Theme: Building health information systems

An economic evaluation of data collection methods for vital statistics

Eliana Jimenez Soto

Kim-Huong Nguyen

Zoe Dettrick

Andrew Hodge

Alan D. Lopez

Working Paper Series • Number 28 • June 2013 • WORKING PAPER

School of Population Health
University of Queensland

Strengthening health systems in Asia and the Pacific through better evidence and practice

An AusAID funded initiative

For the PDF version of this paper and other related documents, visit

www.uq.edu.au/hishub

About this series

The Health Information Systems Knowledge Hub's Working Paper Series is the principal means to disseminate the knowledge products developed by the hub as easily accessible resources that collectively form a lasting repository of the research findings and knowledge generated by the hub's activities. Working papers are intended to stimulate debate and promote the adoption of best practice for health information systems in the region. The series focuses on a range of knowledge gaps, including new tools, methods and approaches, and raises and debates fundamental issues around the orientation, purpose and functioning of health information systems. Generally, working papers contain more detailed information than a journal article, are written in less-academic language, and are intended to inform health system dialogue in and between countries and a range of development partners.

Many working papers have accompanying products such as summaries, key points and action guides. The full range of documents, as well as other resources and tools, is available on the Health Information Systems Knowledge Hub website at www.uq.edu.au/hishub/publication-tools.

This research has been funded by AusAID. The views represented are not necessarily those of AusAID or the Australian Government.

Author details

Eliana Jimenez Soto, Kim-Huong Nguyen, Zoe Dettrick, Andrew Hodge,
School of Population Health,
The University of Queensland

Alan D. Lopez,
Health Information Systems Knowledge Hub,
School of Population Health,
The University of Queensland

Acknowledgments

The authors would like to thank Sonja Firth, Sriram Shankar and Kasia Alderman for their much appreciated assistance with creating and reviewing this paper.

Published by the Health Information Systems Knowledge Hub,
School of Population Health
The University of Queensland
School of Population Health Building,
Herston Rd, Herston Qld 4006, Australia

Please contact us for additional copies of this publication, or send us feedback:

Email: hishub@sph.uq.edu.au
Tel: +61 7 3365 5400
Fax: +61 7 3365 5442
www.uq.edu.au/hishub

Design by Biotext, Canberra, Australia

Contents

Acronyms and abbreviations

CBA	cost–benefit analysis
CE	cost efficiency
CEA	cost-effectiveness analysis
COD	cause of death
CRVS	civil registration for vital statistics
CUA	cost utility analysis
DEA	data envelopment analysis
DCM	data collection method
DHS	Demographic and Health Survey
DMU	decision-making unit
DSP	disease surveillance points
DSS	demographic surveillance sites
EA	efficiency analysis
GGB	Generalised Growth Balance
HMS	health management system
IDS	integrated disease surveillance
ICD	International Classification of Diseases
MCOD	medical certification of death
NHS	National Household Survey
QADI	quality-adjusted data index
QALY	quality-adjusted life-year
SEG	Synthetic Extinct Generations
SRS	Sample Registration System
TE	technical efficiency
TFP	total factor productivity
VA	verbal autopsy
WHO	World Health Organization

Summary

Over recent years there has been a strong movement towards the improvement of vital statistics and other types of health data that inform evidence-based policies (WHO 2011). Collecting such data is not cost free. To date, however, there is no systematic framework to guide investment decisions by donors and governments on methods of data collection for vital statistics or health information in general. Our study draws on the vital statistics and economics literature to build such a framework and systematically assesses the comparative costs and outcomes/benefits of the various methods for collecting vital statistics.

To achieve this we:

- reviewed the existing literature on the available data collection methods (DCMs) for vital statistics
- examined alternative economic approaches for evaluating costs and outcomes, which can be applicable to DCMs
- described the main elements and methodological challenges of undertaking such evaluation
- proposed a framework for a concrete economic evaluation to undertake such an assessment, which includes a transparent account of the main assumptions behind the framework and the analytical tools proposed
- performed a simulation exercise to build a stylised scenario of a hypothetical low-income country to illustrate how our framework might operate.

Key points

Literature review on methods of data collection for vital statistics

There are three major ways to collect data on vital statistics:

- registration of births and deaths within either the whole or a subset of a population
- collection of mortality and cause-of-death indicators via questionnaires administered by censuses and surveys
- continuous recording of births and deaths occurring in health facilities (i.e. facility-based collection).

Mahapatra et al. (2007) developed a systematic framework to assess the quality of data on vital statistics (Mahapatra et al. 2007; Ruzicka & Lopez 1990) which was used to evaluate the quality of cause-of-death statistics that countries report to the World Health Organization (WHO).

The cost of collecting data on vital statistics by different methods is not well documented (Rommelmann et al. 2005a, 2005b).

A systematic assessment of DCMs – the economics approach

We identified two major economic approaches to the systematic assessment of DCMs:

- Economic evaluation involves the systematic and comparative assessment of costs and benefits/outcomes of alternatives. It covers cost–benefit analysis, cost minimisation analysis and cost-effectiveness analysis (CEA). Only CEA was found to be applicable to the economic evaluation of DCMs.
- Efficiency analysis (EA) uses econometric techniques to estimate and compare the productivity (i.e. the relationship between outcomes/outputs and costs) of various units of production. In our case, it could be used to measure the relationship between the outputs and costs of data collection methods for vital statistics.

The three main elements of an economic assessment of DCMs are as follows.

1. Identify the alternatives to be evaluated.
The literature identified various alternative methods to collect data on vital events. The country situation and government policy will dictate which methods to include in the evaluation.

2. Identify and measure the outcomes of DCMs.
DCMs produce data, of which the value/benefit is intangible and difficult to quantify. Any measurement of DCM outcomes will need to capture both the quantity and quality dimensions of data:

- Quantity – One of the basic elements of a sound economic evaluation is the choice of appropriate physical units to measure outcomes (i.e. data). We discuss two options: unit records vs number of people represented. The best option to choose depends on whether there is a long-term government policy of establishing complete civil registration for vital statistics (CRVS).
- Quality – We use a slightly modified version of the framework by Mahapatra et al. (2007) to build a single composite index of data quality for vital statistics. We examine several alternatives for weighting the criteria (OECD 2008).

3. Identify and measure costs of DCMs.
We suggest that the first step is to collect at least basic information to quantify the current costs of DCMs, which should take into consideration apportioning of costs. That is, if Demographic and Health Surveys collect data for a large number of variables, not just vital statistics, what proportion of DHS costs should be allocated to vital statistics? As information on current costing structures improves, cost functions (i.e. how costs vary as coverage changes) can be reliably estimated.

A stylised scenario

We use simulated data to build a stylised scenario and rank alternative DCMs of varying quality according to their cost-efficiency (i.e. the extent to which they produce good value for money). The results of the stylised scenario show that the rankings of the DCMs are not affected by the choice of economics approach (i.e. CEA or EA). However, as shown, the rankings are affected by how quantities are measured.

Our work highlights the challenges of conducting economic evaluations of DCMs. However, we demonstrate that it is a feasible task. We have developed a systematic framework for such an evaluation which can be piloted with country data. The framework is also general enough to be easily extended to other areas of health information systems.

Introduction

Health information, in particular vital statistics and cause-of-death (COD) data, is an essential public good. However, every year, one-third of births are not officially recorded, and three-quarters of all deaths lack either registration or a medically certified cause (Lopez & Thomason 2012). Recent global initiatives have thus called for significant improvements in country systems to register births, deaths and causes of death (WHO 2011; Mahapatra et al. 2007).

The preferred data collection method (DCM) for vital statistics is a complete system of civil registration for vital statistics (CRVS) with medical certification of all deaths (Setel et al. 2007). Medical certification of death (MCOD) involves a medically qualified doctor assigning each death an underlying cause, and a trained coder classifying the death according to the International Classification of Diseases (ICD). In addition to providing essential data on vital statistics, civil registration also serves important public functions such as the establishment of legal identities, which has significant implications for human rights (Mahapatra et al. 2007).

Notwithstanding the well documented social benefits of CRVS (Mahapatra et al. 2007; Lopez et al. 2007; Hill et al. 2007), very few low- and middle-income countries are making substantive progress in advancing this agenda (Hill et al. 2007; Lopez et al. 2011). Most continue to invest in interim measures, such as surveys, to obtain mortality rates and cause-of-death distributions. These interim measures to collect data on vital statistics require funding. However, the decision to invest scarce resources into particular DCMs is often made without even a basic understanding of the cost implications and relative benefits.

To date there have been no documented attempts to undertake a formal comparative assessment of costs and outcomes/benefits of the various methods available to collect vital statistics. In this paper, we illustrate that this is partly explained by the intangibility of the expected outputs (i.e. how do we measure the quantity and the quality of data) and the lack of costing data. Our main objective in this study was to propose a systematic framework for such a comparative assessment, which can provide a starting point for discussions among vital statistics experts, economists and practitioners.

The paper is organised as follows. In Section 2 we present the main findings of our review of the literature on the available DCMs for vital statistics. In Section 3 we briefly describe alternative methods developed in economics to systematically evaluate costs and outcomes, and then introduce the main elements of our proposed framework (the Appendix provides the technical details). In Section 4 we use a simulation exercise to build a stylised scenario of a hypothetical low-income country and illustrate how the proposed framework might operate in-country. We also present results for an alternative scenario. In Section 5 we outline our conclusions.

Literature review on data collection methods for vital statistics

The preferred DCM for vital statistics is a complete system of CRVS that includes MCOD for all deaths (Setel et al. 2007). However, reaching this goal in all countries in a timely fashion, especially in developing countries, will be extremely difficult. The number of countries reporting the registration of more than 90 per cent of all in-country deaths using CRVS increased by only seven from the 1970s to the 1990s (Mahapatra et al. 2007).

Complementary and interim measures therefore need to obtain mortality rates and COD distributions (Hill et al. 2007). We reviewed the literature to identify various DCMs for vital statistics, existing frameworks to assess their quality and any documentation of their associated costs. We searched the MEDLINE and SCOPUS databases using a combination of the MESH terms 'vital statistics' and 'data collection'. We then examined the resulting 256 references published since 1970 for relevance. We checked the reference lists of selected articles to ensure that we had included all relevant papers. We identified three major categories of DCMs for obtaining mortality information when a complete CRVS is not available. These are described below:

i. Partial registration is the registration of births and deaths within a subset of the population. Partial registration is used when there is either incomplete introduction of CRVS or when parts of the population are deliberately sampled to create representative figures.
Rural, remote and disadvantaged populations tend to be under-represented in systems where civil registration exists but is not yet complete (Hill et al. 2007). To use the data from these systems, it is necessary to estimate the coverage and completeness of registration in order to correct for these biases.
In situations where coverage of the total population is difficult, and not likely to be achievable in the near future, it is possible to implement a Sample Registration System (SRS). Such a system involves covering a nationally representative subset of the population by a complete registration system within particular sites (Mozumder et al. 1990; Setel et al. 2005). India's SRS system (Mari Bhat 2002) and China's Disease Surveillance Points (DSP) system (Rao et al. 2005) are examples of this method. In settings where even an SRS has been too difficult to implement, registration can be set up in Demographic Surveillance Sites (DSS) that are representative of a major typology (Pena et al. 2008). The accuracy of data gathered by these methods will depend on the completeness of registration in the target population. Measuring the completeness of a registration system can be difficult and two general methods are available. Cross matching data between multiple sources, known as the 'capture recapture' method (Chao et al. 2001), can be used if there are at least two complete or near-complete datasets, such as a population census and a vital registration system (Vapattanawong & Prasartkul 2011). The second method is based on a comparison of the age distribution of recorded deaths with the age distribution of the population in which the deaths occurred (Murray et al. 2010). Based on their analysis, Murray et al. (2010) recommend three death distribution approaches for these completeness studies: the Generalised Growth Balance (GGB) (Hill 1987), the Synthetic Extinct Generations (SEG) (Bennett & Horiuchi 1984) and combined models using GGB and SEG.

ii. Censuses and population surveys are the second major source of mortality data. Censuses are conducted every 10 years and provide an almost complete record of the total population at the time of the survey (Ruzicka & Lopez 1990). Population surveys, such as the Demographic and Health Survey (DHS), are conducted more frequently (approximately every five years) but survey only a representative sample of the population.
Depending on the questions asked in the survey, mortality rates can be determined either directly or indirectly. Direct methods include asking the respondent about deaths that have occurred in the family in the inter-censual period. Indirect methods include the children-ever-born technique (for estimating deaths under five years) (Rajaratnam et al. 2010) and the sibling survival method (for estimating deaths between 15 and 59) (Obermeyer et al. 2010; Hill et al. 2006). In all cases, death rates can be subject to recall bias and other measurement errors in the reporting (Boerma & Stansfield 2007).

iii. Facility-based data collection relies on recording deaths that occur in health facilities (Hill et al. 2007). These sources of data are heavily dependent on the proportion of the population that use the facilities and are not usually accurate sources for mortality rates, unless the coverage of health services is high (Boerma & Stansfield 2007). Data quality is also a concern because the staff may not be trained in standards for mortality recording (AbouZahr et al. 2007).

Different methods for estimating the COD distribution may be feasible depending on the data collection system implemented. The three ways of assigning COD are presented below, from most to least reliable (Hill et al. 2007).

i. Medical certification of death (MCOD) requires doctors to complete death certificates specifying the COD (AbouZahr et al. 2007). Consequently, it is only feasible in systems where deaths are continually recorded. To generate data of high quality, diagnoses should be compatible with recent revisions of the ICD, limit the number of 'improbable' diagnoses and avoid overuse of generic classifications such as 'undefined' (Mathers et al. 2005).

ii. Verbal autopsy (VA) is used when medical certification of deaths is not feasible, such as where individuals generally die at home without having sought medical attention (Fottrell & Byass 2010). Data is gathered by interviewing families about the circumstances surrounding the death and then using this information to generate a probable broad COD, either through review by medical professionals or through diagnostic algorithms (Soleman, Chandramohan & Shibuya 2006). The validity of a given diagnosis can vary as it affected by recall and cultural sensitivities. VA may be used in systems that collect data both continually and retrospectively; however, if used in conjunction with MCOD, consistency in reporting will be required.

iii. Population-based modelling, the least reliable of the methods presented, may be used where formal recording of causes of death is not feasible. Data regarding risk factors, intervention coverage and age–sex patterns of mortality are used to generate estimates of mortality due to particular causes. Estimates generated using these methods are generally less accurate due to their reliance on assumptions rather than collected data (Johnson et al. 2010)

Ruzicka and Lopez (1990) proposed the first framework evaluating quality of mortality and COD data. It was expanded and applied to contexts such as the SRS of India (Mahapatra & Chalapati Rao 2001) and China's DSP (Mari Bhat 2002). Building on this work, Mahapatra et al. (2007) developed an assessment framework for vital statistics from civil registration systems, which is also consistent with the criteria used to evaluate the quality of official statistics. The authors used the proposed framework to classify countries according to their quality of COD statistics reported to WHO.

The framework has five quality attributes:

- **accuracy**, determined by issues such as coverage, completeness, missing data, use of ill-defined or improbable categories and consistency with other estimates
- **relevance**, measured by the applicability of the data to age, sex and geographical groupings
- **comparability**, encompassing standard coding practices and definitions
- **timeliness**, measured by the production time and regularity of data releases
- **accessibility**, the ease of access to data.

In addition to evaluating the quality of mortality and COD data, it is also important to assess the comparative costs of using different DCMs. However, these costs are not well documented. The only available information relates to costing estimates of specific surveys or surveillance sites design, and annual budget allocations for DCMs available from country offices. However, this information is not comprehensive, and costs are not regularly reported in a way that enables comparability. Cross-country analysis is also more challenging because costing DCMs varies considerably from country to country depending on the manner in which they are administered (AbouZahr et al. 2007).

Indeed, for most DCMs, data on births, deaths and cause of death are not the only (or even the primary) objective of collection. As a result of multiple objectives, the cost of the data for vital events may only be a fraction of the total cost. For example, vital statistics is only a small objective of a survey in health outcomes and service use, but a major focus for a disease surveillance system. As a result, when assessing the cost of collecting vital statistics, we should only factor in the module dealing directly with vital events.

Additionally, DCMs tend to have a very different cost structure in relation to fixed and variable costs. Periodical surveys require one-off investments, whereas facility-based data collection or incomplete civil registration requires regular maintenance in addition to relatively large set-up costs. This information will need to be factored into the apportioning rule for total cost.

The only study documenting the costs and performance of information systems was in the early 2000s in Tanzania (Rommelmann et al. 2005a, 2005b). However, the authors reported numerous information gaps and did not propose a systematic framework to comparatively assess costs and outcomes of the reviewed methods. The study therefore does not constitute a formal evaluation of costs and outcomes, nor does it attempt a CEA.

A systematic assessment of data collection methods – the economics approach

The literature review reveals the absence of formal comparative assessment of costs and outcomes/benefits of competing DCMs. There appears to be no documented attempt to examine and ultimately determine the most efficient method to collect vital statistics data; that is, the optimal mixture of quality and cost. The literature does provide a framework to assess the quality of vital statistics data (Mahapatra et al. 2007); however, this framework still needs to be operationalised so that it can be used to measure the quality of data of available DCMs. On the other hand, no previous study has attempted to systematically measure the cost of collecting vital statistics data.

Two approaches – economic evaluation and efficiency analysis

Taking into consideration the gap in the vital statistics literature, we then turned to those areas of economics which are directly concerned with the systematic assessment of both costs and outcomes/consequences: economic evaluation and efficiency literature. Our purpose was not to examine this prolific literature in detail. Instead, we aimed to identify concrete economic approaches and tools that can be used to undertake an economic evaluation of DCMs with a view to inform investment decisions on DCMs.

Economic evaluation

Cost–benefit analysis, cost minimisation analysis and cost-effectiveness analysis, the three major methods of economic evaluation (Drummond et al. 2005), are briefly summarised below.

- **Cost–benefit analysis** (CBA) is used to measure and compare the social benefits and costs of a particular project or program. CBA uses the net benefit or benefit-cost ratio to rank comparable options (Drummond et al. 2005; Campbell & Brown 2004) and requires that all costs and consequences are expressed in the same monetary units. CBA allows the comparison of costs and outcomes of different alternatives by assigning a dollar value to them. Although widely applied and used by governments around the world, CBA is not directly applicable to the evaluation of DCMs. CBA would require the analyst to assign a dollar value to the benefits of obtaining vital statistics data. In our case, this is clearly unfeasible as it would require measuring the impact that good quality statistics of vital events have on decision-making (i.e. was the data used to inform specific decision-making and if yes, what was their contribution?) and health outcomes (i.e. what proportion of the improvements in health can be attributed to those evidence-based decisions). Additionally, it could be argued that in the case of vital statistics data, there is social consensus of their social added value, which does not need to be established via economic evaluation. In this case, using CBA becomes redundant.
- **Cost minimisation analysis** aims to identify the least costly alternative to achieving the same outcome (Drummond et al. 2005; Muennig 2008). It requires that the alternatives under assessment have the same outcomes (e.g. apply the same questionnaire on vital statistics to 20 000 households) and only differ by inputs/costs (i.e. interviewers filling out paper questionnaires vs interviewers entering data into their computers). Cost minimisation analysis is not applicable in our context since different DCMs have different outcomes (e.g. surveys produce vital statistics representative at national level but not at district level, whereas the opposite might be true for a disease surveillance system).
- **Cost-effectiveness analysis** (CEA) is similar to CBA in that it aims to assess both the costs and outcomes of competing alternatives; however, it does not rely on monetary units to measure the latter (Drummond et al. 2005; Muennig 2008). CEA is used when the objective is to compare options that have similar outcomes/outputs. This is one of the most commonly used economic evaluation methods in health, and analysts usually rely on statistics, such as number of life years gained, to measure the outcomes of the health interventions under assessment. In many instances, the health interventions being evaluated have very different health benefits/outcomes (e.g. heart surgery vs a diabetes treatment). The analyst would then need to develop or rely on indicators, such as quality-adjusted life-year (QALY), which standardise the outcomes to be evaluated and capture the health benefits.[1]

We are in a similar situation when evaluating alternative DCMs. These methods produce different types of vital statistics data of varying quality, according to the different criteria as outlined above. Therefore, to enable the systematic comparison of DCMs outcomes via CEA, one would need to develop a single index that

1 This variant of CEA is commonly referred to as cost utility analysis (CUA). CUA was developed to help decision-makers assess the value of alternative interventions that have very different health benefits.

could then be used as a standardised metric of the DCMs outcomes/consequences. Such an index, similar to QALY, needs to capture both the quality and the quantity of the vital statistics produced by each DCM. We should note that in evaluating health interventions, the number of deaths averted, without adjustments for the quality of life, is sometimes used. However, in our case, an index that does not capture quality would be quite problematic, as it would imply that large amounts of data of very poor quality—and consequently, of limited use—are preferable to smaller amounts of data of very high quality. On the other hand, an index that does not capture quantity would fail to provide us with appropriate physical units to measure outcomes. The latter would be equivalent to trying to quantify the consequences of health interventions by using quality of life, but without measuring the number of years of life gained.

Efficiency analysis

The benchmarking and efficiency analysis literature (Coelli et al. 2005) is also relevant to our research. It looks at measuring outputs, inputs and the productivity of production units (e.g. firms, hospitals, schools, government offices etc.), and uses econometric techniques to estimate and compare the productivity and efficiency of those units. Efficiency in this literature is defined as 'the ability to achieve maximal outputs from a given set of inputs'. We frame our evaluation problem as follows: (a) Each DCM can be considered a unit of production, (b) DCMs produce the outputs of 'good quality vital events data', in which the 'good quality' of data can be measured by different quality attributes and the output is data, (c) The inputs used by each DCM are defined as the resources required to collect the relevant data. For example, human resources, facilities and equipment, and those inputs can be measured in monetary terms and so can be aggregated into a single item of input, which is the total cost associated with each DCM, (d) We use outputs (quantity and quality of data produced) and inputs (costs) to evaluate the productivity of each DCM; that is, how much 'good data' each DCM can produce given the associated cost. Each DCM has a maximum productivity; that is, the highest possible quality data it can achieve given the amount of resources. Therefore, we can rank DCMs from best (most productive) to worst (least productive). In EA, we are thus interested in the relative rank of DCMs in terms of their productivity or cost efficiency.

Main elements of an economic assessment of data collection methods

Both, CEA and EA provide suitable frameworks to evaluate the costs and outcomes of alternative DCMs. They involve three basic elements: (1) identifying the units (i.e. DCMs) to be evaluated, (2) identifying and measuring the outcomes of each DCM and (3) identifying and measuring their associated costs. Each element is discussed in detail in the following sections.

Identify the alternatives to be evaluated

Table 1 summarises our review of the literature to identify alternative systems for collecting data on vital events. Our question is then, should all the identified systems be included in the evaluation? The answer depends on the government's policy in relation to a CRVS. In the first scenario (Scenario A), the long-term policy is to work towards the implementation of a CRVS (i.e. due to the associated social benefits, including human rights), acknowledging that interim measures are required to provide vital statistics data while the system is put in place. In such a case, CRVS should be excluded from the evaluation. Additionally, the criteria used to assess the interim DCM for vital statistics should include the extent to which they enable or undermine the long-term goal of establishing a CRVS. In the second scenario (Scenario B) where no such policy is in place, a CRVS can be included as one of the DCMs to be evaluated. However, since the most important outcomes/benefits of CRVS are not to produce vital statistics or even data but to produce legal identification (for birth, death, marriage etc.), it is important to stress that the evaluation does not assess CRVS as a whole, only the data component. This should also be reflected in the assessment of costs, as discussed below.

Table 1: Data collection methods for vital statistics

	DCM	Description
Complete or partial registration systems	Civil registration	Births and deaths in the population are continuously recorded. Deaths are recorded with a medically certified COD. Yearly statistics are generated based on this data.
	Sample registration	Births and deaths in a representative sample of the population are continuously recorded. In some systems, deaths in the sample are recorded with MCOD. In other cases, deaths in the sample are recorded with a COD assigned using VA. Depending on the method, the resulting data may be grouped either by broad or specific causes of death. Yearly statistics are generated based on this data.
	Demographic surveillance sites	Births and deaths in a non-representative sample of the population are continuously recorded. Deaths in the sample are recorded with a COD assigned using VA. Depending on the method, the resulting data may be grouped either by broad or specific causes of death. Yearly statistics are generated based on this data.

Table 1 – continued

	DCM	Description
Census and surveys	Population census	All households are queried regarding current occupants, as well as details of recent births and deaths. For a system with full VA; that is, a COD exists for each recorded death, a VA questionnaire is used to assign a COD. For a system with partial VA COD, for a representative sample of recorded deaths, a VA questionnaire is used to assign a COD. Alternatively COD distribution may be generated through modelling based on age–sex patterns, prevalence of risk factors and intervention coverage. Statistics are usually generated every 10 years.
	National-level household survey (direct estimates)	A representative sample of households is queried regarding current occupants, as well as details of recent births and deaths. For a survey using VA COD, for each recorded death a VA questionnaire is used to assign a COD. Alternatively, COD distribution may also be generated through modelling based on age–sex patterns, prevalence of risk factors and intervention coverage. Statistics are usually generated every three to five years.
	National-level household survey (indirect estimates)	A representative sample of households is queried regarding current occupants, as well as survival status of siblings and/or children. COD distribution is generated through modelling based on age–sex patterns, prevalence of risk factors and intervention coverage. Statistics are usually generated every three to five years.
	Sub-national-level household survey	A sample of households is queried regarding current occupants, as well as details of recent births and deaths. For each recorded death, a VA questionnaire is used to assign a COD. Statistics are usually generated every three to five years.
Facility-based collection	Facility-based reporting (wide scale)	Births and deaths that occur within medical facilities are continuously recorded. Deaths are recorded with MCOD. Yearly statistics are generated based on this data.
	Facility-based reporting (sentinel sites)	Births and deaths that occur within a representative subset of medical facilities are continuously recorded. Deaths are recorded with MCOD. Yearly statistics are generated based on this data.

Identify and measure outcomes of data collection methods

In general, the aim of DCMs is to produce good quality data for estimating vital statistics. A sensible metric for this outcome would need to incorporate both the quality and quantity dimensions of the data. Intuitively, we know that in measuring the outcomes of the various methods used to collect vital statistics, we are concerned about the quality of those statistics. However, it is less clear how to count those statistics.

Measuring outcomes: quantity

One of the basic elements of a sound economic evaluation is the choice of appropriate physical units to measure outcomes (Drummond et al. 2005) (i.e. mortality can be measured as number of deaths averted). Therefore, any proposed metric for the outcomes of DCMs would need to incorporate, in addition to quality, the quantity dimensions of the data.

Unfortunately, there is no clear guidance regarding the appropriate units to measure the 'quantity of data' produced by DCMs. For example, should we use the number of vital events that get registered as units of measurement. This seems reasonable when quantifying the outcomes of the CRVS. Or should we count the number of people represented in the statistics produced by each method? This might seem reasonable for household surveys.

We propose that the choice of units of measurement is informed by the government policy that relates to CRVS. In Scenario A, where the long-term policy is to establish CRVS, we propose that 'quantity of data' is defined as 'the unit records[2] collected by each DCM'. Such a metric would be in line with the long-term policy of CRVS, which would seek to collect all unit records of vital statistics for a country's population. In this case, the number of unit records could be approximated by the sample size of the DCM.

On the other hand, in Scenario B, where CRVS is not part of the government's long-term policy, it seems reasonable to assume that the government does not place high value on obtaining unit records for all its citizens and so a different metric should be considered. In this event, a metric such as 'the number of people represented by each DCM' might be more suitable and in line with the need to produce vital statistics that are representative of the population, and that can be easily obtained from the documentation commonly provided with each DCM.[3]

Measuring outcomes: quality

The first step in a systematic assessment of quality is to clearly identify those attributes (e.g. relevance and timeliness) that define data quality for vital statistics. The second step would be to operationalise those attributes so that they can be measured for each DCM. That is, for each attribute we need to define clear evaluation criteria (e.g. for timeliness, the evaluation criteria may be the production time and regularity) and provide an adequate measurement scale.

Mahapatra et al. (2007) identified an initial list of quality attributes of vital statistics data and the corresponding evaluation criteria. We have split each of their five assessment categories 'accuracy, relevance, comparability, timeliness and accessibility' into criteria that assess the collection of statistics related to vital events (i.e. births and deaths) and the collection of statistics related to COD (see Table 2). As noted earlier, the quality attributes proposed by Mahapatra et al. (2007) are in line with those included in quality frameworks for official statistics. Thus, although originally developed for CRVS, it seems reasonable to expect that criteria based on this framework can be applied to other methods of collecting vital statistics. However, we need to address three important limitations to this framework in relation to our study.

First, as discussed previously, the scope of the evaluation is defined by the government's policy on CRVS. If the long-term policy is the establishment of a CRVS, then the criteria used to evaluate the interim methods used to collect data on vital statistics should include the extent to which they enable or undermine such a policy. Enabling or undermining the long-term policy of setting up CRVS is not an attribute of the quality of the data produced

2 Each record represents an event (often the case for administrative data) or a person's specific details (often the case for population data).

3 We should note that although small-area estimates are not accounted for by the size of the population represented by the data, they are accounted for in the quality index proposed.

and so is not included in the Mahapatra et al. (2007) framework. It is an effect, however, and a very important one, of alternative methods of collecting vital statistics data. Therefore, we suggest that 'improvement towards CRVS' is added as one of the criteria used to evaluate the 'quality' of the outcomes of DCM.

Second, to provide a numeric measure of quality, we need to establish an appropriate ordinal scale of the potential scores that can be assigned under each criterion. This would allow us to assess each DCM under each quality criterion and provide comparative scores to rank the different DCMs. In our stylised examples described in Section 4, we opt for a range between 0 and 10, in which 0 indicates 'no degree of quality achieved' and 10 indicates 'the highest possible quality achieved'. We stress that this range is indicative only and such information should be produced by consensus among a team of vital statistics experts. Ideally, the same team would assess each DCM and provide scores under each criterion.

Third, since we have multiple criteria, we could reasonably expect that some methods are superior to others when ranked by a particular quality criterion, but not by others. For example, we could expect national representative demographic surveys to cover a wider selection of the target population, but be less detailed in regards to age and COD compared to demographic surveillance systems. Unless the corresponding quality scores can be aggregated in a single index, we will not be able to rank the DCMs based on their outputs. Therefore, we have identified three alternative methods that can be used to aggregate the scores into a single index of quality of DCM. Table 3 presents a brief description, and the main advantages and disadvantages of each method.

Table 2: Assessment framework for vital statistics[1]

Criteria	General vital statistics	Cause-of-death statistics
Accuracy (A)		
	A1- Accuracy of vital event statistics	*A2 - Accuracy of cause-of-death statistics*
Coverage	% of population living in areas where vital event recording occurs	% of population living in areas where COD recording occurs
Completeness	% of events contributing to fertility/ mortality statistics	% of deaths with appropriately certified COD
Missing data	% of key variables with response not stated	% of COD reports for which age–sex data are missing
Use of ill-defined categories	Not applicable	% of deaths classified under various miscellaneous and ill-defined categories
Improbable classifications		Number of deaths assigned to improbable age or sex categories per 100 000 coded deaths
Consistency between cause of death and general mortality		% of COD data points deviating more than 2 (or 3) SDs from general mortality-based predictions
Relevance (R)		
	R1- Relevance of vital event statistics	*R2 - Relevance of cause-of-death statistics*
Routine tabulations	By sex and five-year age groups, based on place of usual residence. Deaths in children under five years tabulated by 0 and 1–4 year age group.	By sex, and at least by eight broad age groups; namely, 0, 1–4, 5–14, 15–29, 30–44, 45–59, 60–69 and 70+ years

[4] Adapted from Table 1, in Mahapatra et al. 2007 (p. 1654)

Table 2 – continued

Criteria	General vital statistics	Cause-of-death statistics
Small-area statistics	Number of vital event tabulation areas per million population	Number of COD tabulation areas per million population
Comparability (C)		
	C1- Comparability of vital event statistics	*C2 - Comparability of cause-of-death statistics*
Over time	Stability of key definitions over time	Consistency in the proportions of cause-specific mortality over consecutive years
Across space	Uniformity of definitions across areas	ICD to certify and code deaths, revision used and code level to which tabulations are published
Timeliness (T)		
	T1 - Timeliness of vital event statistics	*T2 - Timeliness of cause-of-death statistics*
Production time	Mean time from end of reference period to publication	Mean time from end of reference period to publication
Regularity	SD of production time	SD of production time
Accessibility (AC)		
	AC1 - Accessibility of vital event statistics	*AC2 - Accessibility of cause-of-death statistics*
Media	Number of formats in which data are released	Number of formats in which data are released
Metadata	Availability and quality of documentation	Availability and quality of documentation
User service	Availability and responsiveness of user service	Availability and responsiveness of user service
SD= standard deviation		

Although the use of unweighted and weighted averages (where the weights are provided by experts) is relatively straightforward, the creation of a data-driven weighted index using data envelopment analysis (DEA) is more involved. We provide more detailed information about this method and its operation in the Appendix.

Given how little we know about the costs associated with DCMs, an important first step would be for countries to start measuring the cost of their systems. The two basic approaches for measuring costs are bottom up and top down.

Table 3: Alternative methods for aggregating the quality scores into a single index

Methods	Description	Disadvantages	Advantages
Unweighted average –	Sum of all scores divided by the total number of criteria	Assumes all quality attributes are equally important in all circumstances and all settings	Calculation simplicity Does not require a priori knowledge about the relative importance of different quality criteria Aggregation index for individual DCM is independent from the rest
Weighted average – by expert opinion	Weights assigned to each criterion are identified a priori by experts.	The weights are subjective and expert consensus might be difficult to achieve.	Calculation simplicity Reflects the relative importance of different quality criteria, which might vary by setting
DEA-based aggregation[1]	Weights are assigned via an optimisation problem that includes all DCMs to be evaluated (see the Appendix). Used in several applications of multi-criteria decision analysis (MCDA) in a wide range of policy areas, including health.	Requires a large number of observations. If only a handful of DCMs are evaluated, it would require several assessments (i.e. by various experts) of each DCM.	Does not require a priori knowledge about the relative importance of the quality criteria. Data driven No systematic bias toward any quality criteria. Might be useful when there are conflicting opinions on the relative importance of each quality criteria.

Identify and measure cost of data collection methods

There is a vast amount of literature on methodologies for costing exercises, particularly in health services (Drummond et al. 2005). However, the absence of costing data and appropriate costing frameworks for DCM have been identified as a hindrance for development partners and governments trying to make informed decisions on investments in DCM (Cambridge Economic Policy Associates 2013).

Bottom-up costing relies on the 'ingredients approach', whereby the resource items are identified and information on the associated quantities and unit prices is used to obtain the total cost of collecting vital statistics data. The costing instrument would then need to include the cost items, classified in standard and relevant categories such as capital (e.g. buildings, vehicles and equipment) and recurrent costs (e.g. personnel, consumable goods and transportation – car, rental, maintenance and petrol) to survey sites. The level of aggregation would depend on the information available

and the cost structure of each DCM. For example, for surveys, personnel might need to be subdivided into consultants, field supervisors, interviewers, translators and data-entry operators; whereas equipment might include only computers and printers. For each item, appropriate units of quantity (e.g. for field supervisors, number of full-time equivalent person months) and prices (e.g. for field supervisors, their unit prices might include salary, allowances and per-diems) would need to be determined. This information will then be used to estimate total costs for each item (unit prices times quantity) and total costs for producing vital statistics (i.e. the sum of the costs estimated for each item). Bottom-up costing is usually the preferred method, since it can produce relatively accurate estimates (Rommelmann et al. 2005a; Mogyorosy & Smith 2005). However, it is not always feasible. For example, we could reasonably expect that data for bottom-up costing can be easily obtained for surveys, but not for CRVS, which would most likely require a top-down approach. A recent exercise in South Africa illustrates the practical and methodological challenges in capturing even basic cost elements of CRVS (Cambridge Economic Policy Associates 2013). The report authors found that government estimates of direct costs involved in registering a birth were available for staff time and consumables, but not for indirect costs and overheads, which were likely to be substantial.

Top-down costing relies on first estimating the total costs of a system, usually from budget documents (i.e. the total cost of operating the CRVS) and then using appropriate allocation rules to allocate total costs to individual functions (i.e. which proportion of the CRVS costs should be allocated to producing vital statistics). The allocation or apportioning rules chosen will have a significant impact on the costing estimates. They should thus be transparent, documented in detail and follow a clear logic that is easy to justify. Most importantly, the rules should consider the various outputs produced by each DCM and provide a clear justification as to why a relatively large or small proportion of costs should be allocated to the production of vital statistics. For DCMs such as SRS, for which the primary goal is to collect vital statistics, it would be reasonable to assume that most program costs can be attributed to collecting such data. However, this is clearly not the case for other DCMs that produce many outputs, only one of them being vital statistics.

Indeed, most DCMs produce many outputs, only one of them being vital statistics. Therefore a combination of bottom-up and top-down costing would most likely be required. For instance, we could use bottom-up costing to estimate the total cost of a demographic health survey. However, it collects data on many variables, so we would need to use some apportioning rule to allocate a proportion of those costs to the collection of vital statistics data. In this case, we could for example use as the apportioning rule the proportion of questions in the survey that relate to deaths, births and causes of death. If we have reason to believe that some questions will be far more time consuming than others, we could then try to obtain the proportion of the interviewers' time spent on vital statistics questions and use it as the apportioning rule.

Unfortunately, there is no straightforward method to determine the apportioning rule. In addition, the more complex the DCM, and the greater and more varied the outputs it produces, the more difficult the task becomes. For example, as previously mentioned, for civil registration it appears that producing vital statistics is a byproduct of a system that produces other outputs of high social value (e.g. human rights, legal documentation and identification of beneficiaries of government programs). The percentage of the CRVS's costs allocated to producing statistics would largely depend on the relative importance of vital statistics versus other outputs of the CRVS. Similar issues arise in the collection of data through medical records. As vital statistics collection is embedded in existing facility systems and is not the primary goal of medical records, the costs associated with a medical records system will also need to be apportioned. Some of the apportioning rules will need to take into account the proportion of records dealing with births or deaths. An informed estimate based on the identification of cost-sharing activities will be required to allocate a proportion of costs associated with these records to vital statistics.

Disregarding the costing approach used (bottom up vs top down), we should stress that both CEA and EA rely on a systematic framework of comparatively assessing costs for all DCMs being evaluated. It is thus important to ensure that not only the tasks and associated resources required to produce vital statistics in each DCM are included in the analysis, but

also that we are comparing like with like. For instance, survey-based DCMs tend to follow four main phases: (i) preparation, including survey design, sampling, training and piloting; (ii) survey implementation; (iii) data processing and report writing; and (iv) results dissemination. If, for example, we decide to include the cost of analysing the survey data and disseminating the results, we would then need to ensure that the cost of these types of activities is also included when examining other DCMs.

The evidence on costing of DCMs is so limited that little guidance exists beyond the general costing textbooks and manuals (Rommelman 2005a; Mogyorosy & Smith 2005). Additionally, estimating costs when limited or aggregated data only are available, usually involves a considerable number of assumptions and subjective modelling decisions. Therefore, in order to facilitate the replicability of the results and the rigour of the costing exercise, it is important to be as systematic as possible and have detailed documentation of the assumptions used and their justification.

Understanding the costing structure of each DCM (i.e. how much it is costing to collect the observed amounts and quality of vital statistics data) is an important first step. We are also interested in the cost functions of each DCM (i.e. how costs change with changes in output, output being the number of events being recorded or coverage).

Although we know little about the cost functions of the DCMs, there are a few reasonable assumptions that we can make. First, it is likely that those costs would be non-linear and discontinuous. For example, it is reasonable to expect that the costs of CRVS will not increase in a constant proportion to the number of events recorded (i.e. the costs of recording 80% of the population vital events will not be eight times the costs of covering 10%). We can also expect that there will be thresholds/points at which large additional fixed costs are required (i.e. once a certain level of coverage is achieved, significant investments in information technology and a dedicated team of analysts may be required). Second, new technologies are likely to have substantive impacts on the expected costs and feasibility of DCMs, such as CRVS, which are difficult to predict. This is illustrated by the example of the highly ambitious 'Unique Identification' scheme launched in India in 2010, which could in the future be linked to CRVS. Notwithstanding the challenges, this scheme seems to be paving the way for the creation of unduplicated identity numbers linked to basic biometric and demographic data at a reported cost per person enrolled of US$2 (The Economist 2012). Third, from what we know about DCMs, the relationship between costs and output will also likely vary from one DCM to the next. For example, for a survey it is reasonable to assume that the population sample might be the main cost driver, but this is unlikely to be the case for sample registration, where the number of sample units might be the main cost driver. Fourth, cost functions will likely be highly context specific and so will vary significantly between countries. Fifth, in some instances where there are regional variations in coverage (e.g. facility-based reporting covers 80% of the population in a particular province, but only 30% and 10% in others), this information might be used to approximate the cost function of a particular DCM in that country. In this event, it will be possible to predict how costs will change as coverage/number of events recorded increase in that particular setting. However, in many cases this information will not be available.

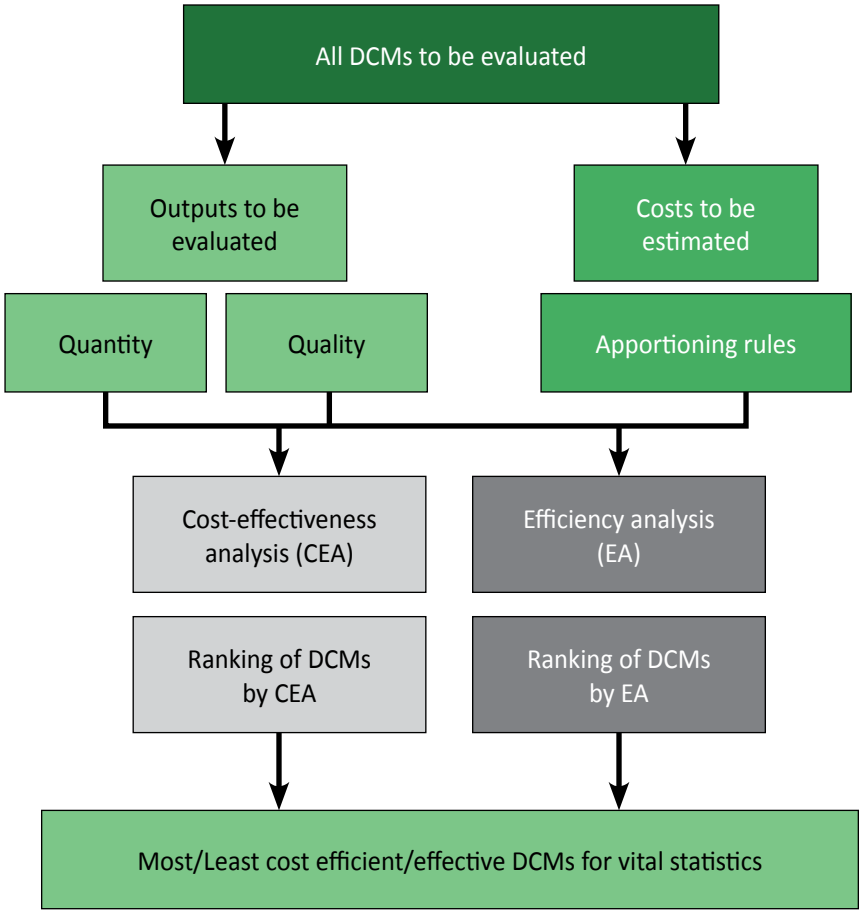
Since we do not have information on the parameters of the cost functions of alternative DCMs, we will only be able to estimate the costs associated with current coverage. This means, we will be able to answer the question of 'which DCM is currently providing the best vital statistics data given the associated costs', but not the question of 'which DCM will be the most efficient to achieve x number of records or y per cent of coverage'. However, if countries understand the current costing structure of their alternative DCMs, this information can be used to examine, at least in hypothetical scenarios, what the most efficient DCM will be to achieve the targeted coverage. To do so, the parameters of the current costing structure can be used to build basic simulation models that also rely on specific assumptions about threshold costs. Other parameters can be used to predict the future costs of scaling-up alternative DCMs. This type of simulation has been used in other areas of policymaking, including health, where simulation models aid decision-making (Jimenez et al. 2012; Winfrey, McKinnon & Stover 2011). However, even the most basic simulation model would require information on the current costing structure of each DCM.

Stylised Scenario A – Assumes a long-term policy of establishing civil registration for vital statistics

In this section, we use a stylised scenario (Scenario A) to demonstrate how all elements of the systematic assessment previously discussed can be applied in-country. A stylised scenario is a simplified presentation of a theoretical concept or model. It is a broad generalisation of reality, which summarises some complicated process or pattern of behaviour. Scenario A represents a hypothetical low-income country with a population of 33 million which has developed DCMs of varying quality operating in parallel. We thus assume that substantial investments have been made in those systems and policymakers are interested in assessing which DCMs are providing the best value for money.

Figure 1 outlines the main steps of evaluation in Scenario A. They resemble the steps for doing this type of economic evaluation in a particular setting. To plausibly approximate the total costs of various DCMs for vital statistics used in Scenario A, we sourced information from the only available study on costing government information systems for social policy (Rommelmann et al. 2005a, 2005b). However, the assumptions used in apportioning those costs are entirely hypothetical but designed to be reasonable. Likewise, the quality scores assigned to each DCM are also hypothetical and based on our assumptions of what might be plausible in such a hypothetical low–middle income country. As a result, our exercise does not have implications for a specific country or setting.

Figure 1: Systematic assessment of data collection methods (DCMs) – the economics approach



Which data collection methods should be evaluated?

We first need to clearly define what is going to be evaluated (i.e. which DCMs for vital statistics), and clearly define the output and costs for each alternative under assessment.

Here we are interested in evaluating the alternative methods currently available to collect vital statistics data. Our hypothetical country has an established CRVS but it remains incomplete. We therefore assume that policymakers are pursuing the long-term policy of establishing a complete CRVS, that is, our evaluation takes place in Scenario A where:

- CRVS is the gold standard to which the country aspires
- CRVS is excluded from the list of DCMs to be evaluated
- a quality indicator is included that captures the extent to which intermediate DCMs are stepping stones towards CRVS
- the units for measuring the quantity of data collected are 'unit records' (events) approximated by sample size/number of participants.

However, to illustrate the differences in results that can be expected when the policy question changes, we also conducted the analysis of Scenario B, where there is no long-term policy of establishing CRVS. In such a situation, the CRVS should be evaluated just like other DCMs.

In addition to CRVSs excluded from the evaluation, our hypothetical country has nine other health and demographic information systems collecting vital statistics data. They are briefly described in Table 4. The objectives of all potential data collection systems in-country would need to be reviewed to decide whether they are collecting vital statistics data and should be included in the evaluation.[4] Additionally, when assessing the objectives of alternative DCMs, it is important to note whether collecting vital statistics data is the main or secondary objective of each DCM. As discussed above, the assumptions used for the apportioning rules of costs should take into consideration the various objectives and outputs, besides vital statistics data, of each DCM.

4 Note that when undertaking this exercise, the country should organise a panel of experts to discuss the feasibility of the evaluation, map all the potential DCMs and then identify those with a primary or secondary objective to collect vital statistics data.

What are the outputs to be evaluated?

Under Scenario A, the long-term policy of establishing CRVS led us to use 'events' (or 'unit records') as the measure of the quantity of output for each DCM, whereas the quality indicator used should capture the extent to which intermediate DCM can be built on for the long-term improvement of CRVS. We approximate 'unit records' by the sample size (last column in Table 4) and use the assessment framework developed by Mahapatra et al. (2007) (see Table 2) to identify the various quality attributes to be measured.

When measuring the quality attributes of each DCM, we require a scoring matrix with the range of potential scores for each attribute. We will use this matrix to measure the extent to which vital statistics data produced by alternative DCMs meet each of the criteria used in the evaluation. Therefore, it is important to ensure that the range of scores and the corresponding cut-offs assigned differentiate the degrees of quality (i.e. very poor, poor, average, good, optimum). Table 5 in Mahapatra et al. (2007) provides an example of cut-offs used to rank CRVS systems. In country applications, it might be advisable for the quality scoring matrix to be developed by expert consensus. One option is developing global standard quality scoring matrices for various typologies of countries. This would facilitate cross-country comparisons and save time and resources.

To implement a straightforward approach for Scenario A and to facilitate discussion, we assumed a basic linear scoring system with a range of [0–10] applied to each quality criteria described in Table 2. For instance, we defined the score range for coverage, as 0 = 0% (or indicators not collected), 1 = 0+ to 10%, 2 = 10+ to 20%, etc. and 10 = 90+ to 100%. So, if a particular DCM in our hypothetical country covers 5 million people, then its coverage rate would be 15 per cent as the coverage target is the country population (33 million) and therefore the score would be 2. Arguably, a finer scoring system (i.e. range of [0–100]) is possible and would provide more variation in the scores of alternative DCMs. However, given the reliance on expert opinion for the quality assessment, these finer score ranges might be deceiving. On the other hand, very broad score ranges (i.e. range of [0–3]) may not give enough variation in the quality scores of alternative DCM and so might not be very informative.

Table 4: Health information systems and their characteristics

Information system	Data collection methods for vital statistics used	Collection objectives	Area of coverage	Number of participants
1. National housing and population census (census)	Population-based census Long form survey	Demographic, poverty, housing, labour for participation and health indicators	Nationwide	33 000 000
2. National Household Survey (NHS)	Household survey	Income and poverty focus, with some demographic and health indicators	Nationwide	23 000
3. Demographic and Health Survey (DHS)	Household, community and facility based surveys	Mortality, fertility and use of maternal and child health service	Nationwide	15 000
4. Vital registration (CRVS)	Population-based forms	Compulsory birth and death registration	> 60% districts	3 000 000
5. Health management system (HMS)	Facility-based forms	Continuous collection of morbidity, mortality, and service coverage	Health facility nationwide	17 000 000
6. Integrated disease surveillance (IDS)	Facility-based forms	Continuous data collection on disease and mortality	Health facility nationwide	3 000 000
7. Demographic surveillance system: Region X (X-DSS)	Population-based census Mortality surveillance using verbal autopsy	Regular documentation of births, deaths, migrations and socioeconomic information	Some districts in Region X	66 000
8. Demographic surveillance system: Region Y (Y-DSS)	Population-based census Mortality surveillance using verbal autopsy	Regular documentation of births, deaths, and health service utilisation	Some districts in Region Y	83 000
9. Adult morbidity and mortality project (MM)	Population-based census Mortality surveillance using verbal autopsy	Regular collection of information on burden of disease and mortality	Some districts across the country	500 000
10. Demographic surveillance system for AIDS (A-DSS)	Population-based census Mortality surveillance using verbal autopsy Population-based HIV surveillance	HIV surveillance and some mortality data collection within the surveillance site	Villages within a specific region	23 000

Given the potential subjectivity of assigning quality scores to each DCM, we suggest that in-country evaluations use a team of experts to assign the scores either by consensus (i.e. using Delphi techniques) or individually (aggregating the individual scores from experts into a single score). For Scenario A, we have assumed a scoring panel of six individuals, providing both a consensus and individual scores.[5] Table 5 presents the hypothetical 'consensus' scores assigned to each DCM under each quality criteria.

5 We use a random data generation process to provide the six individual scores. We used a triangular distribution, with predefined mean, minimum and maximum values, within the range of [1–10].

Table 5: Quality performance using consensus scoring matrix

	Accuracy	Relevance	Consistency	Timeliness	Accessibility	Improvement
Census	2.33	8.00	5.50	3.00	5.67	4.00
NHS*	0.83	2.50	4.50	3.00	6.67	2.00
DHS	1.00	4.00	6.00	5.00	7.67	2.00
HMS	4.83	2.50	5.00	7.00	3.67	8.00
IDS	3.17	1.00	5.50	9.00	4.00	6.00
X-DSS	5.00	6.50	4.50	7.50	4.67	7.00
Y-DSS	6.00	6.50	4.50	8.50	5.00	7.00
MM	4.50	5.00	5.50	7.50	4.00	6.00
A-DSS	5.00	4.50	5.50	8.50	5.00	5.00

To facilitate the discussion and our simulations, we have produced only aggregated scores, that is, the combined scores for both vital event and COD statistics

for the six major criteria listed in Table 5. However, to arrive at those aggregated scores, the experts would need to assess each DCM against each criterion listed in Table 2. For example, the scores for Accuracy would be the result of aggregating the scores provided against Coverage, Completeness, Missing data for vital event and COD statistics, as well as, Use of ill-defined categories, Improbable classifications, and Consistency for COD statistics, which have been identified as the criteria relevant to the Accuracy category.

Overall, as shown in Table 5 for our hypothetical country, DCMs did not score highly for Accuracy, partly because we assumed their coverage to be low. Additionally, in trying to reflect the situation in many countries, we assumed that most DCMs, except for data surveillance systems (e.g. X-DSS, Y-DSS and MM), did not collect information about COD, or did not have recorded deaths with appropriately certified COD (incomplete) or most death records had ill-defined COD (use of ill-defined categories). However, the consistency between COD and general mortality of most DCMs was assumed to be fairly high with estimates assumed to be within two standard deviations of general mortality-based predictions.

Likewise, the scores given against Relevance were not very high, as we assumed that only a few DCMs collected and produced statistics by gender and standard age groups. We assumed surveillance-type systems, such as X-DSS, Y-DSS and MM, were representative at the lowest administrative level within the corresponding surveillance sites, thus receiving relatively higher scores.

Similar to the situation in many low-income countries, in Scenario A the definitions and estimations of mortality and COD indicators changed gradually over time and varied across the surveillance areas. Comparability across space was also assumed to be relatively low, as it might be influenced by the capacity of the staff (both medical workers and interviewers) working on reporting and coding death events. Therefore, most DCMs are given a consistency score between 3 and 6. In assessing the timeliness of the various DCMs, the systems that are typically designed to collect data continuously and produce statistics within a year interval were given the high scores of 9 or 10.

Accessibility of a data collection method was measured by three sub-indicators: (i) the format in which the data are released, (ii) the availability and quality of documentation and (ii) the availability and responsiveness of user services. Most DCMs provide datasets in various formats and with average quality documentation of the data collection process, due to incomplete documentation. The user services are not always available and timely. Therefore, the overall score given to Accessibility ranges between 4 and 7.

The last quality criterion relates to the improvement toward a complete CRVS. Facility-based systems, such as health management systems (HMSs) and integrated disease surveillance (IDS), are given relatively high scores. By contrast, the DHS and National Household Surveys (NHSs) were assigned low scores as their data

collection designs and objectives are further removed from the aims of a CRVS collection (e.g. income and poverty, and maternal and child health indicators).

Which costs should be included?

As previously noted, the associated costs of DCMs for vital statistics, and more generally, for health information systems, are not well documented. In Scenario A, we have used some of the estimates from the comparative cost study in Tanzania as a departure point for the cost estimates of various DCMs. We also assumed that coverage (i.e. the sample size of each DCM) is one of the main drivers of costs. Although we have attempted to produce plausible estimates for the scenario, the costs are 'hypothetical' and do not represent the actual or expected costs in any particular country. The annualised total costs for each DCM are presented in Table 6.

Each DCM has different objectives and outputs, so it is important to allocate the share of total costs to the production of vital statistics data. This is a challenging task for several reasons. First, there is a large degree of uncertainty about the cost of vital statistics collection as a proportion of the total cost of each DCM. We decided to approximate this using the share of vital statistics indicators produced. For instance, of the 13 types of indicators that the HMS collected, two of them relate to vital statistics. Therefore the resource share is approximately 15 per cent. However, collecting data on some types of indicators might require more resources than others. For instance, collecting vital statistics via verbal autopsy instruments tends to be more resource intensive than collecting data on use of the health service. In such cases, the collection of data on two vital indicators (out of 13) might take a resource share higher than 15 per cent, and the apportion rule should be adjusted upwards to reflect this.

Second, in reality there might be different rules to apportion different types of costs (i.e. fixed vs variable). For our exercise, we applied some arbitrary apportioning rules, as shown in Table 6, to the total annual costs derived from Tanzania's costing exercise.[6]

Table 6: Estimated costs for vital statistics

	Total cost	Apportion rule (%)	Cost for vital statistics
Census	$8 500 000	25	$2 125 000
NHS*	$840 000	5	$42 000
DHS	$870 000	10	$87 000
HMS	$750 000	30	$225 000
IDS	$2 200 000	20	$440 000
X-DSS	$4 500 000	15	$675 000
Y-DSS	$200 000	40	$80 000
MM	$220 000	40	$88 000
A-DSS	$100 000	60	$60 000

* Note: The acronyms in this table are spelled out in Table 4.

Undertaking the systematic assessment of data collection methods

As discussed in Section 3, we can use the principles of either CEA or EA to compare the outputs and costs for each DCM. As described below, Scenario A uses both approaches to test the robustness of our results.

Cost-effectiveness analysis

An analyst using CEA needs to develop a single measure of output such as the QALYs used in the health economics literature. We therefore need a quality index that combines the scores given under the six criteria in Scenario A into a single quality measure. Using this composite quality index, information on the associated costs is then used to rank each DCM according to its cost per output ratio.

6 Note that some costs in Table 6 were underestimated since fixed cost was not included in the calculations for some DCMs. In these instances, we have apportioned a relatively larger percentage of costs to account for this potential underestimation of fixed cost.

In the previous section, we briefly discussed alternative weighting methods for producing the composite quality index. Each method is subject to different sources of bias, so we chose to test the robustness of our results using a composite index built with DEA versus one built from the unweighted mean scores (see Table 3). Note that in this exercise, we did not have a team of experts to assign weights to each criterion. We strongly recommend using expert opinion to weight indexes when possible, and that the process of assigning weights is kept explicit and transparent.

Under the two main weighting methods used (unweighted and DEA-based) in Table 7, the left-hand column shows the results for the consensus score by expert panel, and the right-hand column shows the results for the aggregated individual scores from the six experts. Using DEA for the consensus score index fails to adequately distinguish between DCMs, that is, DEA has low discriminatory power in this case. This is mostly due to the low number of observations. The individual scores from the six experts help overcome this challenge by multiplying the number of observations by six, which improves the discriminatory power of the quality index and improving its overall reliability.

The scores for the quality categories in Table 5 measure the quality of the data produced, but fail to account for the quantity of data produced (i.e. the number of vital events recorded). Therefore, the quality index score would need to be multiplied by the number of vital events recorded under each DCM. This would allow us to estimate the quality-adjusted number of vital events produced by each DCM. In the case of census, for example, the unweighted score of 0.748 is multiplied by the number of events recorded as proxied by its sample size (i.e. 33 million), which would give us a quality-adjusted number of events recorded equivalent to w25.9 million.

For Scenario A, we would call the product of quality and quantity of data produced the quality-adjusted data index (QADI). When combined with the cost data in Table 6 (column 4), we can produce the 'cost per QADI' for each DCM. This ratio can be used to rank the alternative DCMs under evaluation, as shown in Table 8.

Efficiency analysis

In contrast to CEA, an analyst using efficiency analysis (EA) techniques does not need to separately build a composite quality index. EA can handle multiple outputs by using the in-built assumptions in DEA to combine information on multiple outputs and costs. As described in the Appendix, DEA uses linear programming to estimate a cost-efficiency index that combines outputs (i.e. quality scores against each attribute and quantity of data) and the associated costs of each DCM. This cost-efficiency index is then used to produce the corresponding ranking of DCMs. The interpretation of this index is straightforward; the range of this index is [0–1] and the closer the index to unity, the more cost efficient the DCM. So we rank the data collections by how close they are to unity. The estimated results and

Table 7: Composite quality index

	Unweighted		DEA-based	
	Consensus matrix	Individual matrix	Consensus matrix	Individual matrix
Census	0.7600	0.7786	1.0000	1.0000
NHS*	0.5200	0.5401	0.8860	0.8843
DHS	0.6844	0.7195	1.0000	1.0000
HMS	0.8267	0.8760	1.0000	1.0000
IDS	0.7644	0.7137	1.0000	0.9773
X-DSS	0.9378	0.9122	1.0000	0.9958
Y-DSS	1.0000	1.0000	1.0000	1.0000
MM	0.8667	0.8760	1.0000	0.9932
A-DSS	0.8933	0.9561	1.0000	1.0000

* Note: The acronyms in this table are spelled out in Table 4.

their associated rankings are presented in the last three columns of Table 8.

It answers the question 'Which DCMs are the most cost-effective/cost-efficient for vital statistics data?'. Our evaluation does not answer the question 'Which

Table 8: Cost per QADI, cost-efficiency index and ranking of DCMs

Data collection methods	CEA: cost per QADI unweighted quality index			CEA: cost per QADI DEA-based quality index			Efficiency analysis: cost-efficiency index		
	Consensus matrix	Individual matrix	Rank	Consensus matrix	Individual matrix	Rank	Consensus matrix	Individual matrix	Rank
Census	0.0847	0.0827	2	0.0644	0.0644	2	1.0000	0.9667	2
NHS	3.5764	3.4434	8	2.0990	2.1030	8	0.0280	0.0248	8
DHS	8.4740	8.0615	9	5.8000	5.8000	9	0.0100	0.0102	9
HMS	0.0351	0.0331	1	0.0290	0.0290	1	1.0000	1.0000	1
IDS	0.2990	0.3203	5	0.2286	0.2339	5	0.1630	0.1585	5
X-DSS	1.2925	1.3288	7	1.2121	1.2172	7	0.0490	0.0485	7
Y-DSS	1.0685	1.0685	6	1.0685	1.0685	6	0.0570	0.0563	6
MM	0.1385	0.1370	3	0.1200	0.1208	3	0.4040	0.3817	3
A-DSS	0.2190	0.2046	4	0.1957	0.1957	4	0.2360	0.2410	4

* Note: The acronyms in this table are spelled out in Table 4.

Ranking alternative data collection methods under CEA and EA

In Scenario A, both CEA and EA produce consistent rankings of the DCMs under evaluation (Table 8). HMS is ranked as the most cost-effective/cost-efficient system, followed by census. Surveillance-type systems usually obtain middle rankings, and surveys like DHS and NHS are consistently the least cost-effective methods to gather vital statistics.

Even though the results were based on simulated data, they seem plausible, particularly given that we use the number of vital events recorded as the measure of quantity of data to be produced. As previously discussed, this choice of quantity was justified based on the long-term government policy of establishing CRVS. Therefore in Scenario A, it is not surprising that HMS is relatively more cost effective/cost efficient than individual surveys that rely on samples of the population, with sometimes low coverage and/or limited small-area statistics. The alternative (Scenario B) where the long-term government policy is not to establish CRVS, and hence CRVS is included in the evaluation, is also presented in Section 5.

Several caveats should be noted with this exercise. First, this exercise is a snapshot evaluation of DCMs.

method will be the most cost-effective/cost-efficient to achieve universal coverage of vital statistics data?'. As discussed throughout the report, answering this would require a good understanding of the marginal costs and the cost functions of the various DCMs, which are largely unknown. On the other hand, we should stress that the first step is to understand the current cost structure of alternative DCMs in light of their outputs. This report proposes a systematic and rigorous way to undertake such analysis, which can be tested in-country.

Second, any systematic evaluation of the outputs of DCMs would need to capture the quality of the data produced. However, quality does not have an intrinsic measurement scale and its quantification is highly subjective. It is thus important to be as transparent as possible and acknowledge that the evaluation will always have an important element of subjectivity.

Third, there is no perfect measure of the quantity of output produced by DCMs and the choice of measure can have a substantial influence on the results and the corresponding rankings. In our case, we have proposed two different measures of quantity of data, which are justified on different grounds.

Stylised Scenario B – Assumes no long-term policy of establishing civil registration for vital statistics

As discussed in Section 4, in Scenario B we assume that there is no long-term policy of establishing CRVS. CRVS is thus included as one of the alternative methods to be evaluated. A major implication of this scenario relates to the choice of measure of 'quantity' of vital statistics data. In this scenario, we use target population or number of persons represented, not number of unit records as used in Scenario A. We expect this choice of measure to have substantive impact on the rankings of alternative DCMs and favour those methods with large coverage and relatively smaller samples, such as nationally representative surveys. They would be measured as producing large output (as measured by coverage) with relatively low input (sampled population).

For this analysis, we follow the same methodology discussed in Section 4, with a few exceptions: (a) we use target population instead of unit records as the measure of 'quantity' of vital statistics data, (b) we exclude the last indicator (improvement toward CRVS') from the quality criteria, (c) we include the costs of CRVS. In regard to the apportioning rule, we assume that although CRVS might have other important functions, such as providing legal identification, its incompleteness prevents it from doing so. Therefore, a relatively generous share of the costs (30%) was decided as the apportioning rule. Results for Scenario B are presented in tables 9, 10 and 11.

Table 9: Quality performance matrix by individual indicators for Scenario B

DCMs	Target population	Accuracy	Relevance	Consistency	Timeliness	Accessibility
Census	33 000 000	2.33	8.00	5.50	3.00	5.67
NHS	33 000 000	0.83	2.50	4.50	3.00	6.67
DHS	33 000 000	1.00	4.00	6.00	5.00	7.67
CRVS	20 500 000	0.33	0.50	4.50	5.00	0.67
HMS	33 000 000	4.83	2.50	5.00	7.00	3.67
IDS	6 000 000	3.17	1.00	5.50	9.00	4.00
X-DSS	510 000	5.00	6.50	4.50	7.50	4.67
Y-DSS	180 000	6.00	6.50	4.50	8.50	5.00
MM	2 600 000	4.50	5.00	5.50	7.50	4.00
A-DSS	23 000	5.00	4.50	5.50	8.50	5.00

* Note: The acronyms in this table are spelled out in Table 4.

Table 10: Quality composite index, estimation by two aggregation methods

Data collections	Unweighted		DEA-based	
	Consensus matrix	Individual matrix	Consensus matrix	Individual matrix
Census	0.7600	0.7776	1.0000	1.0000
NHS	0.5200	0.5471	0.8700	0.8682
DHS	0.6844	0.7335	1.0000	1.0000
CRVS	0.5600	0.3407	0.7740	0.8252
HMS	0.8267	0.8677	0.9360	0.9843
IDS	0.7644	0.7114	1.0000	0.9750
X-DSS	0.9378	0.9178	0.9700	0.9888
Y-DSS	1.0000	1.0000	1.0000	1.0000
MM	0.8667	0.8778	0.9980	0.9863
A-DSS	0.8933	0.9639	1.0000	1.0000

* Note: The acronyms in this table are spelled out in Table 4.

Table 11: Cost per QADI and cost efficiency estimation

Data collection methods	CEA: Cost per QADI unweighted quality index			CEA: Cost per QADI DEA-based quality index			Efficiency Analysis Cost efficiency index			
	Consensus matrix	Individual matrix	Rank	Consensus matrix	Individual matrix	Rank	Consensus matrix	Rank	Individual matrix	Rank
Census	0.0847	0.0828	6	0.0644	0.0644	6	0.0630	6	0.0580	5
NHBS	0.0024	0.0023	1	0.0015	0.0015	1	1.0000	1	1.0000	1
DHS	0.0039	0.0036	2	0.0026	0.0026	2	0.8050	2	0.9113	2
CRVS	0.0196	0.0323	4	0.0142	0.0133	3	0.1930	5	0.1738	6
HMIS	0.0161	0.0154	3	0.0142	0.0135	4	0.5540	3	0.4513	3
IDS	0.1465	0.1574	7	0.1120	0.1149	7	0.0430	8	0.0320	7
I-DSS	0.1654	0.1690	8	0.1599	0.1568	8	0.0490	7	0.0353	8
R-DSS	0.4835	0.4835	10	0.4835	0.4835	10	0.0190	10	0.0128	10
AMMP	0.0270	0.0266	5	0.0234	0.0237	5	0.2940	4	0.2213	4
TANESA	0.2190	0.2030	9	0.1957	0.1957	9	0.0390	9	0.0307	9

* Note: AMMP adult morbidity and mortality, HMIS Health Management Information System, NHBS National Household Budget, TANESA Tanzania and Netherlands Support AIDS Research Center Survey.
For other acronyms see Table 4.

The relative rankings of alternative DCMs by CEA and EA are consistent overall Table 11. For example, both methods ranked NHS as the most cost-effective/ best performing system, followed by the nationally representative sample, DHS. On the other hand, DSSs were identified as the least cost-effective method to gather vital statistics, which is not surprising given their low coverage. Interestingly, CRVS, as a method of collecting vital statistics is relatively cost efficient, particularly as it incorporates COD reporting.

The results for Scenario B in Table 11 compared with those of Scenario A in Table 8 show very different rankings of DCMs. This is to be expected due to the different measures of quantity of vital statistics data produced. For example, surveys are found to be the most cost-efficient methods in Scenario B, yet they are the least efficient in Scenario A. This is however in line with our assumptions of country policy focused on the long-term establishment of CRVS (Scenario A) or on collecting vital statistics data representative at national level (Scenario B).

Conclusion

There have been several calls for global improvements in collecting useable data from health information systems to inform public health policies. Chief among these is the need to improve current standards of information on vital statistics and COD. As yet there is no framework to help countries undertake an economic evaluation of DCMS and establish which alternative provides the best value for money. Indeed, this topic seems under-researched given the large investments made by donors and governments to various DCMs and health information systems in general. This might be partly explained by the numerous challenges discussed throughout the report. These range from problems of methodology, such as the definition of quality, to implementation factors such as the lack of basic costing data.

To the best of our knowledge, ours is the first study that provides a systematic framework to compare outputs and costs of alternative DCMs. This systematic assessment of the elements required a rigorous economic evaluation of DCMs and is also the first in the area of health information systems. Our report examined the elements and challenges of the proposed framework, while also providing some feasible approaches to deal with the challenges. In doing so, we laid out the assumptions in a transparent manner and built a stylised scenario of a hypothetical low-income country to illustrate how the proposed framework might operate in country applications. As such, our study provides information on the basic parameters and modelling assumptions required to build simulation models to aid decision-making on future investments in collecting vital statistics data. Similar types of simulation models, such as LiST or the Investment Case Matrix, have been used to guide decision-making in other areas of health (Jimenez et al. 2012; Winfrey, McKinnon & Stover 2011). Our work is still in its infancy and much remains to be done, particularly to pilot test the proposed framework in an individual country using actual, rather than simulated, data. However our exercise demonstrates that a systematic evaluation of outputs and costs of DCMs is not only necessary, but also feasible.

In these times when the global development community is demanding sound evidence to inform its investments, the lack of information on the relative costs and benefits of investing in vital statistics is hindering global efforts to strengthen CRVS. Our study and the proposed framework are a first, vital step to fill this gap and should prompt country applications that inform decision-making on how much and where to invest resources for vital statistics. These country applications of our framework can be used to engage stakeholders in evaluating the alternatives available and developing investment cases for vital statistics. This will provide a focus around which to rally the currently scattered efforts of the global community.

Appendix – Data envelopment analysis

The Appendix provides a brief technical overview of data envelopment analysis (DEA). It focuses on the details of the methodology relevant to an economic evaluation of data collection methods.

Production frontier, productivity and efficiency

As a precursor to the discussion of DEA in the context of our evaluation framework, it is necessary to introduce basic concepts from production economics. They include: (i) the production frontier, (ii) total factor productivity and (iii) efficiency.

The production frontier

The term 'decision-making units' (DMU) refers to production units that convert inputs into outputs. The set of input–output combinations that are technically feasible for the DMUs is known as the production possibilities set. A production possibilities set for a single-input single-output DMU is depicted in Figure A-1.

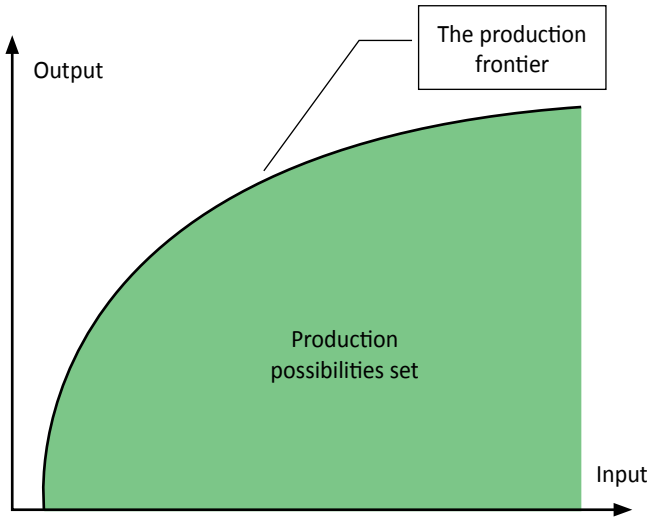


Figure A-1: A single-input single-output production possibilities set and production frontier

The boundary of the production possibilities set is defined as the production frontier (or production function), also depicted in Figure A-2, which summaries the amount of output produced by a given level of input(s). The production function inherits its regularity properties (including monotonicity and concavity) from the regularity properties of the production possibilities set.

When we expand the case to a multiple-input multiple-output technology, a particularly useful representation of the production function is the 'output distance function'. This function gives the 'maximum factor' by which a DMU can radially expand its output vector while using the same input vector. For example, for a particular DMU, if the same amount of input is capable of producing twice the output volume, the output distance is 2. The distance function inherits a number of regularity properties from the regularity properties of the production possibilities set.

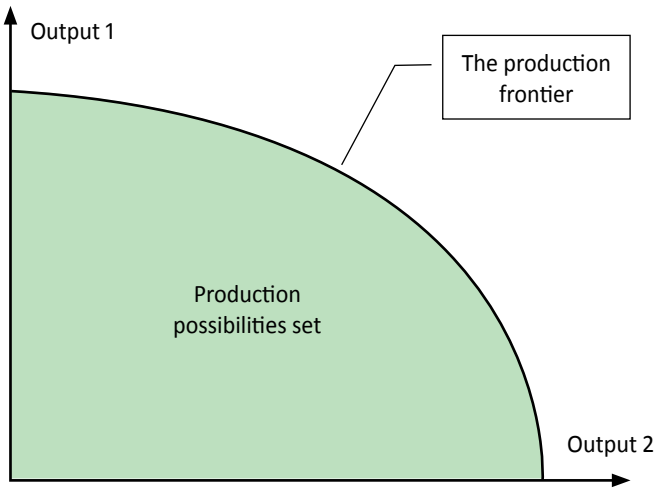


Figure A-2: An output set and the production frontier for a two-output DMU

The production function and output distance functions are defined over input and output quantities, not prices. Other functional representations of the technology, such as cost and profit functions, are defined over prices. For example, the cost function gives the minimum cost of producing a given output vector at a given set of input prices. The cost function inherits a number of regularity properties (including monotonicity and concavity in prices) from the regularity properties of the production possibilities set.

Productivity

Within the DEA approach, productivity of a DMU is commonly defined as the ratio of the output(s) that it produces to the input(s) that it uses. Formally, productivity = output(s)/input(s) or the slope of the line connecting the origin and the DMU, depicted in Figure A-3. In the single-input single-output case, the calculation is straightforward. In the multiple-input multiple-output cases, we need the more comprehensive measure 'total factor productivity' (TFP) (Coelli et al. 2005).

To compute TFP, both outputs and inputs are combined into composite measures that capture all the outputs produced and inputs used. Additionally, one can compute partial measures; for example, labour productivity or machinery productivity. These measures, when considered in isolation, however, can provide a misleading indication of overall productivity.

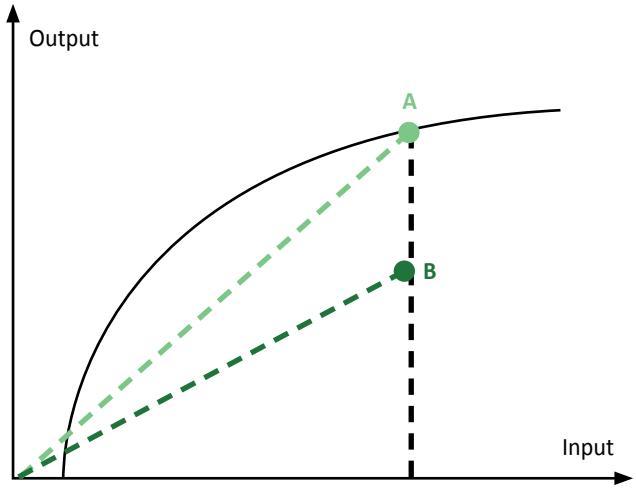


Figure A-3: Productivity measurement for one-output one-input DMUs for observations A and B

Efficiency

Many measures of efficiency are defined in the economics literature but we discuss two that are relevant to our economic evaluation framework. They are technical efficiency and cost efficiency.

Technical efficiency (TE) is the most common efficiency concept used in the literature. It is the conversion of inputs (such as the services of employees and machines) into outputs relative to best practice, that is, the most productive DMUs. In other words, given the current technology, a technically efficient firm will have no wastage of inputs whatsoever in producing the given quantity of output (i.e. maximum possible productivity). A DMU operating at best practice is said to be 100 per cent technically efficient, as represented by observation A in Figure A-3. If a firm is operating below the 'best-practice levels' (e.g. observation B in Figure A-3), then its technical efficiency is expressed as a percentage of best practice.

Cost efficiency (CE) is a measure of the reduction in total cost that is possible while holding outputs fixed. CE can be expressed in terms of the cost function, and formally, is equal to the minimum total cost (i.e. the best practice) divided by the observed total cost.

Other efficiency concepts are often referred to in the efficiency and productivity literature, including 'scale efficiency', 'profit efficiency' and 'allocative efficiency'. We do not discuss these measures in this overview because they are irrelevant to our evaluation framework. Further details can be found in Coelli et al. (2005) and Fried, Lovell & Schmidt (2008).

Data envelopment analysis

Background

The literature describes many different methods for estimating production frontiers, productivity and efficiency. DEA is one popular method, particularly suitable for our purposes because it is a data-oriented approach that produces implicit weights for each DMU (i.e. each data collection method in our framework). Consequently, one does not need to rely on a priori preferences on the relative importance of different quality criteria.

Since DEA was first introduced in 1978, researchers in a number of fields have used this methodology to model operational processes in performance evaluations. Recent years have seen a great variety of applications of DEA in evaluating the performances of many different entities (both public and private) engaged in numerous activities in a wide variety of contexts.

Unlike other methodologies, DEA requires very few assumptions, which facilitates its use in cases involving

complex (and often unknown) relations between multiple inputs and multiple outputs.

DEA model

Let's now formally outline the model for the case of a set of ***N* DMU**, each using one input to produce ***M*** outputs. We can define the output distance function as follows:

$$\max \emptyset$$

$$s.t$$

$$\emptyset q_{i,m} \leq \sum_i \lambda_i q_{i,m} \qquad \forall m$$

$$\sum_i \lambda_i x_i \leq x_i$$

$$\sum_i \lambda_i = 1$$

where $q_{i,m}$ is the amount of output ***m*** produced by the DMU ***i*** ; λ_i is the weighting factor; x_i is the amount of input[9] used; and the restriction of $\sum_i \lambda_i = 1$ allows for variable returns to scale.

The scalar $\emptyset$ represents by how much the production each DMU can increase its output by holding input constant; that is, the proportional increase in output that could be achieved by DMU ***i*** without increasing its input consumption.

The ratio of $1/\emptyset$ defines the technical efficiency score and it takes the value between 0 and 1. If x_i is the total cost, the score is equivalent to a cost-efficiency score.

The efficiency score is illustrated in Figure A-4. All four DMUs produce two outputs, Q1 and Q2, using the same amount of one input (X or total cost). Three DMUs—A, B and C—form the production frontier and D sits below the frontier (i.e. inside the production possibilities set). This implies that A, B and C are fully efficient, and the main difference in their production processes is the mix of outputs that they produce. For example, unit A chooses to produce more Q1 than unit C, while its production of Q2 is around a quarter of that of unit C. However, we can say that 'given their choice of output mix, they are doing the best they can'. On the other hand, D is not fully efficient, that is, it uses the same amount of input as A, B and C but produces substantially less of both Q1 and Q2.

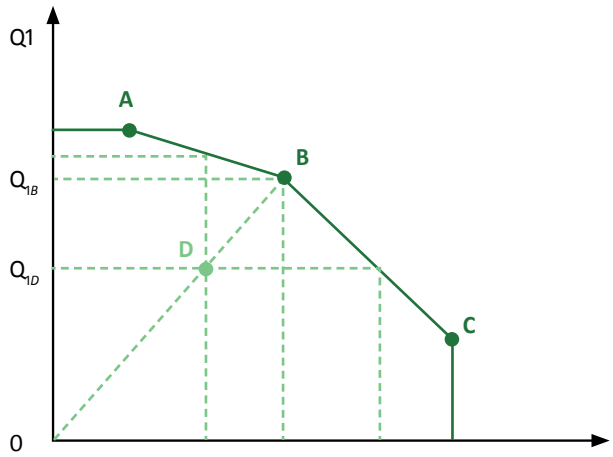


Figure A-4: Frontier constructed using DEA technique: efficient vs inefficient DMUs

Applying DEA to this data, A, B and C will have the efficiency index of 1. Referring to the model, it means $\emptyset$ equals 1; that is, it is not possible for A, B or C to scale up their production without increasing input/cost. On the other hand, the efficiency score of D will be less than 1, and $\emptyset$ takes a value greater than 1 (and consequently, $1/\emptyset$ is less than 1). Indeed, the value of $1/\emptyset$ is measured as the ratio of OD divided by OB (see Figure A-4).

[9] Note that we specify the model with one input only because this applies directly to our evaluation framework In other applications, as outlined above, the model can be expanded to multiple inputs.

[10] Note that we use only two outputs for the graphical illustration. Mathematically, the model can accommodate multiple-output and multiple-input cases just as well as the two-output one-input case.

Note that the efficiency score is the value of the normalised composite index calculated using (i) outputs and inputs (or total cost) of each individual DMU and (ii) the set of weights produced by the linear programming in DEA for each DMU.

Appendix references

Data envelope analysis application in the evaluation framework

In the context of our proposed economic evaluation framework for data collection methods (DCMs), we can apply the DEA methodology by establishing the following.

- Each DCM can be considered as a DMU. They produce the outputs/outcomes of 'good vital events data', which can be measured by the quality attributes (e.g. relevance, accuracy and timeliness). The inputs for that production unit are in amount of inputs required to collect the relevant data by each method, for example, human resources, facilities and equipment. These inputs can then be aggregated into a single item of input, 'total cost'. Thus, we can view each DCM as a multiple-output single-input DMU.
- DMUs can be evaluated and ranked according to their productivity: that is, how much 'good data' a method (i.e. production unit) can produce given the cost. Alternatively, we can think of this question as: 'for a given level of quality data desired, how much money will we need to pay if a particular method is used?'. Different DCMs will have their own maximum productivity, that is, the highest possible quality standard they can achieve. Therefore, they can be ranked from best (most productive) to worst (least productive).
- We are interested in learning which methods of data collection outperform the rest, taking into account all the quality attributes and costs. Those methods that have the highest empirical productivity score are considered more 'efficient' than other DCMs. This implies that we want to measure the productivity and then the efficiency scores of the DCMs to rank them from best (most efficient) to worst (least efficient).

Coelli, TJ, Rao, DSP, O'Donnell, CJ & Battese, GE 2008, *An introduction to efficiency and productivity analysis*, Springer.

Fried, HO, Lovell, CAK & Schmidt, SS 2008, *The measurement of productive efficiency and productivity growth*, Oxford University Press: New York.

References

AbouZahr, C, Cleland, J, Coullare, F, Macfarlane, SB, Notzon, FC Setel, P, Szreter, S, Monitoring of Vital Events (MoVE) writing group, Anderson, RN, Bawah, AA, Betrán, AP, Binka, F, Bundhamcharoen, K, Castro, R, Evans, T, Figueroa, X, George, CK, Gollogly, L, Gonzalez, R, Grzebien, DR, Hill, K, Huang, Z, Hull, TH, Inoue, M, Jakob, R, Jha, P, Jiang, Y, Laurenti, R, Li, X, Lievesley, D, Lopez, AD, Fat, DM, Merialdi, M, Mikkelsen, L, Nien, JK, Rao, C, Rao, K, Sankoh, O, Shibuya, K, Soleman, N, Stout, S, Tangcharoensathien, V, van der Maas, PJ, Wu, F, Yang, G & Zhang S 2007, 'The way forward', *The Lancet, vol. 370, pp. 1791–1799, doi:10.1016/s0140-6736(07)61310-5.*

Bennett, NG & Horiuchi, S 1984, 'Mortality estimation from registered deaths in less developed countries', *Demography, vol. 21, pp. 217–233.*

Boerma, JT & Stansfield, SK 2007, 'Health statistics now: are we making the right investments?', *The Lancet, vol. 369, pp. 779–786.*

Cambridge Economic Policy Associates 2013, Health Metrics Network: Estimating the Cost of the Civil Registration and Vital Statistics System in South Africa.

Campbell, H & Brown, R 2004, *Benefit cost analysis: financial and economic appraisal using spreadsheets, Cambridge University Press: Cambridge.*

Chao, A, Tsay, PK, Lin, SH, Shau, WY & Chao, DY 2001, 'The applications of capture-recapture models to epidemiological data', *Statistics in Medicine, vol. 20, pp. 3123–3157.*

Coelli, TJ, Rao, DSP, O'Donnell, CJ & Battese GE 2005, *An introduction to efficiency and productivity analysis, New York, USA: Springer.*

Drummond, M, Sculpher, M, Torrance, G, O'Brien, B & Stoddart, G 2005, *Methods for the economic evaluation of health care programmes, BMJ Publishing Group Ltd.*

Fottrell, E & Byass, P 2010, 'Verbal autopsy: methods in transition', *Epidemiologic Reviews, vol. 32, pp. 38–55.*

Hill, K 1987, 'Estimating census and death registration completeness', *Asian & Pacific Population Forum, vol. 1, pp. 8–13, 23–24.*

Hill, K, El Arifeen, S, Koenig, M, Al-Sabir, A, Jamil, K, Raggers, H, 2006, 'How should we measure maternal mortality in the developing world? A comparison of household deaths and sibling history approaches', *Bull World Health Organ, vol. 84, no. 3, pp. 173–180.*

Hill, K, Lopez, AD, Shibuya, K & Jha, P 2007, 'Interim measures for meeting needs for health sector data: births, deaths, and causes of death', *The Lancet, vol. 370. pp. 1726–1735, doi:10.1016/s0140-6736(07)61309-9.*

Jimenez, Soto E, La Vincent,e S, Clark, A, Firth, S, Morgan A, Dettrick, Z, Dayal, P, Aldaba, BM, Varghese, B & Trisnantoro, L 2012, 'Developing and costing local strategies to improve maternal and child health: the investment case framework', *PLoS Med, vol. 9: e1001282. doi:10.1371/journal.pmed.1001282.*

Johnson, HL, Liu, L, Fischer-Walker, C & Black, RE 2010, 'Estimating the distribution of causes of death among children age 1–59 months in high-mortality countries with incomplete death certification', *International Journal of Epidemiology, vol. 39, pp. 1103–1114.*

Lopez, AD, AbouZahr, C, Shibuya, K & Gollogly, L 2007, 'Keeping count: births, deaths, and causes of death', *The Lancet, vol. 370, pp. 1744–1746, doi:10.1016/s0140-6736(07)61419-6.*

Lopez, A, Mikkelsen, L, Rampatige, R, Upham, S, AbouZahr, C, Gamage, S, de Savigny, D & Schmider, A 2012, *Strengthening practice and systems in civil registration and vital statistics: A resource kit.* Working Paper Series 19, Health Information Systems Knowledge Hub, School of Population Health, The University of Queensland.

Lopez, AD & Thomason, J 2012, 'If we can't really measure progress towards the MDGs , how do we measure the aid program results', *Development Policy Blog, 9 July, viewed 24 June 2013* <http://devpolicy.org/if-we-cant-really-measure-progress-towards-the-mdgs-how-do-we-measure-the-aid-program-results/>.

Mahapatra, P & Chalapati Rao, PV 2001, 'Cause of death reporting systems in India: a performance analysis', *National Medical Journal of India, vol. 14, pp. 154–162.*

Mahapatra, P, Shibuya, K, Lopez, AD, Coullare, F, Notzon, FC, Rao, C & Szeter, S 2007, 'Civil registration systems and vital statistics: successes and missed opportunities', *The Lancet, vol. 370, pp. 1653–1663, doi:10.1016/s0140-6736(07)61308-7.*

Mari Bhat, PN 2002, 'Completeness of India's sample registration system: an assessment using the general

growth balance method', *Population Studies, vol. 56, pp. 119–134.*

Mathers, CD, Fat, DM, Inoue, M, Rao, C & Lopez, AD 2005, 'Counting the dead and what they died from: an assessment of the global status of cause of death data', *Bull World Health Organ, vol. 83, pp. 171–177.*

Mogyorosy, Z & Smith, P 2005, *The main methodological issues in costing health care services, York, UK.*

Mozumder, KA, Koenig, MA, Phillips, JF & Murad, S 1990, 'The Sample Registration System: an innovative system for monitoring demographic dynamics', *Asia-Pacific Population Journal, vol. 5, pp. 63–72.*

Muennig, P 2008, *Cost effectiveness analysis in health: A practical approach, Josey-Bass,*

Murray, CJL, Rajaratnam, JK, Marcus, J, Laakso, T & Lopez, AD 2010, 'What can we conclude from death registration? Improved methods for evaluating completeness', *PLoS Medicine/Public Library of Science, vol. 7: e1000262.*

Obermeyer, Z, Rajaratnam, JK, Park, CH, Gakidou, E, Hogan, MC, Lopez, AD & Murray, CJL 2010, 'Measuring adult mortality using sibling survival: a new analytical method and new results for 44 countries, 1974-2006', *PLoS Medicine/Public Library of Science, vol. 7: e1000260.*

OECD 2008, *Handbook on constructing composite indicators: methodology and user guide, Paris, France: JRC European Commission, OECD.*

Pena, R, Perez, W, Melendez, M, Kallestal, C &Persson L-A 2008, 'The Nicaraguan Health and Demographic Surveillance Site, HDSS-Leon: a platform for public health research', *Scandinavian Journal of Public Health, vol. 36, pp. 318–325.*

Rajaratnam, JK, Tran, LN, Lopez, AD & Murray, CJL 2010, 'Measuring under-five mortality: validation of new low-cost methods', *PLoS Medicine/Public Library of Science, vol. 7: e1000253.*

Rao, C, Lopez, AD, Yang, G, Begg, S & Ma, J 2005, 'Evaluating national cause-of-death statistics: principles and application to the case of China', *Bull World Health Organ, vol. 83, pp. 618–625.*

Roberts, B, Morgan, OW, Sultani, MG, Nyasulu, P, Rwebangila S, Sondorp, E, Chandramohan, D, Checchi, F & Timmer, A 2011, 'Economic feasibility of a new method to estimate mortality in crisis-affected and resource-poor settings', *PLoS ONE, vol. 6, no. 9, e25175. doi:10.1371/journal.pone.0025175*

Rommelmann, V, Setel P, Hemed, Y, Mponezya, H, Angeles, G 2005, 'Cost and results for information systems for poverty monitoring, health sector reform, and local government reform in Tanzania'.

Rommelmann, V, Setel, PW, Hemed, Y, Angeles, G, Mponezya, H, Whiting, D & Boerma, T 2005, 'Cost and results of information systems for health and poverty indicators in the United Republic of Tanzania'*, Bull World Health Organ, vol. 83, no. 8, pp. 569–577, doi:/S0042-96862005000800009.*

Ruzicka, LT & Lopez, AD 1990, 'The use of cause-of-death statistics for health situation assessment: national and international experiences', *World Health Statistics Quarterly - Rapport Trimestriel de Statistiques Sanitaires Mondiales, vol. 43, pp. 249–258.*

Setel, PW, Macfarlane, SB, Szreter, S, Mikkelsen, L, Jha P, Stout, S & AbouZahr C 2007, 'A scandal of invisibility: making everyone count by counting everyone', *The Lancet, vol. 370, no. 9598, pp.1569–1577, doi:10.1016/s0140-6736(07)61307-5.*

Setel, PW, Sankoh, O, Rao, C, Velkoff, VA, Mathers, C, Goghuan, Y, Hemed, Y, Jha, P & Lopez, AD 2005, 'Sample registration of vital events with verbal autopsy: a renewed commitment to measuring and monitoring vital statistics', *Bull World Health Organ, vol. 83, no. 8, pp. 611–617.*

Soleman, N, Chandramohan, D & Shibuya, K 2006, 'Verbal autopsy: current practices and challenges', *Bull World Health Organ, vol. 84, no. 3, pp. 239–245.*

The Economist 2012, *The magic number, Jan 14 2012, The Economist Newspaper Limited, London.*

Vapattanawong, P & Prasartkul, P 2011, 'Under-registration of deaths in Thailand in 2005-2006: results of cross-matching data from two sources', *Bull World Health Organ, vol. 89, no. 11, pp. 806–812, [Erratum appears in Bull World Health Organ, vol. 89, no. 12, p. 863].*

WHO 2011, Keeping promises, measuring results: Commission on information and accountability for Women's and Children's Health'. <http://www.acdi-cida.gc.ca/INET/IMAGES.NSF/vLUImages/MNCH/$file/Final_Report_Commission_eng.pdf > ISBN 97892156332 8.

Winfrey, W, McKinnon, R & Stover, J 2011, 'Methods used in the Lives Saved Tool (LiST)', *BMC Public Health, vol. 11, Suppl 3: S32, doi:10.1186/1471-2458-11-S3-S32.*

The Knowledge Hubs for Health Initiative

The Health Information Systems Knowledge Hub is one of four hubs established by AusAID in 2008 as part of the Australian Government's commitment to meeting the Millennium Development Goals and improving health in the Asia and Pacific regions. All four hubs share the common goal of expanding the expertise and knowledge base to help inform and guide health policy.

The Knowledge Hubs are funded by AusAID's Strategic Partnership for Health Initiative.

Health Information Systems Knowledge Hub

The University of Queensland

Aims to facilitate the development and integration of health information systems into the broader health system strengthening agenda, and increase local capacity to ensure that cost-effective, timely, reliable and relevant information is available. The Health Information Systems Knowledge Hub also aims to better inform health information systems policies across Asia and the Pacific.
www.uq.edu.au/hishub

Human Resources for Health Knowledge Hub

The University of New South Wales

Aims to contribute to the quality and effectiveness of Australia's engagement in the health sector in the Asia–Pacific region by developing innovative policy options for strengthening human resources for health systems. The hub supports regional, national and international partners to develop effective evidence-informed national policy-making in the field of human resources for health.
www.hrhhub.unsw.edu.au

Health Policy and Health Finance Knowledge Hub

The Nossal Institute for Global Health (University of Melbourne)

Aims to support regional, national and international partners to develop effective evidence-informed national policy-making, particularly in the field of health finance and health systems. Key thematic areas for this hub include comparative analysis of health finance interventions and health system outcomes; the role of non-state providers of health care; and health policy development in the Pacific.
www.ni.unimelb.edu.au

Compass: Women's and Children's Health Knowledge Hub

Compass is a partnership between the Centre for International Child Health, The University of Melbourne, Menzies School of Health Research and Burnet Institute's Centre for International Health.

Aims to enhance the quality and effectiveness of women's and children's health interventions and focuses on supporting the Millennium Development Goals 4 and 5—improved maternal and child health, and universal access to reproductive health. Key thematic areas for this hub include regional strategies for child survival; strengthening health systems for maternal and newborn health; adolescent reproductive health; and nutrition.
www.wchknowledgehub.com.au

The Knowledge Hubs for Health are a strategic partnership initiative funded by the Australian Agency for International Development

Australian AID

Health Information Systems Knowledge Hub
School of Population Health
University of Queensland

COMPASS
Women's and Children's Health Knowledge Hub

UNSW
THE UNIVERSITY OF NEW SOUTH WALES
Human Resources for Health Knowledge Hub

THE UNIVERSITY OF MELBOURNE
THE NOSSAL INSTITUTE FOR GLOBAL HEALTH
HEALTH POLICY & HEALTH FINANCE KNOWLEDGE HUB