
<td>Data on literacy removed for 10% of cases</td>
<td>Information on whether an individual was literate was removed in a randomly selected 10% of cases</td>
<td>Such errors may occur as a result of reporting bias or misclassification error in data capture and data entry</td>
</tr>
<tr>
<td>Roof material</td>
<td>Material used in household roof structure (thatched, corrugated or unknown)</td>
<td>Material used in household roof structure (thatched, corrugated or unknown)</td>
<td>Roof material randomised for 10% of cases</td>
<td>The values ‘corrugated’ or ‘thatched’ were randomly assigned to a random selection of 10% of the cases</td>
<td>Such errors may occur as a result of misclassification error in data entry</td>
</tr>
</tbody>
</table>
Population age and sex composition and all-cause mortality rates were calculated for each dataset and results were compared with the gold standard of the original 10-year data to determine the extent to which the introduction of errors affected the data’s ability to represent the surveillance population. Multivariate Poisson regression models of mortality rate ratios for the 10-year dataset as well as each of the simulated error datasets were also created and compared.

**Cause-of-death Determination**

The Delphi method is an approach used to gain consensus among a panel of experts in order to address a lack of agreement or incomplete state of knowledge [119-121]. The technique utilises the knowledge and values among experts as a basis for decision-making processes and attempts to synthesise qualified opinions [57]. A Delphi-style method was adopted to develop consensus on probabilities of different causes of death occurring at the population level and probabilities of specific signs and symptoms presenting themselves at the population level and in specific causes of death. The technique was also utilised to develop consensus on key conceptual issues of cause-of-death classification and VA usage, both in relation to all deaths (InterVA) and specifically for deaths of reproductive-age women (InterVA-M).

In both instances the group comprised five physicians with extensive clinical experience in resource-poor settings. For refinement of the all-cause InterVA method, the expert panel represented a range of important disciplines of medicine: surgery; maternal and reproductive health; paediatrics; and internal medicine. For development of InterVA-M, all members of the expert panel had extensive experience in clinical obstetrics in resource-poor settings, public health and maternal mortality prevention programmes, as well as first-hand experience and knowledge of VA. In both instances, the experts came from a range of settings in developing or transitional countries where routine death registration is often absent, to encourage a generalised consensus not specific to any one region.

The refined InterVA method was applied to the VA data from the same 189 Vietnamese cases used to validate the preliminary model [94]. Comparisons were made with the cause of death as previously agreed by the two local physicians in Vietnam and with the results from the preliminary model. Rather than discuss sensitivity, specificity and PPVs when comparing physicians and the probabilistic approach, which would imply superiority of the PR approach, Kappa (κ) values were calculated to reflect the level of agreement between the two approaches.

The InterVA-M method was tested and assessed in collaboration with Nouna DSS. Completed VA forms for all deaths in the Nouna DSS since 1992 were scrutinised and those pertaining to females aged 15-49 years extracted and recorded in a spreadsheet. This yielded data on 380 deaths, each of which had been assessed by local physicians. Approximately one-third of these cases (n = 122) were used for initial testing of the probabilistic model to highlight any errors or omissions, which were then presented to and discussed by the expert panel. In addition to the 122 cases from Nouna DSS, archived VA data for 203 obstetric-related deaths from Bangladesh, 18 adult female deaths from Ethiopia and 15 adult female deaths from Ghana were also used in this initial testing to further reduce the chance of modelling the Burkina Faso setting too specifically.
The InterVA-M probabilistic model was then assessed using the hitherto untouched 258 cases of adult female death from Nouna DSS, by comparing the results from the probabilistic model with the original PRs. Given the diversity of the resulting physician diagnoses, and for the purposes of assessing InterVA-M, each individual physician diagnosis was used, weighted according to the number of physicians reviewing the case. This allowed comparison with the InterVA-M model’s possible multiple causes, which were weighted by their likelihoods. Thus, the same original VA questionnaire data were processed independently by the original physicians and the model, giving individually assigned cause(s) of death by both methods. These were then aggregated to CSMFs at the community level. To assess the added value of the free-text information, all the cases were processed with and without information derived solely from the open-ended sections of the VA questionnaires.

Data for Action
Using data relating to the five-year period 1996 to 2000, inclusive, the daily frequency of deaths was calculated from the BRHP database and plotted against time. Based on the CDC definition of a humanitarian crisis threshold, approximate daily (1:10000), weekly (7:10000) and monthly (30:10000) thresholds for the Butajira population were calculated for each calendar year within this period using mid-year populations. Days, weeks and months in which mortality exceeded each threshold level were identified and the threshold levels assessed in relation to their potential usefulness in prospectively identifying humanitarian crises in a timely manner and without excessive sensitivity, which could lead to false positives.

Ethical Considerations
Ethical clearance for on-going surveillance and use of surveillance data was provided by the ethical review boards of the Medical Faculty of Hanoi Medical University in Vietnam, the Medical Faculty at Addis Ababa University, Ethiopia, the Institutional Review Board of CRSN, Burkina Faso and Centre Muraz Ethical Clearance Committee, Burkina Faso. In addition, all individuals participating in the routine surveillance activities or population survey gave their informed consent at the time of data capture.
Image 8: Children playing in rural village in BRHP DSS, Ethiopia.
RESULTS

Detailed findings from each study are presented in the original publications appended at the end of this thesis. The following provides only an overview of the main findings from each study.

Sampling

Overall, 1% samples drawn using different sampling approaches represented the population well. Nevertheless, a degree of variation can be observed both between sampling approaches and different parameters. The ‘targets’ in Figures 8a-f illustrate the precision of different sampling methods in estimating the true population parameter levels. The centre of each target represents zero deviation from the true value with the percentage deviation increasing with distance from the centre. Complementing figures in Paper I, Figures 8a-f illustrate that the reliability of samples between parameters is related to the overall distribution of the parameters in the dataset. The consistent and approximately normal distribution of gender means that the proportion of males in the population was well represented in the samples, irrespective of the sampling approach (Figure 8a). In contrast, the more skewed and inconsistent distribution of educated individuals resulted in few samples adequately reflecting the overall situation in terms of falling within 5 and 10% tolerance of the unsampled value (Figure 8c). In particular, multi-stage DSS-style methods in which individual units were selected randomly are notable for the clear outlying samples (Figure 8d).

Samples drawn at the higher administrative level (larger sampling units) were generally closer to the unsampled population values than samples drawn at the lower administrative level (smaller sampling units) (Figures 8b-c). Random methods of each of the specific sampling methods generally performed better than PPS methods, especially at the higher administrative level (Figures 8d-f).
Figure 8: Average accuracy of parameter estimates derived by taking $20 \times 1\%$ population samples using 7 different sampling methods at 2 administrative levels, ‘ZD’ and ‘Conc.’ (Concession).

Notes: Numbers represent percentage deviation from the ‘gold standard’ value of: A) proportion of males B) proportion aged <$5$ years C) proportion educated D) proportion in poorest quintile E) proportion of adult female residents F) maternal mortality rate. (Simple = Simple random sampling; PPS = Sampling with probability proportional to size; Stratified = Urban/rural stratified sampling; DSS (PPS or Random) = Multi-stage DSS sampling (using either PPS or simple random methods in Stage 2); Dispersed = Dispersed multi-stage sampling). Note: Data in C are presented on a different scale.
Data Quality

Little change in population composition can be observed with the introduction of measurement error and missing data (see Paper II). Figure 9 shows the relative difference in mortality rates based on the datasets with 10% errors in age and 10% of deaths missed compared to the gold standard estimates. The largest differences in mortality rates based on these datasets, relative to the gold standard data, are at the extremes of age. Similarly, in Paper II, Figure 2, these errors have little noticeable effect on age-specific mortality rates between the ages of 5 and 60 years. At extremes of age however, erroneous data, and especially missing death information, result in more noticeable effects and a widening of the gap between erroneous estimates and the ‘true’ value.

Univariate and multivariate Poisson regression models of mortality rate ratios, including all the simulated errors, for the 10-year dataset are not greatly affected by the introduction of errors and missing data (Paper, Table 2).

Figure 9: Difference in mortality rate relative to gold-standard
Cause-of-death Determination

Refinement of InterVA and the development of InterVA-M using the Delphi-style technique proved useful. The computer-based user interface for each model was adapted to allow the entry of VA data as batch files (Figure 10a) and outputs were modified to only show more than one cause of death if the probability of the additional cause(s) was within 20% of the most likely cause (Figure 10b). This is in contrast to the preliminary InterVA model, which always gave the three most likely causes irrespective of probabilities.

As with the all-cause InterVA program, the InterVA-M method was developed to present up to three causes of death, each with a percentage likelihood, and an overall certainty factor, defined as the average of the likelihoods. In addition, the maternal model attributes a likelihood for each case being pregnant at death, dying within 6 weeks of pregnancy ending, or not being recently pregnant.

In refining InterVA and development of InterVA-M, there was strong consensus among the physicians that probabilities of causes of death with large variations in prevalence at the population level between regions, such as Human Immunodeficiency Virus/Acquired Immunodeficiency Syndrome (HIV/AIDS) and malaria, should have the possibility of being adjusted in the model to reflect local disease burdens. To warrant adjustment of the database it was felt that regional variations of disease prevalence should be at least ten-fold. Adjusting the model’s database to reflect low malaria endemicity and HIV/AIDS prevalence in FilaBavi worked well in improving the performance of the InterVA model, where 170 out of 189 (89.9%) InterVA diagnoses agreed with physician diagnoses, showing substantial improvement compared with the preliminary model where 134 cases (70.9%) were in agreement.

The level of agreement between InterVA-M and PRs in 258 adult female deaths from Nouna DSS was 81.8% (n = 211).
Figure 10a: Input batch file for InterVA and InterVA-M

Figure 10b: Text output from InterVA

InterVA-2 Verbal Autopsy Interpretation System

Malaria prevalence set to LOW
HIV/AIDS prevalence set to LOW

run at 05/28/04 15:01:47

ID: 13 cvd

data input
was this an elder 65+ years = yes
was death during wet season = yes
any paralysis/weakness on 1 side = yes
any chest pain = yes
any difficulty breathing = yes
any abdominal swelling = yes
any facial swelling = yes
any diagnosis of heart disease = yes

output
Most likely cause: Pneumonia probability 0.96
Second likely cause: Acute cardiac death probability 0.93
Third likely cause: chronic cardiac death probability 0.91

certainty 93 %
The 100% standardisation of this probabilistic approach to VA interpretation ensures that any differences in resulting population-level CSMFs are a result of actual differences in symptom reporting or disease profiles but not due to differing methods of interpretation. Figure 11 illustrates the model’s capacity to reflect mortality patterns in different communities using data from BRHP [113]. The figure shows higher rates of HIV/AIDS deaths, pulmonary tuberculosis (TB) and cardiovascular mortality in the urban area than in the rural areas, which is a credible finding and cannot be due to differing interpretation of data from rural and urban areas.

The effect of including or excluding exclusively free-text indicators from the Nouna VA questionnaires in the input to the InterVA-M model made no substantial difference to the model’s conclusions, and agreement between PR and InterVA-M was altered by less than 1%.

Figure 11: CSMFs for 54 deaths in Butajira town and 235 deaths in surrounding villages, derived from VAs interpreted by the InterVA model
Source: Fantahun et al. (2006).
Data for Action

In BRHP, a total of 3575 deaths occurred in the five-year period 1996 to 2000, among 259,958 person-years (p-y) of observation, giving an overall mortality rate of 13.8 per 1000 p-y. However, 1324 deaths (37.0%) occurred in 1999, and 800 (22.4%) in 1998, giving a mortality rate of 24.5 per 1000 p-y for 1998 and 1999, compared with 10.4 per 1000 p-y in the remainder of the five-year period.

Paper V, Figure 1 shows the daily frequency of deaths in Butajira during the selected years and highlights the periods during which each level of the humanitarian crises threshold was reached. In all, 199 days exceeded the 1 death per 10,000 population threshold during the five-year period. Almost three quarters (72.9%) of these high mortality days occurred during the 1998-9 period (105 (52.8%) occurred during 1999 and 40 (20.1%) occurred during 1998). The distribution of the other 54 (27.1%) high mortality days is rather evenly spread over the remaining three-year period.

The weekly threshold of 7 deaths per 10,000 population was exceeded on 16 occasions during the five-year period. All of these high mortality weeks occurred during 1998 and 1999 (4 weeks in 1998 and 12 weeks in 1999). The monthly threshold level of 30 deaths per 10,000 population was exceeded on 4 occasions over the five-year period, one month during 1998 and three months in 1999.
DISCUSSION

“[Surveillance methodologies are] distinguished by their practicability, uniformity, and frequently their rapidity, rather than by complete accuracy”

Dictionary of Epidemiology.

As with the summary of the main results given above, detailed discussions of the findings and implications from each individual study are provided in the separate papers appended to this thesis. The following therefore focuses more on the conceptual issues that this work has raised, relating to three interrelated themes of fit-for-purpose methods, cause-of-death determination and interpretation of VA data, and the use of surveillance data for public health action.

Fit-for-purpose Methods for Fit-for-purpose Data

If the purpose of population surveillance is to gain an overall impression of population composition and distribution of risk factors to inform public health action, when are methods and the data derived from these methods ‘good enough’? The different sampling approaches in Paper I and simulated errors in Paper II did not have a major affect on population profiles or the conclusions that one would draw; the innovative approaches to VA interpretation in Papers III and IV, although unlikely to be 100% accurate, provide useful and efficient summaries of population-level cause-specific mortality and address specific data needs; Paper V shows that longitudinal DSS mortality measurement approaches could be invaluable in the prospective detection of important mortality fluctuations. Perhaps these are all features of ‘good enough’ methods that address not only research needs, but are also relevant to policy and public health practice.

Paper I emphasises the need to consider general population distributions and uniformity of certain parameters within localities when selecting sampling methods. Selecting fewer large sampling units generally represents the unsampled population better than drawing a greater number of smaller units. A notable exception in Paper I, however, is the proportion of households in the poorest wealth quintile, where samples were more representative if a greater number of smaller units were selected (Paper I, Figure 5). This may be due to economics being more homogenous within as opposed to between areas, thus a greater number of smaller and potentially more diverse sampling units produces an overall estimate that is more representative of the wider population. Similarly, with regard to the proportion of individuals educated to secondary level or above, the superior performance of PPS approaches over simple random methods at the smaller sampling unit level may result from education being more homogenous in less populated units, therefore by increasing the likelihood of selecting more populous units (which have a more heterogeneous education distribution), PPS methods give more representative samples. These examples illustrate that there is unlikely to be a simple, ‘one-size-fits-all’ approach that can satisfy all needs of survey design and that compromises, which are informed by empirical evidence, need to be made.

The multi-stage DSS-style sampling modelled in Paper I performed inconsistently, producing a wide range of estimates determined by the region within which the second stage of sampling was confined. Outliers are associated with this approach when the urban area of Diapaga was randomly selected in the first of the two-stage sampling method. In particular, these outliers overestimate
the true population value of education level, perhaps due to Diapaga being the provincial town, and subsequently having a concentration of secondary schools and higher demand for an educated workforce. To an extent, dispersed multi-stage modelling between two randomly selected areas and stratification between urban and rural areas appeared to overcome the problem of outliers. Overall stratification methods did not greatly influence the accuracy of samples. This may be due to the conceptualisation of ‘urban’ and ‘rural’ areas within many African settings, where ‘urban’ may subjectively relate to an area being ‘less rural’ than its surroundings. Nevertheless, as illustrated by Figure 11, the advantages of living in towns compared to adjacent villages can often be observed in risk-factor and disease patterns, even if administrative and practical divisions are not easily made [22, 113].

Measurement imprecision as a consequence of sampling should be considered a systematic error, which is often accompanied by degrees of random, non-differential bias in reality. Whilst Paper II illustrates that random errors may not be hugely detrimental to the utility of population surveillance data, non-measurement issues, systematic shortcomings and other sources of bias should be factored into the generalisation and aggregation of surveillance findings, regardless of how well sources of measurement error are documented, controlled for or adjusted. Non-measurement issues may relate to ‘administrative errors’, which may result from increasing the surveillance population and operational complexity. Although methodologically appealing, the use of more widely distributed sampling nodes and bigger population samples for DSS systems, as suggested in Papers I and II, respectively, would likely cause increased administrative error in practice. Furthermore, the purpose of data collection must be borne in mind when planning such surveys. In terms of public health utility, dispersed sampling methods would not only be impractical in terms of intervention measures, but could also diminish the social force that a more unified study population might have as a lever for action.

Paper II separates random and systematic errors. The fact that some events, for example very early deaths or deliberately concealed events, are more likely to be missed or misreported was not factored into the analysis. Whilst it may have been possible, and perhaps more realistic, to simulate either age heaping or more random levels of misreporting, such investigation is beyond the scope of investigating random errors. Furthermore, such errors are only a superficial indicator of the quality of population data and are sensitive not only to a respondent’s ability to recall their ages or date of birth accurately, but also to training procedures for enumerators, where staff may be explicitly discouraged from recording rounded ages. Modelling such errors in Paper II would therefore have diminished the generalisability and relevance of the findings to surveillance operations overall as they would need to relate to specific quality-control procedures.

It may be argued that it would have been more appropriate to compound errors in the BRHP data analysis (Paper II) as they may have accumulated over a ten-year period. In reality, however, random errors are unlikely to accumulate or compound in a longitudinal dataset where regular update rounds and reconciliation of new and existing data mean that errors are often corrected over time. As such, the introduction of randomly distributed errors into the dataset is more likely to reflect the burden of random errors in DSS and other longitudinal population surveillance datasets.
Despite their inconsistent performance in Paper I, multi-stage DSS-style sampling methods are cheap and are often the only realistic option for undertaking research and monitoring activities in rural African settings, even if this is at the expense of statistical precision. Given that the consequences of relatively high degrees of non-differential imprecision on estimates of certain population and mortality parameters would be unlikely to alter population representation to such an extent that would influence public health and policy conclusions (Paper II), operationally simple sampling approaches should not be considered less appropriate than more complex methods as long as care is taken to minimise systematic errors. Understanding the potential advantages and limitations of methods in particular contexts is important for informing appropriate population survey design within the boundaries of financial and logistical constraints. Overall, a reasonable balance between sampling precision and data quality may be satisfactory.

Nevertheless, the standardisation of specific methods of monitoring health development in populations remains an urgent priority for surveillance systems globally. After more than a quarter of a century since the HIV/AIDS epidemic began in southern Africa, there remains ambiguity about the pace and extent of mortality associated with this disease, largely because of the continued absence of temporally and regionally comparable longitudinal cause-specific mortality data [17]. Similarly, there is great uncertainty about whether malaria mortality is rising in Africa and little is known about the pattern of injury deaths, or indeed, the emergence of non-communicable diseases (NCDs) [122-128]. Considerable methodological variation between surveillance sites reflects the lack of standard approaches or best-practice guidelines relating to methodologies in general and DSS methods in particular. Whilst there are logical explanations for this, namely different sites are established for different reasons and the vicissitudes of funding mean that priorities may change over time, this lack of standardisation hinders meaningful comparisons of data between settings and over time.

The standardised DHS approach is a prime example of how standardisation of methods facilitates global comparisons, providing a unique opportunity for comparative analysis of sample designs and sampling errors across many countries [45]. Similarly, application of the standardised WHO STEPwise (STEPS) approach to NCD risk-factor surveillance has enabled scientifically sound cross-country comparisons of risk-factor and mortality burdens using DSS data from three different countries (Ethiopia, Vietnam and Indonesia) [123]. A third example of the benefits of standardised methods is the Expanded Programme on Immunisation (EPI) coverage surveys, which utilise a well-defined field methodology that is explicit in its adequacy for purpose despite imperfections in the methodology [129, 130]. This is not to say, however, that standardised approaches must be rigid in their application. Whilst the DHS, STEPS and EPI coverage surveys conform to a standard model, all aspects of a specific survey are adapted to the particular needs and conditions of a country. Standardisation and best-practice guidelines for population mortality surveillance and routine DSS procedures should relate to the specific objectives of the surveillance activities and not simply to an underlying presumption of the nature of the DSS enterprise overall. Whilst the INDEPTH Resource Kit for Demographic Surveillance Systems provides invaluable guidance for the establishment of new DSSs, perhaps more discussion of evidence-based and fit-for-purpose methods would be appropriate [43].
Cause-of-death determination

Simply counting the number of deaths is not enough to develop understanding of population-level disease profiles and important health transitions. Therefore cause-specific mortality measurement is vital and, for the time being at least, VA methods are the only feasible way of doing this for the majority of the world’s population. Nevertheless, it is wrong to consider VA as an end in itself, but rather a tool to address the widespread lack of cause-specific mortality data as required by end users. To overcome the default assumption of general medical audiences that cause-of-death determination is solely for the purposes of individual-level cause-of-death certification, it is necessary to distinguish between the uses of mortality data and the attributes of VA methods for each use (Table 3). Addressing explicit mortality data needs is relatively uncommon in existing VA literature, where it seems to be assumed that readers will know why VAs are being conducted, often resulting in a further narrow assumption that VA is a direct surrogate for Western-style cause-of-death determination. Using this Western model as the template and gold standard against which all VA developments must be compared has resulted in the slow progression of VA methods. Rather than targeting specific gaps in the understanding of mortality in less-developed countries and considering whether the method is now more or less fit for purpose, VA developments tend to be discussed in terms of whether they meet medical ideals, thereby reinforcing the idea that there is a ‘one-size-fits-all’ solution to long-standing cause-of-death information gaps, which perhaps also causes resistance to computer-based approaches such as InterVA and InterVA-M.

Table 3: Users, purpose and desirable features of VA-derived cause-of-death data in relation to the InterVA and InterVA-M methods

<table>
<thead>
<tr>
<th>User</th>
<th>Mortality data purpose</th>
<th>Desirable features of mortality data</th>
<th>InterVA/InterVA-M</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local Public Health Managers</td>
<td>To effectively plan health services based on local patterns of disease.</td>
<td>Simple and consistent categorisation of top ranking causes of death and public health priorities</td>
<td>Simple and efficient to use</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100% consistent and repeatable</td>
<td>Generates multiple causes of death thus more accurately reflecting the interaction of diseases and more realistically highlighting population-level morbidity and mortality burdens</td>
</tr>
<tr>
<td>Epidemiologists and health service researchers</td>
<td>To describe situations in terms of mortality patterns and to evaluate the effectiveness of interventions across time and regions.</td>
<td>Consistent identification and reliable measurement of specific causes of interest</td>
<td>Good correspondence with ICD-10 categories</td>
</tr>
<tr>
<td>National and global health authorities</td>
<td>To build respective pictures of health patterns</td>
<td>Consistency and reliability from a wide range of sources. Reliable ascertainment of population-level CSMFs. ICD-coding</td>
<td></td>
</tr>
</tbody>
</table>

For a general medical audience, the concept of identifying specific causes of death that are of public health interest out of all possible ‘causes’ along the long chain of economic, social, operational, biomedical and physical events leading to death is rarely considered. Limiting VAs to
a medical model undermines the full potential of VA methods, which may be adapted to any specific point along this chain of events and so may be designed to address specific public health or mortality questions in a way that Western, medical-based models cannot. For example, ‘social autopsies’ are designed to identify failures in social and operational infrastructure systems that contribute to death by gathering information on the social and operational events preceding death. The fact that such an approach is explicit about targeting a specific point along the chain of events leading to death is a useful start in terms of data collection and analysis and may also mean that the routine hoops that traditional VA methods must jump through to be considered valid can be missed. This allows more useful discussions of adequacy for purpose and addressing specific data needs.

VA interpretation is essentially an issue of medical decision-making in which there has been renewed interest following the heightened emphasis on evidence-based medicine (EBM) over recent decades. As previously highlighted, traditional death certification deals with individuals, often with a relatively rich medical case history, whereas VA deals with populations and comparatively sparse, ‘second-hand’ case histories from lay respondents. VA death certification is a public health not a clinical issue, yet a clinical approach is often used to interpret VA data.

Figure 12: Illustration of Bayes’ theorem as applied to the hypothetical situation described. Even when the test is positive for disease B (with true positive results 80% of the time), the probability of disease B is lower than the probability of disease A.
These fundamental differences in clinical versus public health principles should influence the development and application of VA methods and should inform more logical population-based approaches to VA interpretation and validation.

This conflict between clinical and population-based approaches to medicine is not a new one. The inherent practical and cultural differences between EBM and traditional clinical regimes whereby an individual physician’s subjective and independent judgement is what matters, runs deep. Physicians practicing clinical medicine accumulate a vast and idiosyncratic knowledge of medicine based on personal experience and interpretative stories passed on through formal (seminars, journals etc.) and informal (staffroom and golf course conversations etc.) processes during their professional development. As such, it appears to be widely believed in clinical practice that an objective and non-random approach to evaluating evidence does not necessarily achieve the truth. This of course does not mean that physicians practising clinical medicine are ignorant or dismissive of general research findings. Instead, they may find reasons why their patient is atypical and therefore requires treatment different to that recommended by study results and statistics.

The clinical physicians’ review of evidence utilises additional, often implicit information; in other words, the value of the objective, scientific evidence of test results, for example, may be influenced by the physician's personal experiences or beliefs. At a basic level one may assume that determining the value of evidence assists the clinician in determining whether a particular diagnosis is more or less probable. However, from an epidemiological perspective, even subjective probabilities ought to refer to their frequency in a population, otherwise where is the evidence base and how is one meant to monitor, evaluate or plan future population-based health interventions?

Public health is based on the technocratic desire for an objective and scientifically logical approach to decision-making and the interpretation of evidence. The goal of public health, as well as all health service researchers and policy-makers, should be to employ population- rather than individual-based measures to assess the utility and cost-effectiveness of health care. The technocratic public health approach is more qualitative and objective, not least in the respect that, in the traditional clinical approach, each observation must go through an individual physician’s personal, subjective filter before becoming data for analysis. The conflict exists therefore between the professional discretion or the ‘art’ of medicine [131] and medicine based on logic and scientific fact.

The problems of adopting a clinical approach to interpreting evidence may be demonstrated by the following hypothetical example.

A physician is asked to determine an individual’s cause of death. The indicators suggest the individual may have died from disease A or disease B (which is 19 times less common than disease A). A post-mortem test which distinguishes between diseases A and B and provides the true result (true positive) 80% of the time suggests that the individual died from disease B. Either disease A or disease B must be selected as the main cause of death, which should it be? [92]
Literature suggests that the vast majority of physicians would state the cause of death as disease B [92], yet, according to Bayes’ theorem, it is more likely that the individual died from disease A (Figure 12 – the bottom left box, which is larger than the bottom right, represents greater probability). Therefore, from a technocratic, probabilistic point of view, disease A should always be selected as the cause of death in such cases. Presumably the physicians choosing disease B as the cause of death do so either because they do not appreciate the statistics, or they place an additional value on the scientific test result. If VA methods are to be validated in comparison to PR then this additional value is an important factor to consider, although whether or how this should be done is questionable.

The debate between the appropriateness of PR for interpreting VA material versus the use of automated models is likely to continue for some time, and the respective advantages and disadvantages of each must be considered in relation to their particular setting and purpose. Despite the possible subjective interpretations by physicians, they are generally considered to be better at appreciating the nuances of individual VA cases and may also offer greater validity of cause-of-death diagnoses, which is suitable for a one-off VA study. The probabilistic InterVA and InterVA-M approaches may lack some of the subtleties of physician reasoning but, by offering 100% reliability and consistency of cause-of-death diagnoses over time, address the specific needs of researchers and public health surveillance systems assessing changes in causes of death over time.

Whilst individually VA-determined causes may be methodologically easier to compare with individually-certified causes of death from other settings, this does not necessarily imply a need for certainty at the individual level. Rather, it emphasises the need for reliable methods of interpreting CSMFs for known populations. By generating multiple rather than single causes of death, the probabilistic approaches are likely to more accurately reflect the interaction of different diseases and more realistically highlight morbidity and mortality burdens at the community level. This more epidemiological approach may simplify the VA interpretation process substantially and generate consistent population-level cause-specific data as required by the main users. Realistic baseline prevalence of each cause of death and indicator is key for successful functioning of Bayesian approaches to VA interpretation. Delphi-style consensus activities used in the development and refinement of the probabilistic approaches were necessary for deciding realistic baseline prevalence not specific to any one region. Unlike strict Delphi methods, however, the approach used was not anonymous and involved round-table discussions, thus group dynamics and individual personalities may have resulted in the over or under-representation of certain individuals in the final consensus. It is therefore necessary to further apply and evaluate the InterVA and InterVA-M approaches using data from a wider range of settings.

The current cause-of-death lists are not exhaustive, but do correspond well with ICD-10 categorisation of deaths in terms of anatomical systems or pathophysiological mechanisms [87]. Table 4 illustrates the correspondence between InterVA-M’s pregnancy-related causes of death and broad ICD-10 groupings. Ante-, intra- and post-partum haemorrhage have been broadly grouped under ‘haemorrhage’ and such broad categories are necessary where it is unrealistic to expect to be able to distinguish between sub-groups from VA data or where there is substantial overlap in interventions and association with established risk factors. Such broad groupings are both adequate and efficient to guide public health prioritisation.
Table 4: Correspondence between pregnancy-related InterVA-M causes of death and ICD-10

<table>
<thead>
<tr>
<th>InterVA-M Cause-of-death Category</th>
<th>Corresponding ICD-10 Code(s)</th>
<th>Corresponding ICD-10 Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Haemorrhage</td>
<td>O20</td>
<td>Haemorrhage in early pregnancy</td>
</tr>
<tr>
<td></td>
<td>O46</td>
<td>Antepartum haemorrhage</td>
</tr>
<tr>
<td></td>
<td>O67</td>
<td>Intrapartum haemorrhage</td>
</tr>
<tr>
<td></td>
<td>O72</td>
<td>Postpartum haemorrhage</td>
</tr>
<tr>
<td>Pregnancy-related sepsis</td>
<td>O23</td>
<td>Infections of genitourinary tract in pregnancy</td>
</tr>
<tr>
<td></td>
<td>O85</td>
<td>Puerperal sepsis</td>
</tr>
<tr>
<td></td>
<td>O86</td>
<td>Other puerperal infections</td>
</tr>
<tr>
<td>Ruptured uterus</td>
<td>O71.0</td>
<td>Rupture of uterus before onset of labour</td>
</tr>
<tr>
<td></td>
<td>O71.1</td>
<td>Rupture of uterus during labour</td>
</tr>
<tr>
<td>Obstructed labour</td>
<td>O63</td>
<td>Long labour</td>
</tr>
<tr>
<td></td>
<td>O64</td>
<td>Obstructed labour due to malposition and malpresentation of foetus</td>
</tr>
<tr>
<td></td>
<td>O65</td>
<td>Obstructed labour due to maternal pelvic abnormality</td>
</tr>
<tr>
<td></td>
<td>O66</td>
<td>Other obstructed labour</td>
</tr>
<tr>
<td>Pregnancy-induced hypertension</td>
<td>O13</td>
<td>Gestational (pregnancy-induced) hypertension without significant proteinuria</td>
</tr>
<tr>
<td></td>
<td>O14</td>
<td>Gestational (pregnancy-induced) hypertension with significant proteinuria</td>
</tr>
<tr>
<td></td>
<td>O15</td>
<td>Eclampsia</td>
</tr>
<tr>
<td></td>
<td>O16</td>
<td>Unspecified maternal hypertension</td>
</tr>
<tr>
<td>Abortion-related death</td>
<td>O03</td>
<td>Spontaneous abortion</td>
</tr>
<tr>
<td></td>
<td>O04</td>
<td>Medical abortion</td>
</tr>
<tr>
<td></td>
<td>O05</td>
<td>Other Abortion</td>
</tr>
<tr>
<td></td>
<td>O06</td>
<td>Unspecified Abortion</td>
</tr>
<tr>
<td></td>
<td>O07</td>
<td>Failed attempted abortion</td>
</tr>
<tr>
<td>Anaemia</td>
<td>O99.1</td>
<td>Anaemia complicating pregnancy, child birth and the puerperium</td>
</tr>
<tr>
<td>Ectopic pregnancy</td>
<td>O00</td>
<td>Ectopic pregnancy</td>
</tr>
<tr>
<td>Other maternal causes</td>
<td>O95</td>
<td>Obstetric death of unspecified cause</td>
</tr>
</tbody>
</table>

Source: Fottrell et al. (2008).

In the evaluations of InterVA and InterVA-M, the tools used to collect VA data were unrelated to the probabilistic programs, thus not all indicators available in the data were built into the models and vice versa. This highlights the dichotomy between focus on data collection and interpretation. Both InterVA and InterVA-M are capable of processing open narrative and closed ‘tick-box’ data from all sections of VA questionnaires, although identifying and extracting indicators from
open text requires greater medical knowledge and subjectivity, and may be associated with little or no diagnostic advantages [132, 133]. Efforts need to be refocused on standardisation of the complete process to facilitate better comparison of VA data between settings and to improve the consistency of VA-derived mortality estimates. Recent work led by the WHO is perhaps a sign of progress in this area [74].

The cause-of-death list included in the probabilistic models is consistent with the philosophy of establishing priorities across broad causes for which the public health response implications are essentially similar [17]. Once these priorities have been established with adequate certainty, it may be possible to investigate causes in more detail using disease-specific surveillance data, disease modelling or more extensive VA methods. The grouping of HIV/AIDS and TB in one category in the evaluations of InterVA and InterVA-M, however, is likely to mask the potentially large number of TB cases without HIV. Given that the interventions and treatment policies for the two conditions differ considerably, efforts must be made to separate these causes into distinct categories, perhaps through the introduction of additional indicators or further expert review of the model’s probabilities by physicians from settings with high TB prevalence but relatively low HIV/AIDS prevalence. More cause-of-death categories or subdivisions may also need to be introduced into the model, for example, between intentional/unintentional and conflict-/non-conflict-related injuries. This depends largely on the VA study setting and the end-user needs and, in principle, additional indicators and causes prevalent only in certain regions may be added to the model. Similarly, cause-of-death sub-groups may be expanded to meet specific intervention or evaluation objectives.

The importance of subtleties of language and the precision and clarification of definitions and concepts of health and disease to minimise misreporting and facilitate standardised methods of VA interpretation has been previously reported, and was highlighted during panel discussions for both the all-cause InterVA model and the adult-female-specific InterVA-M model [75, 76, 134]. The variability of definitions of abortions, miscarriages and stillbirths among the panel, for example, prevented the subdivision of the broader abortion-related death category in InterVA-M, despite this distinction being useful from a public health perspective. The fact that the significance of various indicators is likely to vary in different regions was also highlighted; for example, in rural areas in many developing countries, delivery at a health facility is indicative of complications since most ‘normal’ births occur at home, whereas in other settings delivery at a health facility is the norm.

Since the above evaluations, the InterVA method has been applied to a community-based study of the affect of varying malaria treatments on malaria-specific death rates in two districts of northern Ethiopia. The simple and quick InterVA method allowed rapid cause-of-death determination of 1,314 deaths with complete consistency between the districts. Proportionately fewer malaria deaths were found at higher altitudes, even though altitude is currently not one of the model’s input parameters, which adds further credibility to the method [135]. A further, separate study in an adult population in Addis Ababa demonstrated high sensitivity and specificity of determining deaths as AIDS-related or not in comparison with hospital records [136].

The InterVA-M method has also been refined and re-evaluated recently using the same Nouna DSS data used in Paper IV [10]. Case-by-case agreement between the refined model and PR was
DISCUSSION

lower than in the preliminary model, but population-level CSMFs showed improvement and compared well with PR.

PDA-based versions of both InterVA and InterVA-M have been developed in a number of languages. Pilot testing of the InterVA-M method in Nouna DSS proved successful. Real-time processing of the VA data on the PDA itself gives the unique opportunity to provide instant feedback to respondents, something that is quite rare in field surveys (Figure 13). Work is currently underway in Burkina Faso to evaluate the data-capture and reporting biases when using InterVA-M with PDA technology.

![Figure 13: InterVA-M interface on PDA showing indicators and real-time diagnosis of likely cause of death](image)

Data for Action

Population surveillance sites in less-developed countries satisfy several necessary components of ideal humanitarian information systems. Their prospective, longitudinal nature means that the completeness of reporting is likely to be greater than for one-off household surveys and VA techniques permit population-level estimates of cause-specific mortality. Paper V demonstrates that the CDC definition of a humanitarian crisis could easily be applied in such surveillance settings for the prospective detection of important health changes. In particular, the weekly threshold level appears to be the most useful in identifying significant increases in mortality in a timely manner without the excessive sensitivity of the daily threshold. The monthly threshold level tested in Paper V would be unlikely to trigger alarms early enough for appropriate investigations and interventions to be designed and implemented.

Failure to use surveillance data for public health action can be considered in terms of ethical principles in accordance with which surveillance and epidemiological studies should be undertaken. However, the concept of surveillance involves populations rather than individuals and the practice of public health and epidemiology is distinct from clinical practice and other forms of medical research [137]. It follows, therefore, that public health surveillance ethics has different domains from those of medical ethics and thus requires careful consideration at the outset of any
surveillance activity. Nevertheless, certain principles are common between the two, not least the need to balance potential benefits against risks; to justify the risks and intrusion of surveillance, the collected information must have a demonstrated utility.

Participants of longitudinal surveillance may be burdened by a loss of privacy, time spent completing interviews and by possible adverse psychological effects such as enhanced grief or anxiety [138, 139]. Given the level of commitment that such surveillance activities can require from communities, it is morally appropriate, if not obligatory, to minimise these risks and maximise potential benefits. This principle of beneficence demands both the protection of individual welfare and the promotion of common welfare, and underlies ethical rules and norms that require public health professionals to act in a timely manner on the information they have and that they make the information available to the public expeditiously [106]. The failure of BRHP to act on data indicating unusually high mortality (Paper V) could be considered contrary to this ethical principle, and thus raises important questions about the ethical responsibilities of those involved in longitudinal population surveillance in developing countries.

In specifying ethical norms in relation to measuring mortality, it is important to draw additional distinctions between epidemiologic research and public health practice [137]. Although some activities can clearly be defined as either research or non-research, for others, such as longitudinal surveillance, it is more difficult [140, 141]. It is debatable whether sentinel surveillance and DSS operations in developing countries are responsible for practical public health action and active surveillance for public health purposes is often a secondary and often neglected function. In relation to crisis detection, however, it is likely that the division between research and practice is not particularly relevant, and the ethical imperative to use and act on the data becomes paramount. The concept of ‘humanness’, which is a deep-rooted moral principle in many African societies demanding humane dealings with others [139], should override the division between research and practice.

Systems for analysing surveillance data and disseminating findings for timely action need much more consideration in population surveillance activities, and in DSS settings in particular. Paper V demonstrates that the prospective detection of health crises in surveillance settings need not be complex. In relation to mortality, simple tally systems may be an adequate starting point. Population surveillance sites in developing countries typically operate in cooperation with local health authorities, universities and local and foreign government ministries. Therefore, key actors in health, development and relief are likely to be receptive to efforts to enhance communication with population surveillance organisations. The responsibility of the surveillance site should be to report data to these other parties, especially in crisis-vulnerable populations when thresholds are reached. In combination with data that may be collected by other parties, such as environmental and meteorological data, mortality information could enhance understanding of environmental and population interrelationships and provide a more complete incentive for public health action.

Paper I shows that, for certain mortality parameters, sample-based population estimates can be representative of the surrounding 100-fold population. Nevertheless, the 1997 and 1998 epidemics observed in the BRHP DSA (Paper V) were not reflected in the national DHS data, nor in the regional DHS figures for the Southern Nations, Nationalities and Peoples Region of Ethiopia in which Butajira district is located [22]. Whether these two major epidemics were so
localised that they should not have been evident in the DHS sample, or whether the retrospective DHS approach tends to simply minimise short-term fluctuations is unknown. DSS approaches have more scope to detect the extent of local variations in mortality, while more dispersed methods, such as the DHS approach, have the possible advantage of averaging out local variations across a region or nation. These are complementary characteristics, since both the extent of local variation and stable trends over time are both important parameters for health planning. These findings further diminish arguments for considering DSSs as purely research-driven enterprises, and suggest that greater recognition of the unique ways in which they may be able to support developing-country health systems is needed.

Further Investigations
The issues raised in this thesis, as well as some additional aspects of surveillance methodologies, warrant further investigation. For example, the effects of differing surveillance intervals is still poorly understood. Also, strict epidemiological data analysis should reflect the sampling strategy employed and it may be interesting to investigate the effects of different strategies on more complex statistical analyses, such as multivariate and multi-level modelling. The practical importance of not taking design effects into consideration when analysing data and how this may influence the usefulness of the data for different end-user perspectives remain important questions.

Further standardisation of surveillance methods is also of great importance to enable reliable comparisons to be made between regions and over time. The next steps in the development of InterVA and InterVA-M will be their application to data from a wider range of sources. It may also be possible and useful to adapt the probabilistic approach to cause-of-death determination in other age sub-groups, such as infants.

Satisfactory measurement of certain aspects of health and risk factors cannot always be achieved by interviews alone, not least because many morbid conditions are asymptomatic. In future research and surveillance it is likely that biomedical and demographic methods will need to be integrated, thus posing further methodological challenges.
CONCLUSION

The publication of this thesis in 2008 coincides with the 60th anniversary of the WHO and 30 years since the Alma Ata Declaration advocated primary health care as the main strategy for achieving health for all by the year 2000 [142]. Despite notable achievements in global health, the 2000 deadline has passed and health for all has still not been achieved, not least as a result of inadequate information on the fundamental epidemiological questions of person, place and time.

Levels of adult mortality are poorly measured and accurate cause-of-death data remain generally unavailable. Very few countries have developed effective and efficient health information systems that take into account the needs of different users, from local to national and global levels [143].

Establishment of registration systems for entire populations is unlikely to occur in the short to medium term. In the meantime, sample-based and sentinel population and mortality surveillance can yield sufficiently reliable and relevant information for programme action, and are well within the means of many developing countries. Both can provide valid and reliable information on causes of death and have the added benefit of providing more complete registration over the longer term. Indeed, such systems represent the only useful alternative to establish the evidence base for health policy and programme delivery for the foreseeable future in much of the developing world. Clearly, however, such activities can only yield useful information when good design principles are followed [17].

There is no simple survey design that can satisfy all data needs. Methodological decisions about surveillance should therefore be a synthesis of all available and relevant knowledge relating to clearly defined concepts of why data are being collected, how they can be used and when they are of good enough quality. A number of mathematical principles have been developed to demarcate what is ‘significant’ statistically, but no comparable principles have been established to indicate what is significant operationally in relation to public health action. Guidance can be sought from textbooks and other sources, such as the INDEPTH network, but rarely does sufficient pilot testing and conceptual thinking on evidence-based methods for specific goals play a large enough role.
Improvement in surveillance systems needed at local, national and international levels, and more integration between these levels, is required to deal with health threats. In this respect, the combination and comparison of existing datasets, such as DSS and DHS, may maximise the potential utility of existing data [22, 115, 144]. In the absence of refined methods and clearly defined responsibilities, however, entirely preventable morbidity and loss of life is likely to continue to occur on a large scale in vulnerable populations [109]. Systems for prospectively monitoring changes in health, and the use of this information for public health action, must become commonplace in routine surveillance operations as it is in randomised control and clinical trials. Explicit discussion of these responsibilities with surveillance communities as well as within the INDEPTH network may help to maintain public trust in, and understanding of, longitudinal surveillance. This in turn may be a significant step towards more widespread, routine vital-event surveillance. Ultimately, the responsibility of mortality surveillance systems lies not only in counting deaths, but also in using the information to make deaths count.

**Image 9:** 'Children sitting on mother’s grave', Kenya
ACKNOWLEDGEMENTS

So many people have helped and encouraged me throughout these PhD studies that naming anybody risks leaving somebody out. Therefore, I will begin with an all-encompassing thank you to all fieldworkers, data-entry and data-management personnel, administrators, drivers, committed respondents, funding agencies and many other dedicated individuals in Vietnam, Ethiopia and Burkina Faso, as well as in Sweden, the United Kingdom and the United States who have contributed to this work. Thanks to everyone whom I have had the pleasure of working with, sharing ideas with and learning from along this journey.

I owe much gratitude to my mentor and supervisor, Professor Peter Byass, who has taught me so much and has guided this work, and whose expertise and good humour have made the PhD process an absolute joy.

Professor Stig Wall has been an endless source of encouragement and guidance since I first joined the department in Umeå. It has been an honour and a privilege to be supported by such an esteemed teacher, renowned scientist and warm-hearted individual.

Anna-Lena Johansson’s faultless organisation and administration has ensured that all of my work in Sweden and elsewhere has run smoothly. I am also grateful to her for introducing me to the pleasures of long walks in the Västerbotten countryside.

I am grateful to Birgitta Åstrom for always taking the time to make me feel so welcome in Umeå and for taking all the hassle out of finding accommodation, opening bank accounts, paying bills…the list is endless!

Thanks to Axel Emmelin for enjoyable times in Burkina Faso and for the privilege of reproducing a few of his wonderful photos in this thesis.

I am grateful to all my friends and colleagues in the Division of Epidemiology and Public Health Sciences, Umeå University. This is such a wonderful environment to work in and the great community spirit and sharing environment has enabled me to learn as much in the fika room as I have in the library! I am particularly grateful to Anders Emmelin, whose door has always been open when I have needed guidance on matters ranging from DSS methods to broken spectacles and jazz, and also to Professors Ulf Högb erg, Hans Stenland and Lars Wein hall for their help and encouragement.

I have been fortunate to visit each of the study areas described in this thesis and have invariably been welcomed with a hospitality that overwhelms me. Even when language has been a barrier, the generosity of my hosts has always enabled me to leave with unforgettable experiences, greater knowledge and more friends than I arrived with. Special thanks go to Mr Agung Nugroho, Dr Asefaw Getachew, Mme Cécile Tamini, Dr Thomas Ouedraogo, Dr Sennen Hounton, Dr Rasmané Ganaba, Dr Issiaka Sombié, Mr Ibrahima Diallo, Mr Henri Somé, Dr Danielle Yugbaré-Bélemsaga, Monsieur Olivier and all the drivers.

It has been a pleasure to work with and to learn from colleagues in Immpact and in the Maternal and Neonatal Health Group at the London School of Hygiene and Tropical Medicine, in particular Dr Veronique Filippi and Professor Oona Campbell, whose encouragement and flexibility with regards to this PhD work are very much appreciated.
ACKNOWLEDGMENTS

I will always be grateful to Dr Heather Roberts at the University of Nottingham for opening the door to the fascinating world of public health and epidemiology and for her encouragement ever since I passed through it.

The co-authors and peer-reviewers of my papers have provided invaluable expertise that has ultimately shaped this work. In particular, I wish to thank Professor Yemane Berhane for taking an interest in this thesis and for providing useful comments and encouragement throughout. Thanks also to Dr Kathy Khan, who provided sound advice on how to write a cover story, and whose own thesis inspired the title of this.

My friend Dr Nawi Ng’s good humour and vast knowledge has taught me much, and I am indebted to him for thoroughly reviewing this thesis at its early stages and for his wise comments that improved it. I am also grateful to Nawi’s family, who made me feel so welcome in Jogjakarta.

My friends and peers Drs Hoang Van Minh, Dao Lan Huong and Fikru Tesfaye have been a pleasure to work and socialise with over the years. I am extremely grateful to Minh and Huong for teaching me much about surveillance and verbal autopsy in FilaBava and to their family and friends for a wonderful experience in Vietnam.

My friends in London, Leeds and elsewhere have been a greater source of inspiration than they probably realise, and their encouragement and interest in this work has been a source of motivation.

This work was made easier by the encouragement I received from the Adcock and Knowles families in Nottingham, and by the constant source of support that I can rely on from my siblings and in-laws, Matthew, Tracey, Sinead, James, Katie and Nick. The recent arrival of my nieces, Isabel and Holly, has been a joyous distraction from PhD life.

Without my parents, none of this would have been possible. Their generous support (both moral and financial) gave me the confidence to take on this work, and their tireless love and encouragement motivate me in everything I do.

Finally, I cannot overstate my gratitude to my best friend and in-house editor, Jo, whose love, patience and support inspires me to do everything in life to the best of my ability.
REFERENCES

REFERENCES

45. TN Lê, VK Verma: An analysis of sample designs and sampling errors of the demographic and health surveys. Calverton, Maryland, USA: Macro International Incorporated; 1997.


95. InterVA: [www.interva.net].


111. A Emmelin, M Fantahun, Y Berhane, S Wall, P Byass: The epidemiology of modern-day famine in Ethiopia: food shortage or inequitable distribution? 2007, SUBMITTED.


