

Towards the next generation of record-linkage studies to advance data quality assessment of civil registration systems in low- and middle-income countries

Romesh Silva[†] and Carla AbouZahr[‡]

IUSSP Panel on Innovations in Strengthening Civil Registration & Vital Statistics Systems

February 10, 2016

Background

Over the last few years, there has been growing momentum to improve civil registration and vital statistics (CRVS) systems in low- and middle-income countries. This is most visible in high-level regional strategies in the Asia Pacific and Africa regions and in the formulation of Goal 16 of the Sustainable Development Goals that recognizes a direct link between CRVS systems, legal identity and sustainable development. Further, international donors have begun mobilizing resources to support improved CRVS systems through the World Bank's Global Financing Facility and associated collaborative funding mechanisms (such as Bloomberg Philanthropies' Data for Health Initiative). However, much work remains to be done in developing a stronger technical foundation to underpin the monitoring and evaluation of CRVS systems improvement. This is particularly the case in relation to the completeness and quality of birth and death statistics, which are key indicators of CRVS system performance.

Record linkage is an important tool for assessing completeness and quality of birth and death records, as has been recognized in the *United Nations Principles and Recommendations on Vital Statistics (Revision 3)*. Promising record linkage studies have been undertaken in high- and middle-income countries to monitor and evaluate the completeness and quality of CRVS data. These initiatives, which have been undertaken both at the national and sub-national levels, have exploited powerful and increasingly accessible machine learning techniques to efficiently and reliably link individual records between different data systems. As CRVS systems improve, countries will need to move beyond coarse methods of evaluation (such as indirect methods) which are subject to notable measurement uncertainty and apply more precise methods (such as record linkage) that can be applied nationally and at sub-national levels and that enable the identification of specific subpopulations that are systematically under-registered.¹ Such efforts will be critical to ensure that the 'data revolution' linked to the 2030 development agenda truly 'leaves no one behind'.

Motivation and Broader Context

The motivation behind this initiative is to learn from current experiences with record linkage in order to develop improved operationalize technical guidance relevant to low- and middle-income countries (LMICs). The guidance should build upon the UN Principles and Recommendations on Vital Statistics (Rev.3) while maximizing the potential for record linkage offered by new technologies and the rapid update of unique individual identification systems around the world. By sharing knowledge and lessons learnt from recent record linkage initiatives in a diverse array of settings, it will be possible to inform practical thinking on how these methods can be integrated into current CRVS improvement and assessment plans in specific LMICs.

[†] Member, IUSSP Scientific Panel on Innovations for Strengthening Civil Registration and Vital Statistics (CRVS) Systems; Statistician, Demographic and Social Statistics Section, Statistics Division, Economic and Social Commission for Western Asia, United Nations. e: silva22@un.org

[‡] Member, IUSSP Scientific Panel on Innovations for Strengthening Civil Registration and Vital Statistics (CRVS) Systems; Independent Consultant.

¹ Murray et al. (2010), when evaluating several variants of death distribution methods, found that the most commonly-used indirect methods of evaluating the completeness of death registration systems were subject to large uncertainty intervals – up to +/-25%.

This initiative is designed to complement the broader CRVS improvement initiatives currently being undertaken by the World Bank, United Nations, regional institutions and foundations, such as the Bloomberg Data for Health initiative. Recently there has been important progress in mobilizing political commitment of governments in LMICs (most notably as part of the high-level ministerial meetings of the APAI-CRVS in Africa and UNESCAP in the Asia Pacific region) and identifying global financing resources (via bilateral donors, the World Bank, and regional development banks). However, more work is needed to develop the methodological underpinnings for evidence-based and context-specific assessment methods of CRVS improvement initiatives.

Experiences in record linkage for CRVS improvement

Record linkage techniques have been effectively employed at subnational levels in LMICs, including the Philippines, Thailand, and Egypt, to identify the magnitude of under-registration of deaths with more precision than is possible with indirect methods or survey estimates.² Further, recent record-linkage between the civil registration (CR) system and a health and demographic surveillance site (HDSS) in rural South Africa indicated that the deaths of vulnerable people, including poorer persons, children under 5 years, and non-nationals, are less likely to be registered by the CR system.³ A study linking mortality data from the South African CR system and the same HDSS, identified the extent and diversity of misattribution of HIV deaths in CR data in a rural setting of South Africa.⁴ Thus, there is a growing body of evidence that record-linkage studies to assess CR systems in LMICs are both feasible and relevant to CRVS improvement efforts.

Recent record-linkage work undertaken by official statistical agencies in Canada, New Zealand and the United States demonstrate the potential for more detailed evaluations of population data systems, including civil registration systems, and the ability of record-linkage techniques to be part of an effective evidence-based feedback loop that guides systems improvement initiatives. In Canada, record-linkage between the civil registration system and perinatal health surveillance system has identified notable variation in death registration data quality across provinces and facilitated in-depth review of registration procedures in different provinces.⁵ Methodological reviews of census coverage evaluation in the United States have recommended that record linkage analyses explicitly measure the 4 main components of coverage error: (1) census omissions, (2) census duplications, (3) erroneous census enumerations, and (4) census enumerations in the wrong location.⁶ Record linkage work in the United States is now less focused on providing aggregate estimates of under enumeration, and more calibrated to identifying deficient documentation processes and proposing better procedural alternatives. A key question for CRVS practitioners is how new advances in record-linkage pioneered in high-income settings might be adapted to LMIC settings.

The initiative that we propose here seeks to synthesize these recent country experiences from a broad array of high, medium and low-income settings into actionable lessons that can inform the next generation of record-linkage assessments in LMICs, as part of ongoing global CRVS scale-up efforts.

Objectives and Strategy

As complementary data systems (including CRVS systems, ID management systems, household survey systems, HDSS and SAVVY systems, etc.) are progressively strengthened in LMICs, there are new opportunities to assess the completeness and quality of these systems through the use of record linkage studies. The objectives of this proposed initiative are to:

- document and share lessons learned from recent record linkage studies;
- strategize about the design and implementation of such record linkage studies in low-income and middle-income countries;
- facilitate a new record linkage projects in 3-4 countries that have well-developed CRVS improvement plans.

This initiative seeks to fill a gap in current technical assistance efforts by convening a group of methodological experts to share recent experiences on CRVS record linkage studies and by exploring how recent technical advances can be customized to LMIC settings.

² Carter et al., 2011; Vappattanawong and Prasartkul, 2006; Vappattanawong and Prasartkul, 2011; El-shalakani, 1985

³ Kabadula et al., 2014

⁴ Joubert et al., 2014,

⁵ Health Canada, 2003

⁶ National Research Council, 2009

Our strategy is to convene a group of statisticians/demographers/epidemiologists from selected LMICs (possible candidate countries include Mozambique, Oman, Egypt, Philippines, Palestine, Morocco, Thailand), technical experts from international agencies (including the World Bank, UN global and regional agencies, and WHO), related research partners (e.g. INDEPTH Network of health and demographic surveillance sites, US National Center for Health Statistics, South African Medical Research Council), and academic experts on birth/death record linkage studies. We propose to hold this meeting immediately following the annual meetings of the Population Association of America (PAA), which will be held from March 31 to April 2, 2016. We would like to exploit the fact that PAA 2016 will be in Washington D.C. and therefore bringing US experts (from the Census Bureau, the National Center for Health Statistics, NORC, Social Security Administration) will be relatively inexpensive and easy (especially if those experts are participating in the PAA-2016). Subject to their availability and the availability of funding resources, we would also like to invite selected experts from Statistics Canada, the Australian Bureau of Statistics, Australian Institute of Health and Welfare, Statistics New Zealand, and Statistics Canada.

Proposed Format

The tentative format of the proposed meeting is:

1. Bring together lessons learnt from recent experiences of high-income countries in carrying out advanced record linkage studies (particularly those that have used probabilistic-based matching and machine learning techniques) and work carried in South Africa, Vietnam, Iran, and at various HDSS sites, and the implications of these experiences for strengthening CRVS system completeness and quality;
2. Share experiences of country representatives in terms of priority issues and available data for a new set of record linkage studies (e.g. Mozambique, Egypt, Philippines, Palestine) and their potential implementation as part of CRVS strengthening efforts;
3. Facilitate brainstorming on design and implementation considerations for future birth/death record linkage studies in selected LMICs and their integration into ongoing CRVS improvement plans.

The enclosed draft agenda provides more details about the structure and format of the meeting, the particular themes to be addressed and the proposed list of participants.

Expected Outcomes and Outputs

The expected outcomes from this initiative are:

- (1) A shared knowledge base of experiences of record linkage studies and their potential for implementation in other settings;
- (2) The formulation of a set of 3-4 draft country projects (possible candidates include Mozambique, Egypt, Philippines, Palestine) drawn from technical discussions at this meeting (that could then be integrated into country-level CRVS improvement plans); and
- (3) Technical guidance and potential collaborative opportunities in the area of death record linkage studies.

The expected outputs from this initial meeting are

- a) summary findings and considerations to be taken into account when introducing record linkage in countries (These findings and considerations will draw direct links between the UN Principles and Recommendations on Vital Statistics and lessons learned from recent record-linkage studies. These summary findings will be written up as a summary paper and included in a journal supplement along with selected papers presented at this PAA side meeting. Possible journals for a supplement include PLoS ONE, the Statistical Journal of the IAOS, Journal of Official Statistics, and Demographic Research.);
- b) a draft set of detailed project proposals that outline pilot studies to advance the customized use of record-linkage methods in improving and sustaining CRVS improvements in selected LMICs. Country representatives and partners from global headquarters and regional offices of international agencies will jointly develop these proposals in consultation with CRVS donor partners.

References

Carter KL, Williams G, Tallo V, Sanvictores D, Madera H, Riley I. Capture-recapture analysis of all-cause mortality data in Bohol, Philippines. *Popul Health Metr*. 2011;9:9.

El-shalakani M. Estimating the completeness of births and deaths registration in Egypt through dual record systems. *Genus*. 1985;41:119–32.

Health Canada. Canadian Perinatal Health Report, 2003. Ottawa: Minister of Public Works and Government Services Canada, 2003. Available at <http://www.hc-sc.gc.ca/pphb-dgspsp/publicat/cph-rspsc03/>

Kabudula et al. Evaluation of record linkage of mortality data between a health and demographic surveillance system and national civil registration system in South Africa. *Popul Health Metr* 2014 12:23.

Murray CJL, Rajaratnam JK, Marcus J, Laakso T, Lopez AD (2010) What Can We Conclude from Death Registration? Improved Methods for Evaluating Completeness. *PLoS Med* 7(4): e1000262.

National Research Council (2009). Coverage Measurement in the 2010 Census. Panel on Correlation Bias and Coverage Measurement in the 2010 Decennial Census, Robert M. Bell and Michael L. Cohen (Eds.). Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Prasartkul P, Vapattanawong P. The completeness of death registration in Thailand: Evidence from demographic surveillance system of the Kanchanaburi Project. *World Health Popul*. 2006;8:43–51.

Prasartkul P, Vapattanawong P. Under-registration of deaths in Thailand in 2005–2006: results of cross-matching data from two sources. *Bull World Health Organ*. 2011 Nov 1; 89(11): 806–812.