United Nations
Department of Economic and Social Affairs

Population Division

Technical Paper
No. 2017/5

Overview of the principles and international experiences in implementing record linkage mechanisms to assess completeness of death registration

This page is intentionally left blank

Population Division

Technical Paper
No. 2017/5

Overview of the principles and international experiences in implementing record linkage mechanisms to assess completeness of death registration

Chalapati Rao and Matthew Kellya

United Nations · New York, 2017

The Department of Economic and Social Affairs of the United Nations Secretariat is a vital interface between global policies in the economic, social and environmental spheres and national action. The Department works in three main interlinked areas: (i) it compiles, generates and analyses a wide range of economic, social and environmental data and information on which States Members of the United Nations draw to review common problems and take stock of policy options; (ii) it facilitates the negotiations of Member States in many intergovernmental bodies on joint courses of action to address ongoing or emerging global challenges; and (iii) it advises interested Governments on the ways and means of translating policy frameworks developed in United Nations conferences and summits into programmes at the country level and, through technical assistance, helps build national capacities.

The Population Division of the Department of Economic and Social Affairs provides the international community with timely and accessible population data and analysis of population trends and development outcomes for all countries and areas of the world. To this end, the Division undertakes regular studies of population size and characteristics and of all three components of population change (fertility, mortality and migration). Founded in 1946, the Population Division provides substantive support on population and development issues to the United Nations General Assembly, the Economic and Social Council and the Commission on Population and Development. It also leads or participates in various interagency coordination mechanisms of the United Nations system. The work of the Division also contributes to strengthening the capacity of Member States to monitor population trends and to address current and emerging population issues.

NOTES

The views expressed in the paper do not imply the expression of any opinion on the part of the United Nations Secretariat.

The designations employed and the presentation of material in this paper do not imply the expression of any opinion whatsoever on the part of the United Nations Secretariat concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.

The term "country" as used in this paper also refers, as appropriate, to territories or areas.

This publication has been issued without formal editing

PREFACE

Methods for assessing the completeness of data from death registration systems have been under constant development since over a century. In the current environment, measures of completeness are required for adjustment of registration data to derive mortality estimates, as well as to monitor and track improvements in registration system performance. Available methods comprise the analytical techniques which apply standard demographic assumptions, and record linkage mechanisms that utilise two or more data sources on deaths for the reference population.

This technical paper summarises the key concepts and principles of record linkage mechanisms, the range of approaches applied to estimate completeness from record linkage and reviews the historical application of these methods according to a typology of study designs to assess completeness of death registration. The paper also discusses various statistical concepts and considerations in analysis of completeness and presents a core set of recommendations for planning record-linkage-based approaches to estimating registration completeness in the new environment of availability of electronic individual death records from multiple parallel sources.

The paper was written by Chalapati Rao (Australian National University) and Matthew Kelly (Australian National University) as background material for the United Nations Expert Group Meeting on the methodology and lessons learned to evaluate the completeness and quality of vital statistics data from civil registration, which was held on 3 and 4 November 2016 in New York. Financial support from Statistics Korea (KOSTAT), on behalf of the Government of the Republic of Korea, to prepare this technical paper is gratefully acknowledged. The paper benefitted from ideas and suggestions provided by Patrick Gerland, Victor Gaigbe-Togbe, Nan Li, Jinho Ho and Romesh Silva and has been revised taking into account helpful comments received from Kenneth Hill, Colin Mathers, Patama Vapattanawong, and Célia Landmann Szwarcwald.

The purpose of the Technical Paper series is to publish substantive and methodological research on population issues carried out by experts within and outside the United Nations system. The series promotes scientific understanding of population issues among Governments, national and international organizations, research institutions and individuals engaged in social and economic planning, research and training.

The Technical Paper series as well as other population information may be accessed on the Population Division's website at www.unpopulation.org. For further information concerning this publication, please contact the office of the Director, Population Division, Department of Economic and Social Affairs, United Nations, New York, 10017, USA, telephone +1-212-963-3209, fax +1-212-963-2147, email: population@un.org

This page is intentionally left blank

CONTENTS

ABBREVIATIONS

BPS	Badan Pusat Statistik
CDC	Centres for Disease Control
CMR	Capture-mark-recapture
CR	Civil registration
CRVS	Civil registration and vital statistics
CS-D	Chandrasekhar and Deming
DDM	Death Distribution Methods
DHS	Demographic and Health Survey
DRS	Dual Registration System
DSP	Disease Surveillance Point
HDSS	Health and Demographic Surveillance Sites
MICS	Multiple Indicator Cluster Survey
MoH	Ministry of Health
MOH ME	Ministry of Health and Medical Education
NOCR	National Organization for Civil Registration
PGE	Population Growth Experiment
POPLAB	Programme of the laboratories for population statistics
SRS	Sample Registration Systems
SVRSS	Sample Vital Registration and Statistics system
VR	Vital registration
WHO	World Health Organization

This page is intentionally left blank

OVERVIEW OF THE PRINCIPLES AND INTERNATIONAL EXPERIENCES IN IMPLEMENTING RECORD LINKAGE MECHANISMS TO ASSESS COMPLETENESS OF DEATH REGISTRATION

Chalapati Rao, Australian National University; Matthew Kelly, Australian National University

1. BACKGROUND AND INTRODUCTION

Information on deaths by age, sex and cause are basic inputs for evaluation of demographic and epidemiological trends in national populations. Primary and secondary analyses of population and mortality data inform various development sectors including labour, housing, health and education, among others. While there are several potential sources for obtaining information on deaths, such as, death registration systems, health surveillance programmes, national censuses and various household surveys, death registration is considered as the optimal source for mortality statistics. This is because in theory, data from death registration systems imply total population coverage[1] as well as potential completeness[2] of death recording (Rao and others, 2005). There are two other parameters of data quality that are implicit in mortality statistics from death registration. Firstly, there is a likelihood for assured validity of recorded variables including age, sex, date of death and cause of death among others, given the legal basis and administrative nature of death registration systems. Secondly, there is a reasonable timeliness in the availability of data, within a timeframe of one-two years from the reference period, given the existence of administrative processes and infrastructure for compilation of vital statistics. In contrast to these theoretical advantages, other mortality data collection platforms suffer from several drawbacks, including recall bias resulting in poor data completeness from censuses and surveys and inadequate population coverage and completeness of health surveillance programmes.

Despite the established theoretical advantages of death registration data, there are widespread deficiencies in the availability and quality of mortality statistics from national death registration systems across the world. Various international assessments of the quality of vital statistics from death registration indicate that about only a third of all countries in the world generate high quality mortality data on a routine annual basis (Mathers and others, 2005; Mahapatra and others, 2007; World Health Organization, 2011) Although most of the remaining countries of the world do have operational death registration systems, the data are deficient due to patchy coverage coupled with incompleteness, thereby of inadequate quality for robust mortality measurement.

In recent times, international development initiatives such as the United Nations Millennium Development Goals and Sustainable Development Goals have created a critical demand for robust empirical mortality data to plan development strategies as well as monitor their impact at national, regional and global levels. In the absence of reliable data, international agencies use a range of demographic models and techniques to estimate completeness of death registration in these countries, and adjust for the same in deriving mortality estimates by age and sex. Such estimation techniques are routinely applied for a large number of developing countries in Asia-Pacific, the Middle East, Africa and the Americas (World Health Organization, 2014). The resultant mortality estimates are generally perceived to be affected by considerable statistical uncertainty, and also lack political relevance in regard to their limited anchorage in local data. Overall, these estimates at best provide a baseline for broad development policy initiatives, but lack the validity and precision required for monitoring their impact over time.

[1] Coverage refers to the population to which the registration practices apply i.e., the population represented in the statistics. For example, registration may be in place only in urban areas/some provinces/sampled areas
[2] Completeness refers to the proportion of estimated deaths in the 'covered' areas that are actually recorded in the registration data. For example, some neonatal/infant/adult/elderly deaths; or from certain socioeconomic sub groups within the covered areas may not get recorded in the registration system

In response to this demand for robust mortality data, there is now an increased focus on strengthening national death registration systems across the developing world, in terms of expanding coverage as well as improving completeness, where applicable. In countries with medium to large populations (>20 million or so); there is an equally critical emphasis to generate reliable subnational data of high quality from death registration systems. Initiatives to strengthen death registration require a primary assessment and understanding of the structure and performance of the existing system. Coverage of death registration can be readily understood from a review of the operational parameters of the death registration and vital statistics system, particularly in terms of the availability (or lack) of vital statistics for various geographical and administrative subunits of the country. A thorough review of the legal framework and the administrative structure and organization of registration services could identify specific impediments to coverage, leading to necessary changes and provision of resources to address the coverage gap.

However, assessment of completeness of death-registration-based mortality data is a more complex matter and has been the subject of demographic research over the past five decades. The principal challenge in measuring completeness lies in establishing an empirical basis for developing an estimate of total number of deaths that would have occurred in the study population during a specified reference period. This estimate of total deaths will serve as the denominator for computing the proportion of deaths actually registered during the reference period; and this proportion is the measure of completeness of registration data. Robust measures of completeness will serve two purposes: first, they will provide factors to adjust observed mortality rates for incompleteness and thereby derive a more plausible estimate of mortality patterns for the study population; and second, reliable completeness measures will enable tracking of improvements in registration system performance and guide interventions to strengthen death reporting and data compilation.

A. Approaches to measuring registration completeness

There are two broad approaches to estimate completeness of death registration, albeit, with their respective strengths and limitations (Preston, 1981). One approach involves the use of indirect demographic techniques applied to data on the distributions of registered deaths by age and sex (also referred to as Death Distribution Methods (DDM)). In this family of methods, statistical models comprising mathematical relationships based on specific demographic assumptions are used to estimate an expected number of deaths by age-sex distribution in the study population. This estimate of expected deaths serves as the denominator to compute the fraction of deaths that were actually recorded in the study population. The other approach involves linkage methods to directly match variables for individual death records from two or more parallel data recording platforms in the study population. The universe of matched and unmatched records serves as a basis for the application of a range of statistical procedures to estimate the total deaths in the study population and subsequently, the completeness of any of the data sources, as required.

The international vital statistics community has been faced with challenges in measuring completeness of death registration for over a century. While record linkage mechanisms were the primary methods employed, there were considerable disadvantages in their implementation. There were logistical considerations in sourcing multiple sets of vital records for larger populations, along with challenges in implementing procedures for linkage/matching of records across the different sources, and finally, limitations in meaningful interpretation of results from such analyses. This led to a reduction in their appeal by the mid-1970s, despite the relatively straightforward computational processes involved. It has been only the Sample (Vital) Registration Systems in India and Bangladesh, and the Disease Surveillance Point System in China, where record linkage mechanisms for evaluating completeness have been sustained on a routine basis.

The limitations in implementing record linkage led to the development of a range of DDMs since the mid-1970s, and these have been tested and widely used since then (Murray and others, 2010). In general, DDMs are far less resource intensive, particularly due to their application on available data from existing systems only, without any need for mounting additional data collection schemes.

However, the limitations in the relevance of some of the demographic assumptions for individual study populations (e.g., constant fertility and mortality for some methods, and absence of migration), as well as uncertainty in accuracy of data from death registration systems (e.g., misreporting of age and reference period for registration data) and in enumerated population from censuses have hampered the overall utility of DDMs. There is also limited institutional capacity for their implementation at the national/subnational level in developing countries. Most importantly, there is inconsistency in outputs of such analysis across populations and over time. Also, there is considerable uncertainty in the resultant measures of completeness, estimated to be in a range of ±25 per cent (Murray and others, 2010). All these reasons have led to the search for more accurate and reliable methods to measure completeness.

Currently, there is renewed interest in the potential for record linkage mechanisms to resolve the challenge of estimating completeness of death registration. Several reasons support such renewal of interest. Firstly, advances in terms of computerised record keeping practices now make available comprehensive electronic datasets of death registration records on a routine basis in some countries, which facilitates processes for record linkage. Such data availability will only increase across the developing world in the near future with the current emphasis on strengthening death registration as well as making use of the advantages of computerisation. Secondly, alternative electronic data sources of death records for record linkage are also becoming available from periodic data collection platforms such as censuses or regular sample survey programmes and can be tailored for specific reference periods for matching and analysis across sources. Finally, improvements in data quality of recorded variables, as well as increasing availability of individual unique identifiers for citizens has enhanced the accuracy of record linkage and matching processes. Hence, the availability of electronic records for linkage and matching vastly reduces the logistical challenges in implementing data collection and compilation for applying record linkage methods, as well as some other constraints specific to matching rules and procedures. Further, the meticulous implementation of conventional record-linkage methods in the routine sample vital statistics programmes of China and India also serve as appropriate examples of the accuracy of such methods and general acceptability of their findings on completeness.

This technical paper summarises the key concepts and principles of record linkage mechanisms, the range of approaches applied to estimate completeness from record linkage and reviews the historical application of these methods according to a typology of study designs. The paper also discusses various statistical concepts and considerations in analysis of completeness and presents a core set of recommendations for planning record-linkage-based approaches to estimating registration completeness in the new environment of availability of electronic individual death records from multiple parallel sources.

B. Conceptual basis of record-linkage methods to evaluate completeness

It should be recognised that the concept of dual-record systems with record linkage to evaluate data completeness has been widely practised in the form of post-census population enumeration for over a century, with derivation of correction factors to adjust census population counts (Whitford and Banda, 2001). However, these experiences do not include analyses specific to evaluation of data on births and deaths. In principle, linking records across data sources offers several avenues for measuring data completeness in any of the sources, as well as for deriving a greater empirical proportionate set of records from the universe of potential events in the study population. Linkage of records identify three subsets of deaths from the two sources: a set of matched records from both sources; and two sets of unique records in either source. Subsequently, a range of methods can be applied to compute the estimated total number of deaths in the study population. Firstly, matched and unique records from all sources could be reconciled to derive the total list and number of observed deaths during the study period. Secondly, based on certain statistical and demographic conditions, the number of events missed by both sources can be estimated, and this number could be added to the other three sets to derive the total estimate of deaths. The total estimate derived by either method could be used as the denominator to compute the fraction of observed deaths in any of the data sources used in the analysis. There are other ‘hybrid’ models that employ a combination of data reconciliation and estimation of missed events.

Table 1 displays the notional model for analysis of completeness of data from a dual systems collection process, as proposed by Chandrasekar and Deming in 1949.

TABLE 1. CONCEPTUAL MODEL FOR DUAL-RECORD SYSTEM EVALUATION OF REGISTRATION COMPLETENESS

Source 2	*Source 1*		
	Reported	*Not reported*	*Total*
Reported	M	U_2	N_2
Not reported	U_1	Z	V_2
Total	N_1	V_1	N

Where:

M = events that are matched across the two sources, i.e., recorded in both sources;
N_1 = total events reported in source 1= M + U_1, where U_1 =records in source 1 not matched in source 2;
N_2 = total events reported in source 2= M + U_2, where U_2 =records in source 2 not matched in source 1;
Z = number of events estimated to be missed by both sources; computed as

$$Z = \frac{U_1 U_2}{M};$$

V_1 = number of events missed by source 1;
V_2 = number of events missed by source 2;
N = estimate of total events = M + U_1 + U_2 + Z.

The computation of the value Z is based on several conditions as follows (Marks and others, 1974c):

a. Independence of the two sources: probability of an event being included in one source is independent of its probability of being included in the second source
b. Homogeneity of reporting in each source: all events in a data source have an equal probability of being recorded in that source
c. Accuracy of reporting in each source: all events recorded in each source are correctly recorded, in regard to being from the same geographic location and time reference to enable compatibility of records for matching
d. Accuracy of matching: the procedures for linking/matching records correctly identify those records that are truly recorded in both sources; as well as correctly identify those that are recorded in only either but not both of the sources. In other words, there are no “erroneous matches” or “erroneous non-matches”.

In actual field operations, it is unlikely that all of the above conditions are fulfilled. Hence, the overall estimate of *N* (the total events in the study population) is likely to be affected by several types of bias, depending on the nature of the events, the processes used in data collection in either or both sources, and data quality in each source. In addition, there is likelihood of statistical error in the estimate of *N* in case there has been any process of sampling (which gets compounded in case of cluster sampling) in either of the data sources. Even if there is no sampling, the reporting of events is subject to stochastic or random variation. Such bias and/or statistical error result in uncertainty in the record linkage based estimate of completeness.

Despite these potential limitations, Marks and others, identified several advantages of the dual system record linkage approach over the indirect demographic techniques for assessing completeness of registration (Marks and others, 1974c). Essentially, the major conditions of the dual-system-record-linkage methods listed above are statistical and relate to the characteristics of the data collection procedures. As a result, a clear understanding of the data collection processes, an assessment of data quality, and clear definition of mechanisms for record linkage provide relevant information to measure bias and error in the completeness estimate derived from such analysis. This is in contrast to the assumptions underlying the indirect techniques, which are demographic in nature and relate to the characteristics of the population under study; and being more generic in nature, do not allow empirical estimation of error in the completeness estimate.

To effectively harness these advantages, Marks and others (1974c), propose that a dual-record-system analysis should be based on

- an efficient study design based on a careful choice of alternatives of data sources;
- a clear understanding of data collection processes, data quality, and matching procedures;
- a thorough analysis and evaluation of study estimates.

Each of these perspectives are influenced by a range of methodological issues, as discussed below.

2. Design of record-linkage studies

Record linkage methods for evaluating completeness are implemented through a cross-sectional retrospective study design. The selection of the two (or more) sources of vital event records that are to be linked are an essential feature of the method. In general, the range of available data sources could be considered to be from one of two broad vital event recording processes; namely, continuous-recording systems that register events as they occur, and periodic data collections, which record details of events that had occurred within a specified reference period. Table 1 presents a broad overview of the typology of mortality data sources according to this categorization, which are used in record linkage studies to assess completeness of registration.

In the current environment, **continuous recording systems** are the preferred data sources for mortality measurement, and therefore serve as the primary source for which a measurement of completeness of death registration is required. The advantages of national level government-operated continuous recording systems are that they operate within a regular legal and/or administrative framework, with assured operational resources from the government, and routine processes for compilation and reporting of vital statistics on an annual basis. Such routine data availability serves as a regular source for deriving mortality estimates, and such estimates could be corrected for bias using measures of completeness of recording. Among continuous-recording systems, national civil registration and vital statistics (CRVS) systems are the optimal source, and hence are the focus of current system-strengthening initiatives in developing countries.

In some countries, national health-information systems have a key role in compilation and analysis of vital records (particularly on deaths and causes of death), and such health based death recording serves is a major alternate primary source to the official civil registration (CR), for mortality measurement. Key examples are in Brazil, Fiji, and Viet Nam, among others (Carter and others, 2011a; Franca and others, 2008; Rao and others, 2010a). However, there is a strong likelihood of inter-dependence in the reporting processes of parallel government systems. Consequently, these multiple sources are potentially more useful for record linkage and data reconciliation, and can be used for simple evaluation of completeness of either source as a proportion of the reconciled dataset. However, a careful understanding of system characteristics of each source would be necessary to evaluate whether the various conditions for applying capture-mark-recapture (CMR) methods are met, before applying the same to estimate completeness.

In countries with large national populations but without efficient civil registration systems as in China, India and Bangladesh, Sample Registration Systems (SRS) have been established since the past four decades, as an alternative interim data sources for estimating vital rates. Sample Registration Systems involve continuous recording of vital events in a nationally representative sample of population clusters. Moreover, as will be discussed later, these systems also implement periodic data collections specifically designed to serve as a secondary source for evaluating completeness of the SRS. More recently, Indonesia has also designed and implemented its national sample registration system, but has yet to establish a secondary source for evaluating completeness (Pratiwi and Kosen, 2013).

Other types of continuous recording systems have only partial coverage, in regard to representation by geography; or by restriction to specific population subgroups. Of these, the "special registration" projects deserve specific description, in that they are established in selected geographic area specifically to record vital events for demographic and epidemiological research. These special registration projects are often supported by a mix of government and international funding sources, and a major international network (INDEPTH) coordinates such special registration projects across several countries in Africa and Asia (Sankoh and Byass, 2012). In practice, data from special registration sites are often used to measure vital rates, but their utility is limited for national mortality estimation, due to their narrow geographic coverage as well as inadequate socioeconomic representation of the national population. There is now a recognition of potential of death records from special registration projects to be used in record linkage mechanisms to evaluate completeness of data from the CRVS for a defined geographic area.

The remaining types of continuous recording systems are in the form of death registers maintained by government programmes for specific age groups (e.g., maternal and child deaths) or by specific disease surveillance programmes such as those for tuberculosis, HIV/AIDS, or cancers, among others. In principle, owing to their very limited scope for recording deaths, records from such registers are largely useful only for improving CRVS completeness through data reconciliation, and can possibly support data quality in regard to improving accuracy of cause of death attribution. Hence, data from continuous recording systems with partial coverage are at best useful as a secondary source for evaluating and/or improving completeness of death registration. In summary, the CRVS and SRS are major primary mortality data sources for which measures of completeness are required.

The second group of sources of death records comprise **periodic data collections** implemented through sample household surveys that record individual or cumulated events across a defined recall period. Such processes are often used to measure vital rates (with considerable uncertainty) in populations that lack efficient continuous recording systems. Most routine surveys listed in table 1 collect important details of individual events such as name, age, sex, address and date of death; all of which could be used for record linkage/matching.[3]

The decennial census is another source collecting eventually details about deaths in households during the 12-36 month preceding period (United Nations, 2014). In general, mortality data collection in censuses is under-reported, primarily due to recall bias, hence census mortality data does not yield reliable vital rates. However, there is very good potential, as demonstrated in a recent research study in Oman, for census death records to be used as a secondary source to provide robust national and subnational measures of completeness of the CRVS. Another important periodic data source is the intercensal sample surveys, primarily designed to assess

[3] The census, as well as many general-purpose household surveys (including earlier rounds of the UNICEF MICS surveys) often collect only cumulated events of 'children ever born and surviving' from women survey respondents of child bearing age. Lack of individual characteristics of births/deaths preclude the use of these data in record linkage mechanisms.

TABLE 2. TYPOLOGY OF DATA SOURCES USED IN RECORD LINKAGE STUDIES TO EVALUATE COMPLETENESS OF BIRTH AND DEATH REGISTRATION

Type of data collection	*Primary source*[1]	*Secondary source*[2]	*Remarks*
Continuous recording systems			
Civil registration	Yes		• Optimal source • annual data on routine basis
Complementary/alternate registration	Yes	Yes	• Health system vital records • Parish registers in Christian societies
Sample registration	Yes	Can serve as a secondary source for evaluating CRVS	• Best alternative to civil registration for statistical purpose • Indian Sample Registration System (Vital Statistics (Sample Registration System) Division, 2007) • Chinese Disease Surveillance Point System (Yang and others, 2005) • Bangladesh Sample Vital Registration System (Bangladesh Bureau of Statistics, 2011)
Special registration	Yes	Can serve as a secondary source for evaluating CRVS or SRS	• e.g., Health and Demographic Surveillance Sites (HDSS) in several countries (INDEPTH Network) (Sankoh and Byass, 2012)
Age based registers		Yes	• Maternal/child health • senior citizens /pensioners' databases
Disease surveillance systems		Yes	• tuberculosis • cancers • injuries • stroke
Periodic data collections			
Census (total population)	Yes	Yes	• Optimal 2nd data source (national coverage)
National sample surveys		Yes	• Census post-enumeration surveys • Intercensal surveys • Demographic and Health Surveys (The DHS Program, 2017) • Routine national health/socio-economic surveys
Special surveys designed to assess completeness		Yes	• Evaluation surveys for sample/special registration • sporadic research based examples

Note:
1 = data source for which completeness needs to be evaluated
2 = data source which will be used to evaluate completeness of the primary source

population dynamics and change. These surveys can also provide sufficient data to estimate completeness using record linkage, as demonstrated by a recent analysis in Thailand (Vapattanawong and Prasartkul, 2011). However, there is a limitation in sample size, which precludes sub-national completeness analyses.

Countries which operate sample registration routinely conduct 'completeness assessment' surveys specifically designed to independently collect vital event data for a defined preceding period within the sample population in India (Vital Statistics (Sample Registration System) Division, 2007), or a subsample thereof in China (Yang and others, 2005; Guo and others, 2015). These surveys are conducted at considerable additional cost, with the sole purpose of providing the second data source for data reconciliation and/or record linkage based assessments of completeness. Such special surveys have been practiced for several decades in India and China, but the design and periodicity in implementation of these surveys vary across these countries. There is a need to review and revise the survey design features, to potentially improve efficiency and possibly reduce costs. Nevertheless, these periodic surveys are a valuable second source of records to evaluate completeness of data from the Sample Registration system.

Other household surveys conducted on a periodic basis are also secondary sources for record linkage studies, especially when they collect individual level data through full birth histories for children, or other details about individual deaths in the household. These include surveys conducted under the Demographic and Health Survey (DHS) programme, United Nations Children Fund (UNICEF) Multiple Indicator Cluster Survey (MICS) programme, the World Health Organization (WHO) STEPS surveys for measuring non-communicable disease risk factors in adults, and other regular health and socioeconomic survey programmes in developing countries. These surveys employ relatively small clusters for data collection, which create major administrative challenges in matching with records from continuous recording systems, particularly if there is poor data quality in either or both sources. Nevertheless, recent expansion in computerisation of survey data, as well as greater attention to data quality have significantly improved the potential for such periodic data collections to serve as secondary sources for evaluating completeness of the CRVS or SRS. Another advantage of other nationally representative surveys listed in table 1 is that they have shorter intervals between rounds and can hence serve as a secondary data source of vital records on a more regular basis, for evaluating completeness, as compared to the censuses or intercensal surveys. However, the relatively small sample sizes of these surveys result in completeness measures that are limited in scope and precision. Despite this limitation, as mentioned earlier, the availability of computerised survey data could enable electronic matching of records with minimal additional costs, and every opportunity for such evaluation should be exploited, given the frequency and regularity of these survey programmes on a routine basis in many developing countries.

In summary, the choice of study design in regard to data sources would usually involve one of the continuous recording systems as the primary source for which completeness is to be evaluated; and one periodic data collection process, which will serve as the second source for record linkage and evaluation of completeness.

An alternate study design utilises multiple parallel continuous recording systems operated by the health sector or other agencies, to apply record linkage mechanisms for measuring completeness. An example of such a study is available from the Philippines, which covered a selection of municipalities on the island of Bohol during 2002-2007 (Carter and others, 2011b). However, such a study design would need careful assessment in regard to the characteristics of selected data sources in meeting pre-conditions for applying statistical methods to estimate completeness. Also, in some countries, there is potential for a range of such combinations of primary and secondary sources, which could be applied in separate studies. A potential hierarchy of combinations of data sources for record linking taking into account the feasibility, logistics and quality of completeness measures is presented later in this paper, to guide any decisions regarding the choice of study design.

3. CONCEPTS OF RECORD LINKAGE

The mechanisms for record linkage as well as subsequent evaluation of completeness are influenced by specific characteristics of each of the data sources. These could include the characteristics of the population represented in each data source (in particular, the sampling design where applicable), and the methods used for data collection and quality control in each source. Also, the specific processes for linking and matching of records across the two sources are an important aspect of the analysis. There are several underlying conceptual issues that need to be considered when undertaking a record linkage analysis of death recording completeness, which are discussed below.

A. Compatibility of data sources

Data sources, across which records are to be linked, should be compatible in regard to the following:

a) Delineation of reference population—there should be clear geographical boundaries or administrative coverage and time reference points, which are common to the data sources to be linked. These common characteristics would reduce the potential for "out-of-scope" events (also known as *coverage error*), which have important implications for the matching process and resultant measures of completeness. Clear documentation of these characteristics is essential to assure compatibility across data sources.

b) Presence of multiple common variables for record linkage. In most instances, even in the presence of common standard unique identifiers (e.g., citizen ID) for individual records in both sources, it is required that additional variables such as name, age, sex, address, mother's name and date of death, are also available in both sources to be utilised to assess or ensure validity of the record linkage.

B. Assurance of data quality of variables

The completeness and accuracy of recorded data variables in all sources is essential to maximise the potential for reliable matching of records. Common problems with data from CRVS include incomplete names and missing or wrong information on age, address, or date of death. Accuracy of variables will reduce the potential for erroneous matches and non-matches, both of which have major implications for the accuracy of the estimated measure of completeness. Such accuracy can be achieved through rigorous training and data quality assurance mechanisms, including standardization of name spellings and address variables, and careful enquiry and confirmation of key dates of events, among others.

C. Issues in matching

c) Mechanism for matching. Manual matching of records has been the common approach for record linkage to assess completeness of death registration up to now However, the availability of computerised datasets from both registration as well as surveys has enhanced the potential for electronic linkage, which is speedy, systematic and cost effective. However, it is likely that electronic matching would need to be supplemented by a certain degree of manual processing in an iterative process to improve electronic record linkage, through manual resolution of minor discrepancies in spelling or other typographical errors and adjudication of special instances.

d) Matching rules or criteria. Matching can be conducted according to explicit or implicit rules. Implicit matching relies on local judgement of record matches by field experts, using a broad set of characteristics on individual events. Such implicit rules are usually difficult to control or replicate, and hence, have low reliability. Nevertheless, they are important and necessary in situations where there are missing or wrong data, which can only be verified or judged based on knowledge of local culture and practices. In the current environment, there is increasing emphasis on explicit rules or criteria and these are essential for electronic record linkage.

Explicit rules can be operationalised in either a **deterministic** approach (requirement for exact adherence to rules); or a **probabilistic** approach, which assigns weights to specific variables with a margin of error or uncertainty, which are summed to an overall weighted score to define probabilistically matched pairs. The probabilistic approach helps account for perceived variations in quality of data in either or both sources.

The choice between using deterministic or probabilistic approaches to data matching should depend on the quality of the data in the sources being matched, and in particular, the presence or absence of key reliable, direct identifiers. In settings where there are a variety of high quality variables upon which matching can be based, deterministic approaches can be most appropriate. Where data are of sufficient quality, the time and resource savings made by using deterministic methods are an advantage. Where information is of lower quality, probabilistic approaches will allow for the identification of a larger number of 'probable' matches improving the potential for data linkage. This is because probabilistic approaches allow for the calculation of partial agreement weights, which can then be assessed as to whether the probability is sufficient that a match can be assigned (Dusetzina and others, 2014). The other advantage of probabilistic approaches in poor data quality settings is that they allow for the possibility of determining likely matches based on relationships observed in the data whereas deterministic approaches require *a priori* pre-determination of a list of variables to be matched before records are linked (Tromp and others, 2011).

e) Measurement of matching error. In setting up the rules/criteria for matching, there would be a trade-off in regard to the number of variables used in matching and the tolerable limits of variation for specific variables; such that the overall process would result in the potentially *minimum net matching error,* i.e., the minimum difference between erroneous matches and erroneous non-matches.

f) Field verification. Finally, after completion of linkage and matching, there should be a mechanism for field verification of matched and mismatched records in at least a sample. Verification could be done through either comparison with an additional independent source or in the case of a birth or death, through actual contact with the household. Results from such verification or validation could be utilised in refining the probabilistic approach towards matching, through identifying the optimal tolerance limits for variables used in matching.

D. Purposes of record linkage

In principle, record linkage can potentially serve several broad purposes. These include linkage to enable data reconciliation to augment the overall set of records for estimating vital rates, linkage to estimate completeness of either data source, or linkage simply to enhance the set of variables for each record, which will facilitate more detailed data analysis. In addition, linkage can identify specific factors that are associated with event recording or non-recording by the civil registration system, thereby identifying the reasons for incompleteness. Such analyses including disaggregated analyses of completeness by various population characteristics (e.g., urban/rural/age/gender), especially at subnational level, can help design of local interventions for system strengthening. Herein what follows, concepts underlying only the first two purposes are discussed, that is, the use of record linkage in relation to estimating vital rates and data completeness.

a. Record linkage to improve data completeness

Record linkage can be used to enable identification of events common to both sources as well as those unique to either source. After verification and/or adjudication of all potentially true events within the study population, these events could be all reconciled into a final list of events for the study population within the defined reference period. The resultant total could be used as a numerator in computing vital rates. Such linkage and data reconciliation could also be attempted with partial data sources such as specific disease registers (e.g., tuberculosis, cancer), or records for specific population subgroups (pensioners, indigenous groups) as described in table 1.

b. Record linkage studies to estimate completeness of a data source

In regard to evaluating completeness, record linkage can be applied in several ways. The simplest method is following up on the reconciliation process described above, to compute the completeness of either source as the proportion of events recorded in it out of the total events from the reconciled list. Record linkage studies could also be used to estimate the completeness of a data source based on assumptions underpinning the "capture-mark-recapture" (CMR) method. Both data reconciliation and CMR methods could be applied independently or in combination to evaluate completeness for subsequent application in population level vital rate estimation, as illustrated in various examples discussed later. Also, such completeness estimates are affected by various sources of bias and/or error, and as far as possible, completeness estimates should be reported with specific margins of statistical uncertainty to enable sensitivity analysis of estimated vital rates.

E. Statistical conditions for applying the CMR method

As mentioned previously, coverage error and problems with data accuracy, which result in matching error can affect the completeness estimate derived from record linkage. In addition, there are two key statistical conditions underlying the CMR method to estimate events missed by both or multiple sources. Firstly, there is a need for homogeneity of capture probability within each data source. This implies that there is no selective exclusion of certain events in the study population due to a specific characteristic such as age group, gender, socioeconomic status, ethnicity or geography. A basic description of the design and data collection process for each data source, particularly in terms of legal and administrative procedures for coverage of all population sub groups, will enable an assessment of the potential for any bias in this regard. If necessary, a stratified analysis of event capture rates in each of the data sources using the related denominator populations for each stratum could also be used to evaluate if there is any gross exclusion of events for any subgroup.

The second condition requires independent probabilities for event capture in either data source. In other words, the condition of statistical independence between the two data sources is fulfilled if the chance of event capture in one source is not related to whether it was recorded or missed by the other source. This essentially implies that the two recording processes are completely separate, with no common involvement of personnel or sharing or information between the two recording systems at any level. For example, the personnel involved in civil registration at the local level have no involvement in the local household survey to record events. As a result, the probability of capture in either source rests entirely on the reporting practices in either data recording process. If there is communication between the two recording systems, it will result in a larger proportion of matched events than the reality and result in an overestimate of the completeness rate, and, hence, eventually an underestimate of the vital rate. Communication between two sources is sometimes intentional and should be factored into any CMR analysis and the resultant "*response correlation bias*" should be estimated, as described later. At a minimum, a detailed description of the process of data capture in each source, specifying organizations and individual involved, processes for data recording, transmission and storage and any planned communication or likelihood of unplanned communication between the two systems should be documented. This will enable a qualitative assessment of the potential independence/dependence between the two sources.

F. Ethics and data confidentiality

Record linkage studies by their very nature involve the use of individual-records and invariably require the use of some identity variables in the linkage process. Hence, the principles of privacy and maintenance of data confidentiality need to be an integral part of such studies. In the case of use of administrative records, suitable gatekeeper approval is required from the data custodian office. In addition, where records from previous primary research studies are to be used, suitable additional ethical clearance may be required from study participants or their representatives, where the original clearance did not envisage such secondary use of research data for record linkage purposes. In most instances, however, a supplementary ethical clearance stating the purpose of the record linkage along with suitable assurances regarding maintenance of data confidentiality would be sufficient. Nevertheless, these issues require careful attention in record-linkage studies.

4. ANALYSIS OF ESTIMATES

A third set of considerations in record linkage studies lie in the evaluation and analysis of estimates of measures of completeness and resultant vital rates. In particular, attention is required in regard to the estimation of precision of the completeness estimate and the potential for bias, variance, and overall error in the estimates. As discussed earlier, a thorough qualitative assessment of the design and data collection procedures of each source will help evaluate the veracity of record linkage and potential for meeting the conditions for estimating completeness by the CMR methods. Further, conduct of field verification of unmatched events as well as a sample of matched events also helps evaluate data quality as well as validity of the matching procedures employed in the analysis. These additional steps can improve the reliability of the information collected on observed events in the two data sources and therefore also confidence in estimating the number of events not recorded in either data source.

In general, Marks and others (1974c), propose that the overall error in the estimate of completeness can be expressed in the form of a “root mean square error” as follows:

$$\text{RMSE}(x) = \sqrt{variance + bias^2}$$

The main source of variance in the estimate would be on account of sampling in one or both of the data sources and the nature of sampling in terms of the sample size and its distribution across strata and/or the characteristics of the cluster design. In general, sampling variance can be estimated from details of the study design, sampling procedures and the study results. However, if both data sources cover the entire population (e.g., civil registration as one source and the population census as the second source) then there will be no sampling variance.

There are multiple sources of bias in the completeness measure mainly arising from: the absence of statistical independence between the two data systems; problems with matching; and the presence of “out-of-scope” events (in regard to time and space) in either or both data systems. However, not all these biases operate in the same direction in regard to the overall measure of completeness, hence the three sources of bias would tend to cancel each other, thereby minimising the overall error. Nevertheless, the study design, organization of data collection, and data evaluation/analysis should be oriented towards minimising the potential for each of these sources of bias.

In regard to error resulting from bias, Chandrasekhar and Deming (1949), proposed the following formula for the standard error of the completeness estimate:

$$SE = \sqrt{Nq1q2/p1p2}$$

Where:

N = total number of events estimated by the method (see table 1)
$p1$ = the probability that an event is recorded in data source 1
$p2$ = the probability that an event is recorded in data source 2
$q1$ = the probability that an event is missed in data source 1
$q1$ = the probability that an event is missed in data source 2

They proposed that this formula is applicable under the assumption of complete statistical independence between the two data sources and that there is no matching bias and there are no out of scope events. Each of these three sources of bias require careful evaluation in regard to the circumstances under which they arise, how they should be assessed and their specific implications for the overall error in the completeness estimate.

A. Bias from lack of statistical independence

Lack of independence occurs as a result of correlation or "sharing" of information between the two sources, which will result in a higher than actual number of matched events. As mentioned previously, such sharing of information could be unplanned, or it may result from a deliberate process in an attempt to conserve resources or improve overall data quality. However, from a statistical perspective, it results in an overestimation of the matched records, and hence, an overestimation of the measure of completeness, and, finally, an underestimation of the vital rate in question.

While statistical independence is often difficult to ensure, there are several ways to limit violations of the independence assumption, or to minimise their impact (Krotki, 1978). Firstly, in regard to data collection, the design should ensure that the personnel involved in the two systems are completely different and that the data collection operations should be through altogether different mechanisms (e.g., civil registration and census, rather than health system records and health surveys), which will promote the potential for statistical independence. Secondly, adequate data quality assurance and control measures should be in place to eliminate/minimise the potential for out of scope events or matching errors. Also, the entire study datasets could be stratified by administrative subunits; or sex and age; and the estimation could be conducted for each stratum separately, followed by a summation of the total number of events estimated for each strata. In summary, if there are no spurious reports or matching errors, then the completeness estimate from the original Chandrasekar-Deming method could be considered a theoretical maximum estimate of completeness. Also, this estimate could be accurate, if there was true statistical independence between the two sources.

More recent studies have shown that the concept of statistical independence may not be directly achievable in the context of studies of human populations, owing to behaviour induced reporting error (on the part of response bias from the household or recording fault in data collection process), rather than any chance-related randomness in reporting error (Blacker, 1977; Chatterjee and Mukherjee, 2013). There have been several attempts over the past four decades to develop methods to measure the effect of statistical dependence between the two data sources and other components of error (bias and variance) in the completeness rate. Table 3 summarises key research publications on this subject.

There are two key observations from this comparison of methods. Firstly, there are very few practical examples using real data from multiple different Dual Registration System (DRS) projects. Secondly, there is no consistency in the approach to deriving the completeness estimate error, arising from lack of statistical independence. Chandrasekaran (1983), and Hook and Regal (Hook and Regal, 1995) propose that a range of completeness measures from different estimation methods is a potential solution to evaluate the likely error in the completeness estimate resulting from lack of statistical independence, rather than computing a variance-based standard error and 95 per cent confidence interval. Hook and Regal suggested that a variance based standard error would result in a range of potential values, which would include a total number of events lower than those that were actually observed, which may not be valid; hence, they were not in favour of this approach.

B. Bias from "out-of-scope" events

The presence of "out-of-scope" events in either or both sources will result in a lower than true set of matched records and, hence, in an underestimation of the measure of completeness, and a resultant overestimation of the vital rate in question. Hence, bias from such problems with the data acts in an opposite direction to that from lack of statistical independence. Where the two data sources apply to the entire country, geographical out-of-scope errors would only arise in case of vital events that occurred abroad. However, such errors could influence the matching process for subnational analysis. Also, in dual-record linkage mechanisms involving sampling, geographical errors are of importance, since the smaller the size of clusters in the sample, the greater the chance for such errors. Commonly, incorrect application of *de facto* or *de jure* rules for recording vital events results in geographical out of scope events.

In general, continuous recording systems have less scope for errors in recording of date, given the usually short interval between date of event occurrence and date of recording, except in the case of delayed registration. However, out-of-scope errors due to date differences are more an issue with periodic data collections, in which there is potential for "heaping" of events around the margins of the reference period e.g., in January and December; which could include some events from outside the time period.

Some data collections have specific challenges in regard to time reference. For example, Viet Nam's annual survey of population change is conducted in the first week of April every year, but asks for information on vital events dating from the annual 'Tet' festival of the previous year, which falls in February, creating confusion across the two annual festivals within the reference period (Ngo and others, 2010). Similarly, Oman's census on 12 December 2010 asked for deaths in the previous calendar year, from 13 December 2009 to the date of the census and this created confusion where the exact date of death of deceased relatives was not known to respondents (Al Muzahmi, 2015). Then again, Pakistan's demographic survey experience has shown that more regular surveying of respondents is not necessarily an advantage. In this setting regular visits led to much time and cost being spent on matching birth and death events due to overlapping time periods. From 1999 onwards, this system moved to one time visits covering a time period of 12 months.

Care needs to be taken during survey design and interviewer training to clearly define the geography, definitions of residence status of deceased and, where relevant, identify specific and unambiguous local calendar time points. Field verification of potentially out of scope events can also help ascertain the nature and degree of such events in the overall dataset.

C. Bias from matching error

Finally, there is a need to evaluate the record linkage procedure for matching error. As mentioned previously, there could be erroneous matches or erroneous non-matches. The difference between the two would provide a measure of the net matching error. If the number of erroneous matches is larger, it would result in an overestimation of completeness and therefore underestimation of the vital rate; with the converse applying in situations where false non-matches are higher.

The reasons for matching error commonly include similarities in names and incomplete or inaccuracies of address or date variables. These phenomena are particularly challenging when conducting electronic matching at national or regional level with explicit rules. In contrast, matching errors can be minimised by the use of implicit rules by local personnel to conduct matching at primary or peripheral sampling units. These implicit rules are based on subjective assessment using additional local information or knowledge that is not documented in the records being matched. While implicit rules have advantages of greater flexibility and use of additional information, these are not clearly defined and not reproducible in all situations. As a principle, matching should be based on explicit criteria and apply implicit rules by well qualified professionals in doubtful cases, with proper documentation of the implicit reasoning for adjudication for either matched or mismatched cases from explicit rules (Marks and others, 1974a).

In general, manual matching is a tedious and potentially costly process in terms of resources as well as time. Recent advances in the development of probabilistic approaches for electronic record linkage and matching have improved the potential for efficient and cost effective processes of this nature. It also facilitates the testing of effect of variations in the 'tolerance limits' for different variables on the overall matching rates and setting the appropriate margins or bounds for the specific variables. Several examples of this approach are available (Joubert and others, 2014; Machado and Hill, 2004; Marks and others, 1974a). These methods are becoming increasingly relevant in the current environment of using large electronic datasets for record linkage, with minimal potential to obtain additional information to verify matching processes. Further research is needed to document standard guidelines and a scientific approach and methodology to adopt probabilistic methods for matching.

D. Sampling error

In addition to bias, as mentioned earlier, sampling error must be also taken into account in computing the precision of the estimate of completeness. Invariably, sampling for continuous death registration or periodic surveys is based on a cluster design, for administrative and logistical reasons. The overall sample size and the size, number and distribution of clusters all influence the overall sampling error. In addition, the heterogeneity across clusters is also an important factor. Overall, larger clusters in the sample are generally more heterogeneous for most characteristics (e.g., age composition of households, prevailing fertility and mortality patterns), and this minimises the effect of the cluster design and, hence, a reduction in the error of the completeness estimate (Marks and others, 1974b). These effects of sampling error, where applicable, should be analysed and factored into estimation of completeness and further implemented in measuring the uncertainty intervals of vital rates.

E. Net error in completeness estimate

In summary, lack of statistical independence tends to overestimate completeness; occurrence of out-of-scope events tends to underestimate completeness and problems with matching can bias the estimate either way, depending on the direction of the net matching error. In effect, the net bias may be compensated by the differential effects of each source of bias. Overall error should always include any variance arising from sampling.

Despite this potential for bias and error, statistical methods for measuring completeness using record linkage, offer a distinct advantage in enabling estimation for different age groups in contrast to indirect demographic techniques, which produce a completeness estimate invariant by age. Such subgroup analyses from record linkage mechanisms allow finer adjustment of mortality rates for estimating summary mortality measures such as life expectancy at birth and age-sex specific risks of mortality, as observed from studies in Oman and Viet Nam (Al Muzahmi, 2015; Hoa and others, 2012). Further, such disaggregated analysis can also be extended to specific population subgroups for which there are greater problems with registration completeness and could therefore be targeted through interventions to strengthen completeness.

TABLE 3. SELECTED ARTICLES DESCRIBING METHODS TO MEASURE BIAS AND ERROR IN COMPLETENESS ESTIMATES FROM DUAL-RECORD SYSTEMS ANALYSES

Author/Year	*Source*	*Title*	*Example*	*Methods/Results*
Seltzer and Adlakha (1974)	International Program for Population Statistics, UNC, Chapel Hill, United States of America. Reprint Series 14, 1974	On the effect of errors in the application of the Chandrasekar-Deming technique	Theoretical example	Proposes methods for estimating net relative bias from out of scope events, lack of independence, matching errors, and interactions between these three sources. No details on measurement of standard error of completeness estimate.
Greenfield (1976)	Journal of the Royal Statistical Society. 139 (3) 389-401	A revised procedure for Dual-Record Systems in Estimating Vital Events	Malawi Population Change survey, 1972	Greenfield estimate accounting for lack of independence; 10 per cent higher than standard DRS estimate; no standard error measurement.
Raj (1977)	Journal of the American Stats, Assoc. 72 (358) 377-81	On Estimating the Number of Events in Demographic Surveys	Theoretical example based on 3 scenarios of completeness	Detailed methods for estimating bias from lack of independence; sampling variance; and total mean square error of completeness.
Nour (1982)	Journal of the Royal Statistical Society. 145 (1) 106-116	On the estimation of the Total Number of Vital Events from Dual Systems	Malawi Population Change survey, 1972	Nour estimate accounting for lack of independence; 4.7 per cent higher than standard DRS estimate; presents method for estimating sampling variance.
Chandrasekaran (1983)	Cairo Demographic Centre Working paper 6	On two estimates of the number of events missed in a dual-record system	Theoretical examples, Indonesian Vital Registration Project	Compares Greenfield method and standard DRS, identifies that error is inversely proportional to the completeness estimate; two estimates provide a plausible range of completeness.
Hook and Regal (1995)	Epidemiologic Reviews. 17(2); 243-264	Capture-recapture methods in Epidemiology	Examples of multiple source data on diseases	Implements separate models for two-source/three-source analysis of missed events, accounting for dependence in each source combination. Proposes use of range of completeness estimates from different methods, rather than any specific variance based standard error calculation of CI.
Ayhan (2000)	Journal of Applied Statistics. 27 (2) 157-169	Estimators of vital events in dual-record systems	Theoretical example based on two sample sources	Method accounts for lack of independence, no details on measurement of standard error of completeness estimate, or of sampling variance.
El-Khorazaty (2000)	Environmetrics; 11; 435-448	Dependent dual-record system estimation of number of Events: a capture-mark-recapture strategy	Vital registration and sample survey data for Egypt, 1974-75	Method accounts for dependence, but assumes no geographic or matching error. Paper compares completeness estimates from data reconciliation; standard DRS; Greenfield; Nour; and El-Khorazaty. No methods for estimating standard error or variance.
Chatterjee and Mukherjee (2013)	arXiv:1311.3812v3[stat.ME. https://arxiv.org /abs/1311.3812	Approximate Bayesian solution for estimating population size from a Dual-record system	Malawi Population survey, 1972	Models account for variations in behavioural response causing dependence between sources. Includes method for estimating Standard error and 95 per cent CI of completeness.

5. HISTORICAL REVIEW OF RECORD LINKAGE STUDIES

In the backdrop of the theoretical concepts of record linkage mechanisms and various principles in their implementation with regard to data collection practices as well as analysis for measuring completeness provided above, a historical review of such studies that have been conducted over the past five decades is presented here, to understand practical experiences and lessons learned from their implementation. This review is presented according to five broad combinations of data sources used in record linkage studies for evaluating completeness of death registration, as summarised in table 4, with relevant examples for each combination of data sources. More details of all studies mentioned in this summary are presented in the Annex tables 1-4.

A. Special registration with household surveys

As mentioned previously, the term "special registration" is applied to any initiative that records vital events, outside the purview of the national civil registration system. The initial phase of the approach towards establishing special registration was launched in the 1960s, under the Population Growth Experiment (PGE) programme (also referred to as the Programme of the laboratories for population statistics—POPLAB) conducted in several countries in Africa, Asia and Latin America, supported by the United States Agency for International Development (Linder, 1971a). The programme aimed to test the use of dual-record systems to measure vital rates, with record linkage mechanisms as a process to evaluate and adjust data for completeness of birth and death registration in a selection of sites in several countries. All these projects were conducted during the period 1961-1975.

Annex table 1 provides more detailed characteristics of a selection of specific projects of this nature conducted in several countries and more details are available from other publications, as listed in a detailed bibliography of the dual-record system method (Carver, 1976). There was considerable variation across the field sites in regard to sampling design, data collection procedures for both data sources and analytical methods. Table 5 below describes some of the various strategies and different options that were tested in regard to methodology of implementation of the dual-record system method, as documented in comparative analyses of studies across several countries (Myers, 1976; Seltzer, 1969). Key issues that were tested include various definitions of study population, sample distribution and size of population clusters, variations in methods for data collection through direct household contact and through key community informants, and periodicity of household surveys. Field experiences were varied across countries for all these characteristics, but in general, provided valuable lessons in conduct of such demographic measurement.

It is informative that not one of the PGE studies involved the official civil registration system as one of the data sources for testing the dual-record systems method. In general, the results of these studies did not yield conclusive evidence on the actual success of the method in estimating either completeness of death recording or in measurement of vital rates. Instead, the studies identified considerable logistical challenges and costs in implementation, as well as methodological issues in regard to establishment of statistical independence, in matching and in measurement of error. As a result, this approach in setting up experimental population research sites with dual-recording procedures to measure vital rates fell into disfavour after the mid-1970s. Despite the challenges in implementation, the dual-record design was continued by the government of India, through its nationally representative sample registration system with in-built mechanisms of periodic surveys with record linkage procedures to assess completeness. Since this approach served as the only viable option for reliably measuring and monitoring vital rates in large populations with CRVS systems under development, this design was also taken up by the governments of Bangladesh and China.

Overall, the key lessons from the PGE studies were that continuous recording methods were found to be more successful in generating data for measurement of vital rates and that the health sector and community contacts had a key role in the notification and recording of vital events. The studies also identified that periodic half yearly or annual retrospective surveys were a viable alternate source for deriving vital event records for evaluating completeness, but careful attention needed to be paid to issues such as cluster sampling, maintenance of independence in data collection, data quality and thorough analysis (Scott, 1974). In general, mounting special registration efforts to measure vital rates is a costly, resource-intensive exercise and in the current environment, it would be useful to integrate DRS methods into the implementation of routine official civil registration systems. Despite all these drawbacks, it should be borne in mind that the PGE studies served an important function in regard to introducing the practice of compiling vital statistics in most of the countries where these studies were conducted. In particular, the establishment of data collection platforms for compiling vital event records served an important function of institutional capacity development as well as human resources for such demographic activities.

The POPLAB programme also convened a series of international consultations and workshops during the period from 1960-1980, bringing together population scientists and government officials from the study countries as well as international experts, to share experiences and plan strategies to strengthen national vital statistics programmes. The POPLAB programme has also produced an extensive series of project documents, field manuals and technical papers on the design, methods and results from the dual-record-system activities, as well as several comparative analysis reports that serve as a comprehensive resource on the dual-record-system methodology for future applications.

More recently, such a design of special registration with household survey-based record linkage was implemented to test death registration reforms in Indonesia during 2005-2010 (Rao and others, 2011; Rao and others, 2010b). The Indonesian experience suggests that special registration sites with independent "completeness" surveys could be implemented as a pilot study or precursor to test major interventions to strengthen death registration and mortality statistics.

The Indonesian experience has identified that in such initiatives, the household survey, which will generate the secondary source of events to measure completeness should be carefully designed to ensure appropriate geographical coverage as well as common reference period with the primary source, to minimise bias from out-of-scope events. Further, the survey sample should be optimised to ensure that the record linkage and analysis is sufficiently detailed to measure levels of completeness by population subgroups (e.g., age, sex, region). Also, missed events identified during the survey should be probed to ascertain reasons for being missed by the registration system. Such reasons will provide inputs to modify the design of the official civil registration system, to strengthen its operations and data quality.

In summary, while the independent surveys may seem resource-intensive, their design and outcomes will fulfil the purposes of accurate estimation of vital rates as well as improve the overall performance and sustainability of the CRVS.

B. National sample registration with periodic surveys

The national sample registration systems of India and China are the key examples of dual-record-linkage systems specifically established for routine measurement of vital rates, which are corrected for completeness of birth and death registration. Both systems share this basic concept in design, particularly in regard to their being established for compiling vital statistics independent of civil registration. Annex table 2 provides details on some key characteristics of the two systems, which are in continuous operation for about five decades in India and about four decades in China. In both countries, data collection and compilation operate as "stand-alone" systems independent of the national CRVS system, as recommended by Linder (1971b). The primary data source is through a continuous recording system, which, in the Indian SRS, is operated by the Registrar General of India administration using a local recorder in each sample site and by local monitoring stations operated by the China Centres for Disease Control (CDC) in the Chinese Disease Surveillance Point (DSP). However, there is a basic difference in the operation of the independent survey for compiling the secondary data source.

TABLE 4. SUMMARY OF SELECTED RECORD LINKAGE STUDIES CONDUCTED DURING 1960-2015, ACCORDING TO COMBINATIONS OF DATA SOURCES

Data source combination	*Countries/year*	*Study methods*	*Completeness measurement*
Special registration with periodic surveys	1960-1975 Pakistan, Egypt, Liberia, Malawi, Philippines, Columbia, Morocco, Turkey, Kenya 2006/07 Indonesia	• Time bound projects (-3 years) in listed countries during 1960-1975 • Tested data collection strategies e.g., direct household contact; using local key informants; combinations • Tested range of recall periods (1,3,6, 12 months) • Indonesian studies in 2006-2007 as sentinel sites, later transformed into national SRS (see below)	• Completeness; estimated by CMR method (ranging from 53 to 90 per cent in different settings); no 95 per cent CI • completeness for 2006 by data reconciliation (no 95 per cent CI); in 2007 by CMR method (with 95 per cent CI)
National sample registration with periodic surveys	India – SRS since 1970 Bangladesh- SVRS - 1980 China DSP since 1990 Indonesia since 2014	• India & Bangladesh – continuous recording in sample clusters with total coverage in routine periodic surveys; • China – continuous recording in sample clusters with triennial subsample completeness surveys; • Indonesia – continuous recording in sample clusters with subsample completeness survey of 2014	• India data reconciliation used to measure mortality, completeness not reported • Bangladesh – completeness measured by CMR method, results not reported • China: completeness estimated by CMR method, results reported with uncertainty intervals for 2015 • Indonesia 2014: Survey findings affected by data quality • fresh survey planned in 2017
Civil registration or death notification with periodic data sources	Thailand (2006) Oman (2010) Philippines (2012-14)* Palestine (2017)*	• Thailand: civil registration and intercensal survey • Oman: civil registration & national census • Philippines and Palestine – civil registration and census (studies yet to be implemented)	• Thailand: completeness by CMR method, with 95 per cent CI; • Oman: completeness by CMR method with 95 per cent CI
Multiple sources with overlapping recall periods	Philippines 2006/7 Viet Nam 2008/9 Kiribati (2001-2009) Tonga (2000-2009)	• Philippines: Civil registration/health system/ parish • Viet Nam: civil registration; health system; peoples committee plus additional partial sources; • Kiribati – civil registration; health information system; reproductive surveillance • Tonga –civil registration; health information system;	• Philippines: CMR method; with 95 per cent CI • Vietnam: completeness by variant of CMR method with 95per cent CI • Kiribati: completeness by, data reconciliation; no 95 per cent CI • Tonga: completeness by CMR method; No 95 per cent CI
Civil registration with special registration	South Africa 2006-09	• Civil registration and HDSS; electronic linkage with deterministic & probabilistic matching;	• completeness not measured due to 'out-of-scope' coverage

FIGURE 1. FEATURES OF POPLAB STUDY METHODOLOGIES

Features of POPLAB study methodologies

1. Sampling methodologies and definition of study population
   - All POPLAB sites used cluster sampling permitting economies in field operation
   - All sites except Turkey decided to include both citizens and non-citizens in the study population
   - All POPLABS except Morocco adopted a modified de jure definition for study population, Pakistan had previously used a de facto definition
   - All sites conducted baseline surveys to define study population, except Colombia and Morocco where the first retrospective six-month survey filled this role

2. Continuous recording data collection methods
   - All sites used some mixture of household visits and community contacts for the continuous recording data source
   - Philippines divided study area to test 3 cluster types each using 1) bimonthly household visits only, 2) community contacts only 3) both at the same time
   - Colombia also designated some samples for household visits and others to rely on community contacts
   - Community contacts used most extensively in Colombia and Philippines
   - In Kenya community contacts were provided a financial incentive to report events

3. Assessment of community contacts vs household visits for continuous recording
   - In Colombia nurses and housewives were very effective community contacts in rural areas but not in urban areas.
   - Overall community contacts from the health sector report births most effectively
   - Overall community contacts seem to work best in rural areas where relationships are more stable

4. Comparison of continuous recording versus household survey
   - Overall household visits have yielded most vital event reports in the continuous recording system.
   - In all sites continuous observation yielded results closest to the CS-D estimates except for births in Morocco and in one of the Philippine areas.
   - In the Indian SRS retrospective surveys have yielded higher completeness
   - Events missed by both methods (statistically derived) ranged from 4-5 per cent for births and up to 13 per cent for deaths (in Kenya)

5. Recall periods and survey frequency in retrospective surveys
   - For retrospective surveys most common reference period 6 months, but intervals of 3, 12 and 18 months have also been trialled. During some experimentation overlapping recall periods were also trialled.
   - six month recall periods produced higher yields of vital events than 12 month periods. Experiments in Philippines showed though that combining yearly or even 18 monthly surveys with more frequent household visits in the continuous recording would produce the same completeness outcome
   - three monthly retrospective surveys however were of less value as they are not cost effective and they lead to duplicated event reporting and respondent fatigue
   - Philippines and Kenya have each developed supplementary samples to guard against respondent fatigue

6. Relationship between continuous recording and retrospective surveys and ensuring independence
   - In all sites interviewers and recorders not assigned to work in the same area at the same time.
   - In Philippines alertness premium paid to the recorded or interviewer who first records a vital event

The Indian SRS conducts dual system survey to cover the entire SRS population every six months by an independent surveyor (Lingner and Wells, 1973). Moreover, the six-monthly survey also includes a complete population enumeration component, which provides accurate denominators for computation of vital rates. On the other hand, the Chinese Disease Surveillance Point system implements the independent survey to collect the secondary set of vital events only in a small subsample, and only once in approximately three years (Guo and others, 2015; Yang and others, 2005). The sub sample comprises all households within one rural township or urban street in each DSP site, which is deemed representative of the economic level of the respective DSP. In each selected township/street, an initial list of vital events in the reference period (usually three years) is prepared by resident group leaders, based on recall. Subsequently, this list is tallied with data on reported deaths maintained by the public security department, the family planning department, and the maternal and child health department, to develop a reconciled list. Each household on the list is approached by field surveyors who verify the event and update the record. Unlike the Indian SRS, the DSP independent survey only visits households from where a death had been either recalled by the resident group leader or had been recorded by one of the local data recording sources, rather than a visit to each household in the township/street. This creates the potential for events to be missed in all the available sources. Also, unlike the SRS, the DSP relies on local population estimates from the China CDC in each site, rather than conducting its own regular population enumeration to keep track of denominators for accurate vital rates.

In both systems, statistical independence between the continuous recording and the periodic survey is not assured, due to intended collaboration between data collectors in the two processes. As a result, the Indian SRS only uses a reconciled list after record linkage to measure vital rates (Padmanabha, 1981), without attempting to estimate the number of events potentially missed by both sources. However, the Chinese DSP uses the CMR method completeness factor to adjust the recorded data to estimate vital rates. In its most recent report, the Chinese DSP also reports the 95 per cent confidence interval of completeness by sex, age, and geographical region, although the method used to derive the standard error is not reported. For instance, it would be useful to know whether the standard error accounted for sampling variance in the secondary survey study population. While both systems are considered to generate the most reliable vital statistics for their respective populations, there are several drawbacks, mostly in terms of limitations in regard to sample size for subnational mortality estimation, and considerable operational costs. Currently, the Chinese DSP is being redesigned to address the limitation of sample size, along with alignment with the routine Ministry of Health Vital Registration System that covers a larger fraction of the national population.

Bangladesh has also implemented a Sample Vital Registration and Statistics system (SVRSS) which has been in operation by the Bangladesh Bureau of Statistics for about three decades (Bangladesh Bureau of Statistics, 2011). This too is independent of the national civil registration system, which is still under development. The Bangladesh SVRSS operational methodology is similar to that of the Indian SRS, except for the fact that the independent household surveys are conducted once in 3 months, and that the capture-mark-recapture method is used to assess completeness and adjust vital rates. However, completeness measures are not reported. In 2002, the system was redesigned with a new sampling scheme, increasing the total sample units to 1000, covering a population of over 1 million people. However, there have been some inconsistencies in regard to data quality, as a result of which the resultant vital statistics are of limited utility at national level.

Another example of dual-record systems of this nature is the recently established Indonesian Sample Registration System. This is an extension of the ‘special registration’ project described earlier, the lessons from which were expanded from pilot sentinel sites into a nationally representative sample. The current sample covers 128 sub districts and a total population of roughly 8.7 million people (Pratiwi and Kosen, 2013). The unique nature of the Indonesian SRS is that the local compilation of vital statistics is through a consultation of multiple stakeholders including the civil registration system, which also benefits through augmenting its records from information on events recorded in other sources. It relies on local population estimates for each sample site from the Statistics Office (*Badan Pusat Statistik (BPS)*). A methodology involving the local BPS office in conducting the independent household survey as the secondary data source for evaluating completeness has been trialled, but this needs to be standardised for routine implementation.

In summary, the success of the Indian and Chinese systems suggests that Sample Registration Systems are potential interim steps for measuring vital rates in countries with large populations, while efficient national civil registration systems are under development. Also, given the uncertainty in the indirect demographic techniques, dual-record-linkage methods have been demonstrated to be a viable alternative for estimating completeness of vital registration (VR) in these systems.

C. Civil registration with periodic data sources

In the current environment, the civil registration system is the primary continuous recording system that requires completeness assessment in many developing countries, particularly those with well-established systems or rapidly improving systems. Periodic national level household surveys offer the best option for secondary sources for record linkage mechanisms, given the potential for statistical independence. However, the range of design characteristics of these surveys affects their utility for exercises in evaluating completeness. For instance, the national census is the optimal survey design with its national coverage, if there are specific modules to record recall-based information on vital events. On the other hand, other nationally representative surveys such as the DHS and MICS surveys, health behaviour and risk factor surveys, and socio-economic surveys are usually limited by sample size considerations. Two examples of the use of survey based information for evaluating completeness are described here, with potential lessons for replication in other countries where such secondary data sources are available or can be made available in the coming years.

Thailand is a key example of a country with a long-standing history of vital registration, along with a robust programme of decennial sample intercensal surveys as a secondary source for record linkage. The survey in 2005 was particularly advantageous since it also recorded the unique citizen identifier of each reported death, which facilitated the record linkage and CMR analysis. However, a detailed analysis applying record linkage methods found that 11.3 per cent of events could not be matched due to missing identifiers, and this could have biased the measured completeness of 91 per cent (Vapattanawong and Prasartkul, 2011). Moreover, the sample size limited the completeness analysis to the national level. Nevertheless, the relatively clear basis for statistical independence between the civil registration and the intercensal survey based data collection is an advantage of this study design for record linkage based assessment of completeness.

More recently, a record linkage study was conducted in Oman, using the national birth and death notification system database as the primary source and deaths recorded in the 2010 census as the secondary source (Al Muzahmi, 2015). Given the national coverage of both data collection systems, the study yielded completeness estimates for sex and broad age groups, as well as for major regions of the country. One challenge faced in this study was that the census data on deaths (based on a 12-month recall) was roughly 10 per cent lower than deaths from the death notification system and this could affect the results from the linkage study. However, as in the Thai study described above, there is a greater potential for statistical independence between the data collection processes of the two data sources. In the Oman case, there was no unique personal identifier for each death, hence the actual matching process involved a combination of electronic and manual processes with explicit rules, supplemented with the use of implicit rules in about 10 per cent of the cases, which needed local inputs to adjudicate on potential matches.

As mentioned earlier in regard to the potential for "out-of-scope" events, there was some potential for confusion in regard to the recall period in the census, given that the date of the census was on 12 December 2010. However, the data analysis revealed that there was potentially accurate recall of the date of death, as determined from the absence of specific heaping of deaths at either the beginning or the end of the reference period, which would have been indicative of problems with recall bias. From another perspective, however, there was a problem in the reporting of the age at death, with considerable age mis-statement in both vital registration and census, necessitating the use of implicit criteria (setting tolerance bounds of five years around age at death) and using local opinion for verification of matching based on additional criteria such as sex and address variables.

These studies have demonstrated the viability in using national data collections such as the census or intercensal surveys as secondary data sources for assessing the completeness of vital registration through record linkage mechanisms. The likelihood of statistical independence allows for applying the capture-mark-recapture method to estimate completeness and, particularly in the case of the census, the national coverage enables disaggregated measures of completeness by sex, broad age group and subnational divisions. While other national surveys may be limited in their scope for such detailed analysis of completeness, the availability of electronic data on vital events from such surveys along with reliable matching information could be used for evaluating completeness with minimal logistical resources. However, an important aspect that requires attention is the emphasis on survey data quality in terms of minimizing the potential for recall bias, as well as accuracy of recorded variables.

D. Multiple data sources with overlapping recall periods

In several countries, there are multiple continuous recording systems that record births and deaths. Most commonly, the systems include the national civil registration system and the national health information system, which both have a mandate to maintain death registers. Such parallel data sources offer prospects for data reconciliation across sources to improve completeness (as described in the next section), or to undertake analyses using capture-mark-recapture methods to evaluate completeness.

In recent times, attempts have been made in several studies to utilise multiple data sources to improve the measurement of vital rates. In a study conducted in the Pacific Island nation of Kiribati, records from civil registration and the health information system were linked and reconciled to measure mortality rates. Owing to the small population size, the matching exercise was conducted on a dataset aggregated across ten years (2000-2009) (Carter and others, 2016). Both data sources were found to be considerably incomplete when compared to the reconciled data set, limiting the potential to apply CMR methods, since the adjustment factor yielded implausibly high mortality rates. In another study on Tonga, death records were reconciled across three sources—the civil register, the health information system, and a partial data source—the reproductive health surveillance system (for neonatal and infant deaths). Given the dependence between the two health data sources, they were merged, and a standard CMR analysis using the civil registration and the merged health data source, was conducted to evaluate completeness of the civil registration and derive adjusted vital rates (Hufanga and others, 2012).

A study in the Philippines in 2006-2007 linked data from civil registration, health centre records and local church parish records (Carter and others, 2011b). A detailed review of the data collection procedures in the three data sources revealed communication between the civil registration and health centre records, potentially violating the assumption of statistical independence. The analytical protocol tested several two-source models across different combinations, as well as a three-source CMR analysis according to the method proposed by Hook and Regal (1995). The authors eventually chose to use an average of the estimate of completeness derived from all the analysis models. A multiple source analysis in Central Java, Indonesia in 2007, using information from the civil registration, health information system and a special independent survey, found that data reconciled across the civil registration and health datasets accounted for more than 90 per cent of all reconciled deaths, indicating that routine processes of data reconciliation across these two sources would improve the overall data completeness (Rao and others, 2010b). However, lack of independence between the civil registration and the health system records precluded the use of capture-mark-recapture methods. Also, the independent survey data was of low quality, hence could not be used in the completeness analysis.

Finally, a detailed study using multiple sources was conducted in a nationally representative sample of population clusters in Viet Nam in 2009. The three overlapping data sources were the civil registration system operated by the Ministry of Justice, the health information vital event registers maintained by the Ministry of Health and the Commune Population and Family Planning Committee household registers (Hoa and others, 2012). A combination of data reconciliation and DRS analysis was used to evaluate completeness as well as measure vital rates. The sample size was adequate to measure completeness by sex and broad age groups. The analysis applied the bootstrapping technique to measure the standard error of the completeness estimate.

E. Civil registration with special registration

As described in table 1, special registration sites operated under the INDEPTH Network are the only viable source of continuous vital statistics for these areas in several countries where civil registration systems are yet to be established on a national scale (notwithstanding the use of censuses or surveys to collect retrospective data at discrete time points). In these countries, as civil registration systems undergo development, the special registration sites offer the potential to become secondary data sources to evaluate the completeness of civil registration data. An attempt at such analysis was made in South Africa, but the record linkage process could not lead to an evaluation of completeness of the civil registration (Joubert and others, 2014). This was because the geographical and administrative boundaries of the special registration area did not coincide with those of the local civil registration system, as a result of which, there was potential for "out-of-scope" events to bias the completeness analysis. In addition to this fundamental issue, there were also challenges in the quality of data variables in the civil registration system, which is a key factor that affects the record linkage process. Nevertheless, the analysis indicated the potential for such an approach over the coming decade or so, to evaluate the progress in the development and data quality of civil registration in other African and Asian countries with demographic sites operated under the INDEPTH network (Sankoh and Byass, 2012). Care would be needed to reconcile the definition and coverage of the two data sources to ensure their compatibility for the record linkage and, if possible, the completeness analysis.

F. Other early experiences with record linkage for evaluating completeness of vital statistics

The earliest information on the utility of linking vital records from two different sources to assess completeness of birth and infant death registration is documented in the "Rules for Statistical Practice" by the American Public Health Association in 1908 (as cited in United Nations (1955)). These rules required the comparison of lists of registered births and infant deaths with independent lists of same events as may be available from newspaper reports and cemeteries. It was realised that such lists may be biased in regard to covering more affluent as well as more urban segments of the population. There is no report on the results of such record linkage efforts. During 1943-1945, the New York City Public Health Department conducted a linkage study between foetal deaths recorded in the civil register and a special data collection initiative under the Emergency Maternity and Infant Care programme, and identified that only 56 per cent of EMIC foetal deaths were registered. (as cited in United Nations (1955)).

Other secondary sources for linkage and assessment of completeness that were used included the "Teachers Return of School Enrolments" in Canada, mail survey lists of births from households in the United States of America (1934-1935), and household contact censuses in Scotland (1916), Canada (1931 and 1941), and the United States of America (1940 and 1950), Panama and Venezuela (1950), Chile and Paraguay (1952), Yugoslavia (1951—birth and death registration), and Ceylon (now Sri Lanka) in 1953 (United Nations, 1955). All these record linkage studies only evaluated completeness as a proportion of census reported events that were officially registered, without accounting for events missed by both sources. The potential for under-reporting bias in both sources, which affects the reliability of the estimate, were noted. The findings varied from 81 per cent in Scotland, 88 per cent in Ceylon, improvement from 92 per cent to 98 per cent in the United States of America between 1940 and 1950; and 98 per cent in Canada. Findings from the efforts in other countries are not available.

G. Contemporary record linkage experiences with data reconciliation to improve completeness

In most countries, several parallel vital event recording systems exist, some of them being only partial data sources, as defined earlier. There have been a few illustrative case experiences of routinely linking records across these multiple systems to identify any potentially missed records in the civil registration data, followed by data reconciliation to strengthen the overall completeness of vital statistics. Examples from Iran, Brazil and Turkey demonstrated the processes and results of such data reconciliation efforts.

In Iran, until 2002, there were two separate agencies that compiled mortality data, the Iranian National Organization for Civil Registration (NOCR) and the Iranian Ministry of Health and Medical Education (MOH ME) (Khosravi, 2008). In general, the data from the two systems showed discrepancies in regard to completeness and data quality. Since 2002, a new comprehensive mortality statistics system is operated by the MoH ME (Deputy of Health). At the district level, to achieve complete registration, the District Health Centre sources and reconciles mortality information from all health sector institutions including hospitals and maternity units, rural health centres and health houses in remote areas, urban health centres, and bureaus of legal medicine. In addition, the District Health Centre also accesses records from all local cemeteries. To avoid duplication of recorded deaths, data from the District Health Centre are further cross checked with data from the local office of the National Organization for Civil Registration, according to identity variables, to remove duplicates and update records in both data sources. A Farsi computer programme is implemented at the District Health Centre and, all data for each case, including personal information of deceased (first name, family name and father's name), address, age, sex, date of death, area of residence and causes of death, are entered into the computer programme.

Data are sent from the district to the province level, where nominated medical universities from the MoH ME perform data quality assessment checks and primary data analysis, and provide feedback to District Health Centres for query and updating of records. At the end of the year, data are sent to the central MoH ME office for analyses and publication. Substantial progress has been made in integrating data from various systems into a single coordinated death registration and statistics system. The revised system also includes detailed procedures to improve the accuracy of recorded causes of death. In summary, the mortality statistics programme reforms started as a pilot project in one province in 1997, but within a decade was successfully expanded to cover all 30 provinces by 2007 (Khosravi, 2008). It is an illustrative experience of system development and standardisation in pilot sites, followed by incremental expansion. The effect has led to the availability of improved vital statistics at both the national and subnational level.

Another country that has consistently used record linkage mechanisms with data reconciliation to improve the overall completeness of death registration is Brazil. Despite a long-standing history of vital registration, the compilation of mortality statistics remained a problem in Brazil. In the 1970s, the Brazilian Institute of Geography and Statistics (IBGE) was tasked with compilation of vital records from the Civil Registration Office in each municipality, while the Ministry of Health (MoH) commenced the compilation of data on causes of death from hospitals. Since then, the Ministry of Health has forged close collaboration with civil registrars to register deaths and their causes for events that occur outside of hospitals, and merge the data with hospital events to generate a standard mortality data base in the MoH Mortality Information System (Franca and others, 2008).

In addition, in North Eastern States experiencing the greatest deficiencies in death registration, over the past two decades, the Ministry of Health had undertaken numerous studies to actively search and record infant deaths from several alternate sources, including primary health care units, institutes of forensic medicine, ambulance service, burial sites, funeral homes and midwife groups (Szwarcwald, 2008; Szwarcwald and others, 2014). These studies used record linkage mechanisms to identify deaths missed by the official civil registration and subsequent data reconciliation enabled accurate measurement of infant mortality rates. An additional measure that has enhanced overall completeness was the introduction of a provision for government health funding to municipal health offices to be disbursed only upon satisfactory submission of mortality data for the preceding reporting period. These initiatives, in combination with the record linkage and data reconciliation mechanisms, have vastly improved the overall completeness of Brazilian mortality data to over 90 per cent at the national level, with some subnational variations.

Turkey provides a third example of record linkage and data reconciliation being used as a strategy to strengthen completeness of national mortality statistics. The civil registration system in Turkey records death events using the Central Population Administrative System (MERNIS) reporting form both urban and rural areas, and statistics from MERNIS forms are compiled by the Ministry of Interior. In urban areas, the Turkish Statistical Institute operates another system—TURKSTAT Death Registration System (DRS), in which a separate form is used for reporting deaths to TURKSTAT by civil registrars. Neither system was capturing the entire set of death events in the country. Since 2009, the MERNIS and TURKSTAT-DRS records are linked and reconciled at the level of the Provincial Health Directorate and recent analysis of the overall merged datasets has identified very high levels of data completeness, albeit using indirect demographic techniques (Ozdemir and others, 2015). Other recent analyses have confirmed this improvement in data completeness, with some deficiencies in registering deaths of those over 75 years of age (Yayla and Cavlin, 2017).

A final example of the use of record linkage and data reconciliation is available from the Republic of Korea, which has implemented a detailed record linkage programme since 1999 to evaluate infant mortality. A range of sources, including the civil register, hospital records, family registers (mother and child health registration) and crematorium records are reviewed and linked to identify infant deaths. The programme identified under-reporting of about 50 per cent of infant deaths in 1999, with reconciliation of records to more accurately measure infant mortality. Under-registration of infant deaths continues to be an issue in the Republic of Korea, although it has now reduced to 37.6 per cent (Lee and others, 2016).

In summary, all these countries have implemented linkage and reconciliation with demonstrated improvements in data completeness, without undertaking any additional analysis to estimate events that were potentially missed by either or several different data sources used in the review. These experiences of Brazil, Iran, the Republic of Korea and Turkey demonstrate the value in the use of record linkage to strengthen completeness of data from death registration systems, and to incrementally help to improve data quality.

6. Lessons learned from five decades of record linkage studies

The lessons learned from use of record linkage methods and subsequent analytical techniques to assess completeness of vital registration and/or improve the quality of mortality statistics over the past five decades are rich, extensive and varied. While it will be difficult to completely document all lessons learned from each of the previous attempts at such methods, the key issues are summarised here.

A. Lessons from special registration based dual-record systems

The POPLAB experimental studies during the 1960-1980 period, identified that conducting specifically designed and funded activities in small scale discrete population clusters with dual or multiple vital event recording processes is not a viable approach for either measurement of vital indicators or to evaluate vital statistics in either source. However, the experience that continuous recording systems usually performed better, particularly with close contact with a network of local informants, is an important lesson for strengthening vital event notification and registration in the current context of strengthening local civil registration and vital statistics systems. Two additional important lessons from the POPLAB studies are the importance of accurate recording of all data variables in any data source of vital events, and the need to establish clear definitions of population coverage, reference time periods and data collection mechanisms to assess statistical independence of data sources, as a precursor to applying capture-mark-recapture analyses to evaluate data completeness. Finally, these population growth experimental studies helped build considerable local institutional and individual capacity in demographic research and field experience in death recording and compilation of statistics, and the approach and successes in such capacity building serve as important lessons when initiating vital statistics projects in previously uncounted populations.

B. Lessons from sample registration systems

The sample registration systems in India, Bangladesh and China have demonstrated success as special registration dual-record systems, as a result of sustained implementation over several decades. These systems have been sustained by intensive national government programme operations, given their critical importance as the only viable local data on vital statistics in these countries. Although the mechanisms of dual-record operations vary across these three countries, the experiences provide important evidence on the practical success in implementation. One of the key lessons from the Indian SRS is the reliable quality of the data, arising from meticulous attention to implementation of record linkage and data reconciliation, along with regular denominator population enumeration. However, a major drawback has been that in India, the SRS has not expanded the overall size of the sample population under surveillance (this has remained around 0.5 per cent of the national population), owing to the extensive resources for operating the dual-recording systems across the entire study population. This has critically limited the potential of the SRS to be used to reliably estimate mortality patterns by age, sex and cause at subnational level.

On the other hand, the Chinese DSP has demonstrated the value in undertaking incremental expansion of the study sample over time, without intensive attention to the secondary data source, or regular population enumeration. However, repeated evaluations of completeness have indicated that overall data completeness has remained only around 85 per cent or so since three decades, with true completeness likely to be lower, owing to the lack of statistical independence. Also, there is no adjustment for sampling variance in the completeness estimate from DSP and this is a critical element in assessing precision/uncertainty of the completeness estimate. In the case of the Bangladesh SVRSS, although the capture-mark-recapture methodology is used, completeness measures are not reported.

The key lesson from sample-based dual-record systems is that there is a need to establish independence between data sources in order to implement capture-recapture methods to assess completeness, or that there should be more rigorous data analyses to account for potential bias due to lack thereof. In regard to data collection, there should be a process for field verification of at least a sample of events to measure any net matching error. Close attention is required to measure denominator populations, particularly where there is potential for regular migration. Completeness measures must be reported with error margins, which should also take into account effects of sampling variance, as well as matching error.

Finally, findings from record linkage and field verification must be analysed to identify reasons for events being missed by the continuous recording data source and these should be addressed through appropriate interventions to strengthen the completeness of the routine data collection process. Such interventions will enhance the quality of routine data from the sample-based system, and this could reduce the need for resource intensive dual-record systems; as well as create the potential to expand the enhanced quality data collection processes in an augmented study population for subnational vital statistics. From a broader perspective, there is a need for integration of these sample-registration systems with national civil registration, such that there could be common reporting and recording procedures to reduce operational costs in registration, but there could be more focussed attention on data quality and compilation of statistics in a representative sample of clusters, the coverage of which could be gradually scaled up over time.

C. Lessons from record linkage across multiple parallel data sources

The key lesson from applying record linkage methods across multiple sources to measure completeness is that it is very challenging to establish independence across all sources. However, an analytical approach has been devised through statistical models that account for dependence across various combinations of the multiple data sources, along with the recommendation to average the completeness measures derived from each of the different multiple source analytical models. The range of completeness measures could serve as the confidence interval for the point estimate of completeness. However, another aspect to note from this methodology is that one or more data source may only be partial, hence not be appropriate to include in the multiple source models for evaluating completeness.

D. Lessons from record linkage with data reconciliation

Record linkage with data reconciliation across two or more sources to derive a numerator to estimate vital rates, as in the processes employed in Brazil, Iran and Turkey, are viable options, without evaluating or correcting for completeness. This can be of particular value if one or more individual data sources are, *a priori,* believed to be of relatively high completeness. Also, in such instances it is often difficult to establish statistical independence, as required for implementing the capture-recapture analysis. In the context of strengthening civil registration and vital statistics, record linkage with data reconciliation is a major intervention required to improve data quality and, as mentioned earlier, can be additionally beneficial if reasons for not recording events in the routine registration system are identified and addressed to strengthen data compilation practices. Regular data verification with accurate identification and deletion of duplicates will improve the accuracy of estimation/measurement of vital rates.

7. CONCLUSIONS AND RECOMMENDATIONS

Record linkage mechanisms are of proven value to improve the overall universe of registered vital events and also to enable the estimation of data completeness of either source, based on certain conditions. Record linkage provides specific information to detect biases in the data from the two sources, which is essential for measurement of error in resultant completeness estimates. This is in contrast to the indirect analytical methods, the assumptions for which are demographic in nature and are dependent on the characteristics of the underlying population, rather than the actual nature of recording of events. At the same time, it must be appreciated that meeting the condition of statistical independence of data sources, in order to implement the capture-mark-recapture method to evaluate completeness, is a considerable challenge. Probably, the most important limitation is that the probability of non-reporting of events in either or all sources is a result of human error, rather than chance, and this restricts the statistical assumptions in the computation of events missed by both sources. Nevertheless, within the environment of imperfect data from current civil registration and vital statistics systems, record linkage methods are a useful tool which can be applied alongside other demographic analytical techniques, to understand the potential range of data completeness. The following recommendations can be considered when planning and undertaking record-linkage-based methods to measure or improve completeness of vital statistics.

1. The detail and accuracy of information recorded in each data source for critical variables, including those for identity of individual registered (name, age, sex); location (street address, village, suburb, post code); and event (date of birth/death/registration); is of paramount importance for compilation of accurate vital statistics as well as for enabling linkage with other data sources.

2. Accuracy of names, in particular, is very important, both in terms of correctness of spelling as well as in recording of complete name(s) as per any available official records (e.g., birth certificate, passport, driving license).
3. Wherever feasible, unique identifiers should be included (e.g., citizen ID, social security number, tax file number), to improve the potential for matching. However, this could be subject to local government rules in regard to variables that could be included in civil registration records for vital events.
4. Similar levels of detail and accuracy of information is necessary and should be ensured in all other data compilations of individual vital events, for correct matching.
5. Computerisation of civil registration records, preferably at the point of registration, can considerably enhance the potential for data verification and accuracy, since it can enable timely local enquiry and updating of records. This will enhance the degree of accuracy in matching.
6. In regard to the choice of primary and secondary sources, a defined hierarchy of study designs based on statistical independence, costs and maximizing sample size to improve potential for sub-group analysis, can be considered as follows:
    (a) CRVS as primary source with census based recall of births/deaths as secondary source is the optimal study design, the main limitation being that such analyses are possible only once a decade.
    (b) The next best option to evaluate CRVS is to conduct record linkage with routine intercensal surveys or other routine/periodic nationally representative sample surveys. Sample size limitations would restrict the scope for subgroup analysis, but these evaluations could be conducted more frequently, based on national survey programmes at country level.
    (c) Depending on feasibility in regard to subgroup analysis, completeness should be evaluated according to various comparative subgroups including gender, urban/rural divisions, subnational administrative divisions and by major age groups e.g., 0-4 years; 5-14 years; 15-49 years, 50-69 years; and 70 years and above. This will improve the utilisation of completeness factors in adjusting vital rates. However, this will also require a high degree of accuracy in the quality of recorded variables in both data sources, to minimize non-matches owing to cross-over to the neighbouring subgroup due to data errors.
    (d) Where CRVS systems are generally weak, it is recommended to implement system strengthening activities in a sample of geographic/administrative areas to develop a "sample vital statistics" compilation programme, to measure vital rates. In such an environment, there are options to conduct periodic special surveys to establish the secondary data source for record linkage studies. In contrast, there may be other parallel vital event recording systems, which could be used for record linkage with data reconciliation and/or completeness assessment using capture-mark-recapture methods.
    (e) Where nationally representative population samples are not feasible due to cost or technical resource constraints, special registration in targeted surveillance sites as conducted under the INDEPTH network could serve as the primary data source, with a periodic special survey or linkage with other local parallel data source to evaluate completeness of the special registration data.
7. Computerisation of vital events recorded in all sources will also enable the potential for electronic record matching and this should be promoted by all data collection agencies.

8. In regard to measurement of infant mortality, record linkage should be attempted between birth registers and death registers, as well as records from reproductive and child health programmes, to improve the capture of vital events.

9. In all instances, where CRVS is the primary source, record linkage should be followed by data reconciliation to strengthen the overall set of registered events and contribute to system strengthening efforts, as observed from the success stories of Brazil, Iran and Turkey.

10. Any record linkage study should have clear documentation in regard to description and definitions of study population, data sources, reference period, matching variables and rules for matching. Use of explicit deterministic/probabilistic criteria along with tolerable margins and with any implicit judgement/adjudication of matches, should be adequately described, to enable evaluation of bias and interpretation of results.

11. All record linkage studies, whether electronic or manual, should include a field component to verify a sample of unmatched as well as matched events. This will enable measurement of net matching error to be accounted for in the evaluation of error in the completeness estimate. Findings from verification of matching are also required in some of the methods for assessment of statistical independence between sources.

12. Record linkage studies to evaluate completeness should report the standard error of the estimate, which should include an assessment of bias from statistical dependence (see table 3 and associated text), matching error and out-of-scope events, as well as any variance arising from sampling methodology. Where applicable, cluster design effects should also be accounted for in estimation of variance.

13. It is recommended that additional research should be undertaken to assess the utility of the multiple methodologies that have been developed to account for statistical dependence, bias and error in the application of the capture-mark-recapture technique (see table 3). This research should apply all available methods on select sample datasets from dual-record systems and, where feasible, validate the completeness estimate from different methods against a reference standard.

14. Range of completeness measures from different methods could be used to conduct sensitivity analysis to estimate the potential range of population-level vital rates.

15. In addition to their use in evaluating completeness, record linkage mechanisms should also be used to identify specific subgroups and/or population characteristics that are associated with lower registration, leading to design of specific interventions to enhance registration of vital events in these subgroups.

16. In the current and future environment of availability and utilisation of electronic data sources, greater attention should be paid to the principles of ethics, privacy and data confidentiality in record linkage studies to evaluate completeness of vital statistics.

References

Abernathy, J., and H. Booth (1977). *The dual record system: retrospective survey subsystem*. Laboratories for Population Statistics, Scientific Report Series, No. 29, North Carolina, USA, University of North Carolina, Chapel Hill.

Adlakha, A., and others (1977). *The dual record system: Sampling design - The POPLAB experience*. Laboratories for Population Statistics, Scientific Report Series, No. 30, North Carolina, USA, University of North Carolina, Chapel Hill.

Al Muzahmi, S.N.K. (2015). Mortality patterns in Oman: Demographic and Epidemiological Review. PhD Thesis, Epidemiology Dept., University of Queensland.

Ayhan, O. (2000). Estimators of vital events in dual record systems. *Journal of Applied Statistics*, vol. 27, No. 2, pp. 157-169.

Bangladesh Bureau of Statistics (2011). *Report on Bangladesh Sample Vital Registration System 2010*. Dhaka: Statistics Division, Ministry of Planning.

Blacker, J.G.C. (1977). Dual record demographic surveys: A reassessment. *Population Studies*, vol. 31, No. 3, pp. 585-597.

Carter, K.L., and others (2016). Mortality and life expectancy in Kiribati based on analysis of reported deaths. *Population Health Metrics*, vol. 14, No. 1:3, pp. 1-7. doi: 10.1186/s12963-016-0072-6.

Carter, K.L., and others (2011a). Mortality trends in Fiji. *Aust N Z J Public Health*, vol. 35, No. 5, pp. 412-20. doi: 10.1111/j.1753-6405.2011.00740.x.

Carter, K.L., and others (2011b). Capture-recapture analysis of all-cause mortality data in Bohol, Philippines. *Population Health Metrics*, vol. 9, No. 1:9, pp. 1-9. doi: 10.1186/1478-7954-9-9.

Carver, J.S., ed. (1976). *Bibliography on the Dual Record System*, *Laboratories for Population Statistics, An Occasional Publication*. North Carolina, USA: University of North Carolina, Chapel Hill.

Chandrasekar, C., and W.E. Demings (1949). On a method of estimating birth and death rates and the extent of registration. *Journal of the American Statistical Association*, vol. 44, No. 245, pp. 101-115.

Chandrasekaran, C. (1983). *On estimates of the number of events missed in a dual record system*. Cairo Demographic Centre Working Paper, No. 6, Cairo, Egypt.

Chanlett, E., and M.D. Fichet (1976). *The dual recording system: Vital event recording subsystem*. Laboratories for Population Statistics, Scientific report Series, No. 27, North Carolina, USA, University of North Carolina, Chapel Hill.

Chatterjee, K., and D. Mukherjee (2013). Approximate Bayesian solution for estimating population size from dual record systems. *arXiv preprint arXiv:1311.3812*.

Dusetzina, S.B., and others (2014). *Linking Data for Health Services Research: A Framework and Instructional Guide (Prepared by the University of North Carolina at Chapel Hill under Contract No. 290-2010-000141)*. Methods for Effective Health Care AHRQ Publication, No. 14-EHC033-EF, Rockville (MD), USA, Agency for Healthcare Research and Quality (US).

Eisele, T.P., and others (2003). Evaluating the completeness of demographic surveillance of children less than five years old in Western Kenya: A capture-recapture approach. *American Journal of Tropical Medicine and Hygiene*, vol. 69, No. 1, pp. 92-97.

El-Khorazaty, N.M. (2000). Dependent dual record system estimation of number of events. *Environmetrics*, vol. 11, No. 4, pp. 435-448.

Fellegi, I.P. (1974). *Some aspects of the survey design for the Moroccan dual record experimental design*. Laboratories for Population Statistics, Scientific Report Series, No. 16, North Carolina, USA, University of North Carolina, Chapel Hill.

Figueiroa, B.D.Q., and others (2013). Analysis of coverage in the Mortality Information System in Olinda, Pernambuco State, Brazil. *Cadernos de Saude Publica*, vol. 29, No. 3, pp. 475-484.

Franca, E., and others (2008). Evaluation of cause-of-death statistics for Brazil, 2002-2004. *International Journal of Epidemiology*, vol. 37, No. 4, pp. 891-901. doi: 10.1093/ije/dyn121.

Frias, P.G., and others (2010). Evaluation of data on mortality and live births in Pernambuco State, Brazil. *Cadernos de Saude Publica*, vol. 26, No. 4, pp. 671-681.

Greenfield, C.C. (1976). A revised procedure for dual record system in estimating vital events. *Journal of the Royal Statistical Society*, vol. 139, No. 3, pp. 389-401.
Guo, K., and others (2015). Propensity score weighting for addressing under-reporting in mortality surveillance: A proof-of-concept study using the nationally representative mortality data in China. *Population Health Metrics*, vol. 13, No. 1:16, pp. **1-11**. doi: 10.1186/s12963-015-0051-3.
Hoa, N.P., and others (2012). Mortality measures from sample based surveillance: Evidence of the epidemiological transition in Vietnam. *Bulletin of the World Health Organization*, vol. 90, No. 10, pp. 764-772.
Hook, E.B., and R.R. Regal (1995). Capture-recapture methods in epidemiology: Methods and limitations. *Epidemiologic Reviews*, vol. 17, No. 2, pp. 243-264.
Hufanga, S., and others (2012). Mortality trends in Tonga: An assessment based on a synthesis of local data. *Population Health Metrics*, vol. 10, No. 1:14, pp. 1-11. doi: 10.1186/1478-7954-10-14.
Joubert, J., and others (2014). Record-linkage comparison of verbal autopsy and routine civil registration death certification in rural north-east South Africa: 2006–09. *International Journal of Epidemiology*, vol. 43, No. 6, pp. 1945-1958. doi: 10.1093/ije/dyu156.
Khosravi, A. (2008). Evaluation of completeness and data quality of mortality information systems in Iran. PhD Thesis, Epidemiology Dept., University of Queensland.
Krotki, K.J., ed. (1978). *Developments in Dual System Estimation of Population Size and Growth*. Edmonton: The University of Alberta Press.
Lee, J.-Y., and others (2016). Enhancing Korean infant mortality statistics by linking administrative data and surveys. Paper presented to the United Nations Expert Group Meeting on the Methodology and Lessons Learned to Evaluate the Completeness and Quality of Vital Statistics Data from Civil Registration. 3 and 4 November 2016, New York, USA.
Linder, F.E. (1971a). *The concept and the program of the laboratories for population statistics*. Laboratories for Population Statistics, Scientific Report Series, No. 1, North Carolina, USA, University of North Carolina, Chapel Hill.
__________ (1971b). Vital event numeration system as a new tool for measuring population change. *Bulletin of the International Statistical Institute*, vol. 44, No. 1, pp. 377-393.
Lingner, J.W., and H. Wells (1973). *Organization and methods of the dual record system in India*. Laboratories for Population Statistics, Scientific Report Series, No. 9, North Carolina, USA, University of North Carolina, Chapel Hill.
Machado, C.J., and K. Hill (2004). Probabilistic record linkage and an automated procedure to minimize the undecided-matched pair problem. *Cadernos de Saude Publica*, vol. 20, No. 4, pp. 915-925. doi: /S0102-311X2004000400005.
Mahapatra, P., and others (2007). Civil registration systems and vital statistics: Successes and missed opportunities. *Lancet*, vol. 370, No. 9599, pp. 1653-63. doi: 10.1016/S0140-6736(07)61308-7.
Marks, E.S., and others (1974a). Choosing appropriate techniques: Matching error. In *Population Growth Estimation: A handbook of vital statistics measurement*, pp. 101-120. New York: The Population Council.
__________ (1974b). Choosing appropriate techniques: Sampling and sampling error. In *Population Growth Estimation: A handbook of vital statistics measurement*, pp. 124-132. New York: The Population Council.
__________ (1974c). *Population Growth Estimation: A handbook of vital statistics measurement*. New York: The Population Council.
Mathers, C.D., and others (2005). Counting the dead and what they died from: An assessment of the global status of cause of death data. *Bulletin of the World Health Organization*, vol. 83, No. 3, pp. 171-7. doi: /S0042-96862005000300009.
Murray, C.J., and others (2010). What can we conclude from death registration? Improved methods for evaluating completeness. *PLoS Med*, vol. 7, No. 4, pp. e1000262. doi: 10.1371/journal.pmed.1000262.

Myers, R.J. (1976). *The dual record system: An overview of experience in five countries*. Laboratories for Population Statistics, Scientific Report Series, No. 26, North Carolina, USA, University of North Carolina, Chapel Hill.
Ngo, A.D., and others (2010). Mortality patterns in Vietnam, 2006: Findings from a national verbal autopsy survey. *BMC Res Notes*, vol. 3, No. 1, pp. 78. doi: 10.1186/1756-0500-3-78.
Nour, E.S. (1982). On the estimation of the total number of Vital Events from Dual Record Systems. *Journal of the Royal Statistical Society*, vol. 145, No. 1, pp. 106-116.
Ozdemir, R., and others (2015). Reliable mortality statistics for Turkey: Are we there yet? *BMC Public Health*, vol. 15, No. 1, pp. 545. doi: 10.1186/s12889-015-1904-1.
Padmanabha, P. (1981). Use of Sample Registration Systems in studying levels, trends and differentials in mortality: The experience of India. Data bases for mortality measurement. Proceedings of the United Nations / World Health Organization Working Group on Data Bases for Measurement of Levels, Trends, and Differentials in Mortality. Bangkok, United Nations publication, Sales No. E.83/XIII.3.
Pratiwi, E.D., and S. Kosen (2013). Development of an Indonesian Sample Registration System: A longitudinal study. *Lancet*, vol. 381, No. Special, pp. 118.
Preston, S.H. (1981). Use of direct and indirect techniques for estimating the completeness of registration systems. Data bases for mortality measurement. Proceedings of the United Nations / World Health Organization Working Group on Data Bases for Measurement of Levels, Trends, and Differentials in Mortality. Bangkok, United Nations publication, Sales No. E.83/XIII.3.
Raj, D. (1977). On estimating the number of events in demographic surveys. *Journal of the American Statistical Association*, vol. 72, No. 358, pp. 377-81.
Rao, C., and others (2011). Tuberculosis mortality differentials in Indonesia during 2007-2008: Evidence for health policy and monitoring. *International Journal of Tuberculosis and Lung Disease*, vol. 15, No. 12, pp. 1608-1613. doi: 10.5588/ijtld.11.0018.
Rao, C., and others (2005). Evaluating national cause-of-death statistics: Principles and application to the case of China. *Bulletin of the World Health Organization*, vol. 83, No. 8, pp. 618-625. doi: /S0042-96862005000800016.
Rao, C., and others (2010a). Compiling mortality statistics from civil registration systems in Viet Nam: The long road ahead. *Bulletin of the World Health Organization*, vol. 88, No. 1, pp. 58-65. doi: 10.2471/BLT.08.061630.
Rao, C., and others (2010b). Mortality in Central Java: Results from the Indonesian mortality registration system strengthening project. *BMC Research Notes*, vol. 3, No. 1, pp. 325. doi: 10.1186/1756-0500-3-325.
Sankoh, O., and P. Byass (2012). The INDEPTH Network: Filling vital gaps in global epidemiology. *International Journal of Epidemiology*, vol. 41, No. 3, pp. 579-88. doi: 10.1093/ije/dys081.
Scott, C. (1974). The dual record (PGE) system for vital rate measurement, some suggestions for further development. International Population Conference. Liege, International Union for the Scientific Study of Population.
Seltzer, W. (1969). Some results from Asian population growth studies. *Population Studies*, vol. 23, No. 3, pp. 395-406.
Seltzer, W., and A. Adlakha (1974). *On the effect of errors in the application of the Chandrasekar-Deming Technique*. Laboratories for Population Statistics, Reprint Series, No. 14, North Carolina, USA, University of North Carolina, Chapel Hill.
Szwarcwald, C.L. (2008). Strategies for improving the monitoring of vital events in Brazil. *International Journal of Epidemiology*, vol. 37, No. 4, pp. 738-744. doi: 10.1093/ije/dyn130.
Szwarcwald, C.L., and others (2014). Correction of vital statistics based on a proactive search of deaths and live births: Evidence from a study of the North and Northeast regions of Brazil. *Population Health Metrics*, vol. 12, No. 1:16, pp. 1-10.
The DHS Program (2017). Demography and Health Surveys. United States Agency for International Development. Available from http://dhsprogram.com/. Accessed Jan. 2017.
Tromp, M., and others (2011). Results from simulated data sets: Probabilistic record linkage outperforms deterministic record linkage. *Journal of Clinical Epidemiology*, vol. 64, No. 5, pp. 565-72. doi: 10.1016/j.jclinepi.2010.05.008.

United Nations (1955). *Handbook of Vital Statistics*. Studies in Methods, No. 7, ST/STAT/SER.F.7, New York, NY, Department of Economic and Social Affairs, Statistical Office of the United Nations, Sales No. 55. XVII.1.

__________ (2014). *Principles and Recommendations for Population and Housing Censuses, Revision 3*. Statistical Papers, No. 67, ST/ESA/STAT/SER.M/67/Rev.3, New York, NY, Department of Economic and Social Affairs, Statistics Division.

Vapattanawong, P., and P. Prasartkul (2011). Under-registration of deaths in Thailand in 2005-2006: Results of cross-matching data from two sources. *Bulletin of the World Health Organization*, vol. 89, No. 11, pp. 806-812. doi: 10.2471/BLT.10.083931.

Vital Statistics (Sample Registration System) Division (2007). Sample Registration System. Vital Statistics Division, Office of the Registrar General of India. Available from http://www.censusindia.gov.in/Vital_Statistics/SRS/Sample_Registration_System.aspx. Accessed 20 June, 2007.

Whitford, D.C., and J.P. Banda (2001). Post-enumeration surveys (PES): Are they worth it? Symposium on Global Review of 2000 Round of Population and Housing Censuses: Mid-Decade Assessment and Future Prospects. New York, 7-10 August 2001, Department of Economic and Social Affairs, United Nations Secretariat.

World Health Organization (2011). Annex Table 4: Data sources and Methods for estimation of mortality in 2008 by cause, age and sex. In *Causes of death 2008: data sources and methods*, pp. 17-23. Geneva: World Health Organization.

__________ (2014). *WHO Methods and data sources for country level causes of death, 2000 - 2012*. Global Health Estimates Technical Paper, No. WHO/HIS/HIS/GHE/2014/7, Geneva, World Health Organization.

Yang, G., and others (2005). Mortality registration and surveillance in China: History, current situation and challenges. *Population Health Metrics*, vol. 3, No. 1:3, pp. 1-9. doi: 10.1186/1478-7954-3-3.

Yayla, Z., and A. Cavlin (2017). An evaluation of Turkey's new death registry system. International Population Conference. Cape Town, International Union for the Scientific Study of Population.

ANNEX TABLES

ANNEX TABLE 1. CHARACTERISTICS OF DUAL-RECORD SYSTEMS INVOLVING SPECIAL REGISTRATION AND SPECIAL PERIODIC SURVEYS

Country	*Year*	*Primary source* *Special registration death records*	*Secondary source* *Survey death records*	*Reference population*	*Completeness assessment method*	*Completeness results - Registration*	*Completeness results - Survey*	*Notes*	*References*
Pakistan	1962-65	Continuous recording based on combination of household visits and consultation with informants	Full annual household enumeration using de facto household. Reference period 12 months	12 clusters (10 rural 2 urban) in each of East and West Pakistan, each cluster consisting of 5000 people, total population 120,000	Chandra-Deming de facto method	77-84 per cent	56-61 per cent		(Chanlett and Fichet, 1976; Marks and others, 1974c; Seltzer, 1969)
Egypt	1965-66	Continuous recording of vital events over 18 months by recorders responsible for 2 to 3 villages.	3 surveys conducted, 6 months apart	Stratified sampling of villages of 8 provinces of lower Egypt according to village population size. Total sample 100,000 persons or 1 per cent of population of lower Egypt	Capture-mark-recapture with Chandrasekaran adjustment method	62 to 70 per cent across stratum	79 to 88 per cent across stratum	Vital events identified by a combination of household visits and non-formal key informants	(Krotki, 1978)
Liberia	1969-71	Local registrars made monthly household visits	Surveys conducted in 6 monthly household visits	70,000 persons in 200 clusters (100 each urban and rural), or 5 per cent of national population	Capture-mark-recapture with Chandrasekaran adjustment method	66 per cent	65 per cent		(Marks and others, 1974c)
Malawi	1970-72	Local registrars made monthly household visits	Surveys conducted in 6 monthly household visits	30,000-person sample, (half urban) of total national population of 4.7 million	Capture-mark-recapture with Chandrasekaran adjustment method			Sample half urban while population is only 5 per cent urban	(Krotki, 1978)

Country	*Year*	*Primary source* *Special registration death records*	*Secondary source* *Survey death records*	*Reference population*	*Completeness assessment method*	*Completeness results - Registration*	*Completeness results - Survey*	*Notes*	*References*
Philippines	1971-73	Study Area A: bimonthly house to house visits with community contacts Study Area B: Same as above but types of community contacts were modified. Monetary incentives used for event identification	Varied between semi-annual and annual Reference periods of 6 months and 12 months with 18 months for one survey Surveys used to update household lists from previous round	Study site A: 2 small study areas with population 40,000. Study Area B: Covers most of Misamis Oriental province. 13 clusters totalling 40,000 population, 8.5 per cent of study area population. Study area population of 473,000,1.3 per cent of national population	Capture-mark-recapture with Chandrasekaran adjustment method	73 to 88 per cent in urban and 89 to 100 per cent in rural areas	49 to 91 per cent in urban and 79 to 97 per cent in urban areas	1. Respondent fatigue in survey collection led to the need for a supplementary sample being drawn	(Abernathy and Booth, 1977; Adlakha and others, 1977; Chanlett and Fichet, 1976; Myers, 1976)
Columbia	1972-73	In half of each of the clusters in each study zone data collected in house to house visits, in the other half relied on key community informants followed up by household visits	Varied between quarterly, semi-annual and annual surveys. Reference periods varied between 3, 6, and 12 months. Information collected independently in each survey round	40 clusters from two study zones (one coastal, one in interior mountainous region) totalling 70,000 population, around 4.1 per cent of study area population (1.7 million) which is 7 per cent of national population	Capture-mark-recapture with Chandrasekaran adjustment method	Varied from 53 to 83 per cent in urban and 44 to 81 per cent in rural areas	Varied from 42 to 55 per cent in urban and 38 to 59 per cent in rural areas	1. Respondent fatigue reported in registration system	(Abernathy and Booth, 1977; Adlakha and others, 1977; Chanlett and Fichet, 1976; Myers, 1976)

Country	*Year*	*Primary source* *Special registration death records*	*Secondary source* *Survey death records*	*Reference population*	*Completeness assessment method*	*Completeness results - Registration*	*Completeness results - Survey*	*Notes*	*References*
Morocco	1972-73	In urban areas, monthly household visits were carried out, in rural areas mix of household visits supplemented by community contacts	Semi-annual surveys with 12-month reference periods. Information collected independently in each survey round	56 clusters totalling 84,000 population representing 2 per cent of study area population. Study area population (4.3 million) is 28 per cent of national population	Capture-mark-recapture with Chandrasekaran adjustment method	66 per cent in urban and 71 per cent in rural areas	70 per cent in urban and 62 per cent in rural areas	Co-existence in sample areas of census and civil registration operations risks confusion and lack of independence of results due to frequent surveys	(Abernathy and Booth, 1977; Adlakha and others, 1977; Chanlett and Fichet, 1976; Fellegi, 1974; Myers, 1976)
Kenya	1973-75	Two monthly household visits plus informal community contacts. Later experimented with community contacts vs household visits.	Varied between semi-annual and annual Reference periods of 6 months and 12 months (alternatively). Information collected independently in each survey round	72 clusters with 93,000 population. Study area is one narrow strip southeast of Lake Victoria with population 3.6 million, 33 per cent of national population	Capture-mark-recapture with Chandrasekaran adjustment method	61 per cent in urban and 79 per cent in rural areas	58 per cent in urban and 55 per cent in rural areas	1. Respondent fatigue in both data collections 2. Community contacts receive an incentive payment.	(Abernathy and Booth, 1977; Chanlett and Fichet, 1976; Myers, 1976)
Turkey	1974-75	Only household visits used, 3 visits to each household per month.	Semi-annual Reference period 6 months. Surveys used to update household lists from previous round	National sample with 256 small clusters covering 90,000 people, or 0.2 per cent of the national population.	Capture-mark-recapture with Chandrasekaran adjustment method			1. Respondent fatigue reported in registration system	(Chanlett and Fichet, 1976; Myers, 1976; Seltzer, 1969)
Kenya	2000-01	Households visited twice per year and migrations, births and deaths recorded (HDSS system)	Survey of sample of households in study area Reference period 6 months	60,000 rural residents in Rarierda Division and 70,000 residents of Wagai and Yala divisions. Survey covered 5,000 households in each area	Capture-mark-recapture with Peterson maximum likelihood estimator used to estimate missed events	Sensitivity: 49-62 per cent for neonate deaths, 72-78 per cent for post neonatal child deaths	6 per cent of male and 8 per cent of all estimated female deaths recorded	Study focused on child deaths	(Eisele and others, 2003)

Country	*Year*	*Primary source* *Special registration death records*	*Secondary source* *Survey death records*	*Reference population*	*Completeness assessment method*	*Completeness results - Registration*	*Completeness results - Survey*	*Notes*	*References*
Indonesia	2007-08	Continuous recording of vital events over 2 years. Community deaths identified by families through community volunteers, hospital deaths by routine death notifications	One off independent household survey in random selection of villages	638 rural villages in 5 provinces in Sumatra, West Kalimantan, Sulawesi and Papua islands	Capture-mark-recapture with Chandrasekaran adjustment method	68 per cent			(Rao and others, 2011)

ANNEX TABLE 2. CHARACTERISTICS OF DUAL-RECORD SYSTEMS BASED ON NATIONAL SAMPLE REGISTRATION AND ROUTINE PERIODIC SURVEYS

Country	*Year*	*Title of national programme*	*Routine recording system*	*Periodic Survey*	*Survey data collection method*	*Reference population*	*Completeness assessment method*	*Completeness results*	*Remarks*	*Reference*
India	1965 onward	Sample Registration System of India	Continuous recording by specially appointed local registrar	Periodic half yearly survey by the local SRS supervisor	6 month recall of vital events recorded by independent investigator	Nationally representative sample clusters. Current sample covers 8.4 million 8700 clusters	Not measured or reported on routine basis	Indirect demographic techniques identified 90 per cent completeness in 2001	The Indian SRS operates independent of the CRVS with separate funding and human resources.	(Padmanabha, 1981)
China	1990 onwards	Chinese Disease Surveillance Point (DSP) System	Death register maintained at township hospitals in each DSP site	Periodic, 3 yearly, surveys for evaluation of completeness of death registration in only a sample of each DSP.	Selected sample of townships/ streets where key informants create lists of deaths, followed by household visit	64 urban and 97 rural surveillance sites in all 31 provinces currently covering 98 million people, or 7 per cent of the total Chinese population.	Capture-mark-recapture with Chandrasekaran adjustment method	Completeness in 2011 varied from 83 to 92 per cent across geographic regions	Surveys contact only households on lists of events from institutions and key informants. 3 yearly surveys create risk of recall bias	(Guo and others, 2015; Yang and others, 2005)
Bangladesh	1980 onward	Sample Vital Registration and Statistics System	Monthly household visit by local registrar	3 monthly household visit by SRS supervisor	3 month recall of vital events recorded by independent investigator for matching	Nationally representative sample of 1000 primary sampling units; population coverage of 206522 households	Chandrasekar Deming adjustment method	Completeness not reported		(Bangladesh Bureau of Statistics, 2011)
Indonesia	2014 onward	Indonesian Sample Registration System	Lists of deaths compiled every quarter by key informants	Special household surveys in sample of villages	Household surveys recording deaths in the previous 12 months	Nationally representative sample of 128 sub-districts covering a population of 8.4 million	Capture-mark-recapture with Chandrasekar adjustment method	Survey conducted in 2014 could not be utilised due to problems with survey data quality	Design and implementation of survey in 2017 being planned with focus on sample size and distribution	(Pratiwi and Kosen, 2013)

ANNEX TABLE 3: CHARACTERISTICS OF DUAL-RECORD SYSTEMS BASED ON ROUTINE CIVIL REGISTRATION AND ALTERNATE DATA SOURCES

Country	*Year*	*CRVS data source*	*Survey description*	*Survey data collection method*	*Reference population*	*Variables matched*	*Completeness assessment method*	*Complete-ness results*	*Remarks*	*References*
Thailand	2005	Ministry of Interior death records	Periodic intercensal sample survey of population change covering 82,000 households	Quarterly household visit over 12 months, to record occurrence of deaths	CRVS - national coverage; survey - Nationally representative cluster sample	13-digit national personal identification number	Capture-mark-recapture with Chandrasekar Deming adjust-ment method	CRVS completeness 91.3 per cent	• 11.3 per cent records not matched due to missing PID • Study identified reasons for missed events in both sources • Study underway periodically since 1960s with some variations in study design	(Vapattanawon g and Prasartkul, 2011)
Oman	2010	Ministry of Health Death registration database	Decennial census of 2010	Recall of household deaths in 12-month period preceding census	National coverage of both sources	Age, sex of deceased; date of death; address variables	Capture-mark-recapture with Chandrasekaran adjustment method	CRVS completeness 95 per cent	Name of deceased not available in census records; hence matching undertaken with name of head of household	(Al Muzahmi, 2015)
South Africa	2006-2009	Civil registration	HDSS surveillance	Annual updates of vital events in study area	Agincourt HDSS, located in the Bushbuckridge district of Mpumalanga province in rural north-east South Africa, population 87,000	national identity number (a unique 13-digit number), surname, sex, day/month/year of birth, day /month/year of death, village name and institution/place of death	Record matching and reconciliation	CR 85 per cent complete, HDSS site registered 61 per cent of deaths recorded in CR	Capture-recapture method not applicable due to geographical out-of-scope events due to different boundaries in CR data	(Joubert and others, 2014)

Country	*Year*	*CRVS data source*	*Survey description*	*Survey data collection method*	*Reference population*	*Variables matched*	*Completeness assessment method*	*Complete-ness results*	*Remarks*	*References*
Korea	1963-1965	Civil registration	Household survey	One-off survey with reference period up to 2 years	100,000 people		Record matching	CRVS – 43 per cent births, 33 per cent deaths. Survey – 28 per cent births, 23 per cent deaths		(Seltzer, 1969)

ANNEX TABLE 4: MULTIPLE DATA SOURCES WITH OVERLAPPING RECALL PERIODS

Country	*Year*	*First data source*	*Second data source*	*Other data sources*	*Reference population*	*Matching variables*	*Completeness assessment method*	*Completeness estimate*	*Remarks*	*References*
Kiribati	2000-2009	Civil registration	Health Informatio n System		National population of Kiribati, approximately 103,000 people	Name, island of death, date of death, gender	Death record reconciliation	Civil registration 59 per cent of unique deaths and HIS recorded 64 per cent of total deaths		(Carter and others, 2016)
Tonga	2001-2009	Civil registry	Health Informatio n System	Reproductiv e health surveillance system	National population of Tonga, approximately 103,000 people	Name, age at death. Date of death, same place of death, same place of residence	Capture-mark-recapture with Chandrasekaran adjustment method	53 per cent for Civil registration data, 66 per cent for HIS data, and 40 per cent for Reproductive Health Surveillance System (2005-2009)	HIS and Reproductive Health Surveillance system data were co-dependent	(Hufanga and others, 2012)
Philippines	2006-7	Civil registration	Church parish records	health centre records	Residents of 5 municipalities and one city in Bohol	Full name age at death, date of death, municipality of residence and sex	Capture-recapture	Civil registration and health centre records 77 per cent complete. Infant deaths had lower completeness being under reported by civil registration data by 62 per cent	Civil registration and health centre records were co-dependent so capture recapture analysis was performed with both two source and three source models	(Carter and others, 2011b)

Country	*Year*	*First data source*	*Second data source*	*Other data sources*	*Reference population*	*Matching variables*	*Completeness assessment method*	*Completeness estimate*	*Remarks*	*References*
Indonesia	2006/7	Continuous recording by community volunteers, and hospital deaths notifications	One off household survey in sample of villages	Civil registration	Survey sample 10,467 rural households in 12 villages and 14,607 urban households in 13 city wards	the name(s), age at death, gender, address, and date of death	Data reconciliation	73 per cent rural, 55 per cent urban	The survey was conducted in conjunction with a electoral household register update	(Rao and others, 2010b)
Vietnam	2008-9	Justice department records	Commune health centre records	Commune Population and Family Planning Committee records	Stratified multi-stage cluster design sample. Total 192 communes, 668,142 households, 2.6 million people (3 per cent of national population)	name, sex, age, month of death	Capture-mark-recapture with Chandrasekaran adjustment method	Overall completeness was 81.3 per cent but in infants was very low	Commune health centre and Population and Family Planning Committee records dependent, matching combined these two sources and matching with Justice Department records	(Hoa and others, 2012)
Brazil	2008-2009	Civil registration (Mortality Information System)	Registry offices	Hospitals, primary health care units, institutes of forensic medicine, ambulance service, burial sites, funeral homes and midwife groups	Probabilistic sample of 133 municipalities of 17 states of Amazonia and Northeast Brazil	Not reported	Correction factors calculated and applied to civil registration mortality estimates	VR completeness for infant deaths 85 per cent overall (range—74-89 per cent) For all deaths 94.4 per cent (85 per cent in North and 88 per cent in Northeast Brazil)	Proactive search technique involves actively seeking death reports from any available source and then later matching Although CRVS completeness was high, the study found problems with data flow.	(Figueiroa and others, 2013; Frias and others, 2010; Szwarcwald, 2008; Szwarcwald and others, 2014)