
Methodological musings

Using routine health information systems for well-designed health evaluations in low- and middle-income countries

Bradley H. Wagenaar,^{1,2*} Kenneth Sherr,^{3,2} Quinhas Fernandes⁴ and Alexander C. Wagenaar⁵

¹Department of Epidemiology, School of Public Health, University of Washington, Seattle, WA, USA, ²Health Alliance International, Seattle, WA, USA, ³Department of Global Health, School of Public Health, University of Washington, Seattle, WA, USA, ⁴Department of Monitoring and Evaluation, National Directorate of Planning and Cooperation, Ministry of Health, Maputo, Mozambique and ⁵Department of Health Outcomes and Policy, College of Medicine, University of Florida, Gainesville, FL, USA

*Corresponding author. Department of Epidemiology, School of Public Health, University of Washington, 1959 NE Pacific Street, Seattle, WA 98195, USA. E-mail: wagenaarb@gmail.com

Accepted on 5 March 2015

Abstract

Routine health information systems (RHISs) are in place in nearly every country and provide routinely collected full-coverage records on all levels of health system service delivery. However, these rich sources of data are regularly overlooked for evaluating causal effects of health programmes due to concerns regarding completeness, timeliness, representativeness and accuracy. Using Mozambique's national RHIS (*Módulo Básico*) as an illustrative example, we urge renewed attention to the use of RHIS data for health evaluations. Interventions to improve data quality exist and have been tested in low- and middle-income countries (LMICs). Intrinsic features of RHIS data (numerous repeated observations over extended periods of time, full coverage of health facilities, and numerous real-time indicators of service coverage and utilization) provide for very robust quasi-experimental designs, such as controlled interrupted time-series (cITS), which are not possible with intermittent community sample surveys. In addition, cITS analyses are well suited for continuously evolving development contexts in LMICs by: (1) allowing for measurement and controlling for trends and other patterns before, during and after intervention implementation; (2) facilitating the use of numerous simultaneous control groups and non-equivalent dependent variables at multiple nested levels to increase validity and strength of causal inference; and (3) allowing the integration of continuous 'effective dose received' implementation measures. With expanded use of RHIS data for the evaluation of health programmes, investments in data systems, health worker interest in and utilization of RHIS data, as well as data quality will further increase over time. Because RHIS data are ministry-owned and operated, relying upon these data will contribute to sustainable national capacity over time.

Key words: Epidemiology, evaluation, health information systems, health policy, health systems research, impact, implementation, methods, research methods, survey methods

Key Messages

- Routine health information systems (RHISs) are regularly overlooked for evaluating causal effects of health programmes due to concerns regarding completeness, timeliness, representativeness and accuracy.
- Intrinsic elements of RHIS data (numerous repeated observations over extended periods of time, full coverage of health facilities and numerous real-time indicators of service coverage and utilization) enable the use of research designs that are superior to those relying on intermittent community sample surveys in their causal attribution of effects to an intervention or policy change.
- Given limited global resources, the high costs of intermittent community surveys have to be weighed against improving RHIS for evaluations, detailed measurement of implementation strength, primary health care organization, operations research, continuous quality improvement and resource allocation.

Introduction

Routine health information systems (RHISs) are present in nearly all low- and middle-income countries (LMICs) and are recognized as the backbone of facility-level microplanning and higher-level (district, provincial, regional, national) decision-making, resource allocation and strategy development (AbouZahr and Boerma 2005; World Health Organization 2007; Mutale *et al.* 2013). However, using RHIS data for micro and macro-level evaluations of programme interventions or policy changes has been questioned, largely due to concerns with data quality (Rowe *et al.* 2009, 2007; Hazel *et al.* 2013). The ‘gold standard’ for evaluations continues to be intermittent community surveys—such as the demographic and health survey (DHS) or the multiple indicator cluster survey (MICS)—despite their high cost, low frequency, inability to provide district-level estimates (Fernandes *et al.* 2014) and reliance on external organizations without building sufficient in-country capacity for independent continuity (MEASURE DHS; Short Fabic *et al.* 2013; United Nations Children’s Fund (UNICEF)).

Our objective is to urge renewed interest in using RHIS data for country-led programme and policy evaluations. Questions about RHIS data quality are not unfounded; however, there are rapid, effective and inexpensive methods for improving RHIS data in LMICs, and data quality concerns may decrease over time. Less attention has been given to the advantages of RHIS data for robust programme evaluation, which are optimal for quasi-experimental research designs with strong validity protections and associated causal inference. Quasi-experimental designs are of particular advantage in LMIC settings, where alternative randomized controlled trials are often expensive, slow, not feasible, lacking in external validity and/or ethically questionable (Biglan *et al.* 2000; Victora *et al.* 2004; Rothwell 2005). Using Mozambique’s national RHIS (*Módulo Básico*) for illustration purposes, we describe methods and explore potential benefits of using RHIS data for micro or macro-level health evaluations.

RHIS data completeness, timeliness, representativeness and accuracy

Improving health evaluations in LMICs is complicated by defining ‘treatment’ and ‘comparison’ areas amid myriad global health initiatives and continuous programme and policy changes. Recommended enhanced analyses focus on districts as the unit of analysis, and call for continuous data availability and integration from numerous data sources (Victora *et al.* 2011). Existing national RHIS records already contain the essential design elements required to populate integrated continuous monitoring and evaluation databanks, but RHIS data are often ignored due to concerns of their ‘completeness, timeliness, representativeness and accuracy’ (Rowe 2009), and are substituted

with intermittent surveys primarily organized and implemented by international organizations.

Case studies of poor RHIS data quality are widespread and cut across multiple programmes including immunization (Ronveau *et al.* 2005; Lim *et al.* 2008; Bosch-Capblanch *et al.* 2009), Human immunodeficiency virus (HIV) (Mate *et al.* 2009) and malaria (Rowe *et al.* 2009). Others have raised questions about the quality of under-5 mortality data compared with ‘gold standard’ community surveys (Amouzou *et al.* 2013). In contrast, others have successfully used RHIS data to evaluate interventions focused on these same programmes: malaria (Chanda *et al.* 2012), immunization (Verguet *et al.* 2013), the Global Fund activities around HIV/ Acquired immunodeficiency syndrome (AIDS), Tuberculosis and Malaria, among others (The Global Fund to Fight AIDS, Tuberculosis, and Malaria 2009). Although RHIS data quality concerns are substantiated, we argue for a more nuanced and balanced view of the interpretation of high-quality data in the context of programme and policy evaluations.

First, RHIS data quality concerns are not equal across time, indicators, health programmes, or levels of data input or aggregation (health facility, district, provincial, national). In many LMICs, RHIS data quality is increasing rapidly through improved data management, use of electronic data systems for input and aggregation, and the simple act of measuring data quality through quality assessments (Mphatswe *et al.* 2012; Mutale *et al.* 2013; Wagenaar *et al.* 2015). Fostering RHIS data use has been linked to increased data quality over time, suggesting that the continued dismissal of RHIS data for widespread application may hinder sustained data quality improvements (Cibulskis and Hiawalyer 2002; Braa *et al.* 2012).

In Mozambique, we have described high availability and reliability of five RHIS indicators (antenatal care, institutional births, DPT3, HIV testing and outpatient consults), including between 97 and 100% availability and 80% reliability overall (with 70% of all data disagreements originating from two urban health facilities; when excluded, overall reliability increased to 92%) (Gimbel *et al.* 2011). Additionally, electronic patient tracking systems for vertical HIV programmes were implemented in concert with the Mozambique RHIS, with completeness and reliability having been described to be over 90% (Lambdin *et al.* 2012).

This high reliability contrasts with other elements of the RHIS, such as data on malaria, facility-level resources (number of beds, staffing, equipment/supplies) or mental health services, which are notoriously poor in Mozambique (Chilundo *et al.* 2004). These indicators lag behind others for multiple reasons, including inconsistent use (facility-level health resources), historic under-valuation (mental health services), difficulties with diagnosis in the clinical encounter (malaria), unlinked data systems (laboratory vs clinical diagnosis for malaria) and high complexity (over 40 HIV indicators).

RHIS critiques based on data timeliness are unwarranted compared with 'gold standard' community surveys. For example, the 2011 Mozambique DHS sampled households from June to November 2011, but the final report and dataset were not available until March 2013 (Instituto Nacional de Estatística 2013). Mozambique's RHIS (*Módulo Básico*) contains health-facility data for every public clinic in the country from 1998 to present, continuously updated with less than a 1-month data lag. Since over 98% of the population in Mozambique uses public-sector health facilities, population coverage of health services (and the RHIS) is high (Sherr *et al.* 2013).

Assessing RHIS data validity in LMICs is more challenging than completeness, reliability or timeliness. Using independent observers in the clinical encounter or during exit surveys can be prone to the Hawthorne effect (McCarney *et al.* 2007). RHIS data can be compared with community surveys as the 'gold standard' at provincial or national levels, which has demonstrated high correlation in Mozambique (Gimbel *et al.* 2011). Yet, this validity assessment method may not yield reliable estimates when RHIS quality is known to be heterogeneous at the district or health facility level. In addition, the validity of community surveys is compromised by testing effects (changes in behaviours, self-reports and systems due to repeated surveys), social desirability biases, methodological discrepancies across surveys and survey teams over time, interviewer-induced error and non-validated survey questions (Fernandes *et al.* 2014). In contrast, recall and other reactive biases are not an issue given the continuous nature of RHIS data collection in the clinical encounter. Data errors and biases are limited and largely consistent over time with quality oversight through repeated use of the same data system, indicators and consistent staff. In areas with data errors, biases and/or limited sensitivity/specificity, RHIS data trends will accurately represent utilization patterns and can still be used for high-quality evaluations of intervention effects (assuming consistent errors across time) (Rowe *et al.* 2007).

In light of RHIS data quality heterogeneity, we propose that data quality assessments (DQAs) be embedded in pilot data collection activities at focal clinics, or a random sample of clinics/districts, to determine current data quality in preparation for an evaluation of a new delivery model, intervention, health strategy or policy. If data quality is unsatisfactory, simple, rapid data improvement approaches can be implemented during the preparation period, paired with follow-up DQAs throughout the intervention's time course. Indicators for an evaluation framework can be selected from those in the RHIS that achieve predetermined standards for high quality (reliable, concordant, available and valid if quantifiable), with DQAs ongoing throughout the project to account for changes in data quality over time (Mutale *et al.* 2013).

With increased use, data quality monitoring, and management and information technology interventions, RHIS data could become the new 'gold standard' for health programme evaluations. Although reliable and valid RHIS data are prerequisites for robust analyses, more attention is needed to identify research designs that maximize causal inference across different data sources. The structure inherent in RHIS data allows for robust quasi-experimental time-series designs that are not possible with intermittent community surveys, which argues for prioritizing investments in RHIS data for micro and macro-level evaluations.

Design elements favouring RHIS data for health evaluations—an example from Mozambique

Like many LMICs, RHIS data are routinely collected in clinical encounters at all government health facilities in Mozambique.

Monthly programme data are aggregated at the health facility level, sent to the district level for aggregation across facilities, and aggregated across higher-level administrative levels to the national level, resulting in a continuous time-series of repeated monthly counts of multiple health indicators. This framework of full-coverage monthly repeated observations is ideal for controlled interrupted time-series (cITS) analyses, the strongest quasi-experimental design to evaluate the effects of health interventions (Biglan *et al.* 2000; Shadish *et al.* 2002). For example, to evaluate the effects of a pilot intervention to increase institutional birth attendance by providing free newborn kits at the time of delivery implemented at Vanduzi health centre in July, 2009, a traditional evaluation may employ a controlled pre-post or difference-in-differences design (Figure 1A) using 'gold standard' DHS and MICS data (we recognize evaluations using community survey data are usually limited to assessing intervention effects at the provincial-level or higher—consider this an illustrative example). Alternatively, using RHIS data, a cITS design could be used, with associated design and analysis advantages (Figure 1B and C).

First, the cITS approach maximizes statistical power, resulting in 138 repeated observations of the number of monthly institutional births (90 pre- and 48 post-intervention), allowing for high-resolution detection of trends (history, maturation, seasonality) before, during and after intervention implementation that is not possible with two observations using community survey data (design: Figure 1B; time-series data: Figure 2). This ability to accurately control for temporal trends, multiple concurrent programmes and fluid random shocks in data series has been repeatedly advocated for in global health (Lim *et al.* 2008; Rowe 2009; Victora *et al.* 2011). As full-coverage RHIS records can be considered census data of facility-based or facility-driven outreach care and treatment, there is no need for traditional frequentist statistics as required with sample surveys. This has important on-the-ground implications as simple provincial, district and health facility-level analyses (such as graphing and visualizing means or proportions over time) can be undertaken without the need for complicated techniques to account for the sampling approach in multi-stage community surveys. More complex analyses can easily control for nested correlations (within facilities, districts, provinces and national), as well as autocorrelation across repeated measures.

Second, RHIS data can be aggregated or disaggregated to all administrative levels of the health system or country. Evaluations can be designed for health-facility, district, provincial or national levels using the same design/analytical framework allowing for intervention effects to be assessed where an a priori theory of change indicates the effect will be most visible. The ability to analyse data at all levels of aggregation is essential because the practical implications and relevance of evaluation data depends on the level of analysis, as does the sensitivity to detect an intervention effect; for example, it would likely be impossible to observe and attribute facility-level changes in institutional births as a result of the pilot intervention at Vanduzi clinic using two provincial-level community survey data points. In addition, RHIS data allow assessing if an intervention is effective at some clinics, districts, or time points, but not effective or harmful in others. This is essential information that should interface with theories of the intervention effect, and which is missed using community surveys with low time resolution or a lack of data below the provincial level. The level of data analysis is also related to statistical variance; with RHIS data, if variability is high at the health facility level, data can be aggregated to the district or provincial level to eliminate noise in estimates.

Standard Pre-Post Design with Treatment and No-Treatment Control Group (2 observations)

$$(A) \quad \begin{array}{c} O_{2008 \text{ MICS}} \quad X \quad O_{2011 \text{ DHS}} \\ \hline O_{2008 \text{ MICS}} \quad O_{2011 \text{ DHS}} \end{array}$$

Interrupted Time-Series Design with Treatment and No-Treatment Controls (138 observations)

$$(B) \quad \begin{array}{c} O_{\text{Jan '02}} \dots O_{\text{Sep '05}} \dots O_{\text{Jun '09}} \quad X \quad O_{\text{Jul '09}} \dots O_{\text{Jul '11}} \dots O_{\text{Jun '13}} \\ \hline O_{\text{Jan '02}} \dots O_{\text{Sep '05}} \dots O_{\text{Jun '09}} \quad O_{\text{Jul '09}} \dots O_{\text{Jul '11}} \dots O_{\text{Jun '13}} \end{array}$$

Interrupted Time-Series Design with a Nonequivalent Dependent Variable (138 observations)

$$(C) \quad \begin{array}{c} O_{A1} \dots O_{A45} \dots O_{A90} \quad X \quad O_{A91} \dots O_{A114} \dots O_{A138} \\ \hline O_{B1} \dots O_{B45} \dots O_{B90} \quad X \quad O_{B91} \dots O_{B114} \dots O_{B138} \end{array}$$

Figure 1. Diagrams of potential health programme or policy evaluation research designs using routine community survey data (A) vs RHIS data (B/C). Data observations over time are represented by Os; X represents a programme or policy intervention of interest; A represents an outcome of interest that the programme or policy intervention is expected to affect; B represents a non-equivalent dependent variable not expected to change as a result of the intervention but expected to respond to other potential confounding factors in a similar way as the variable of interest. Adapted from Shadish *et al.* 2002.

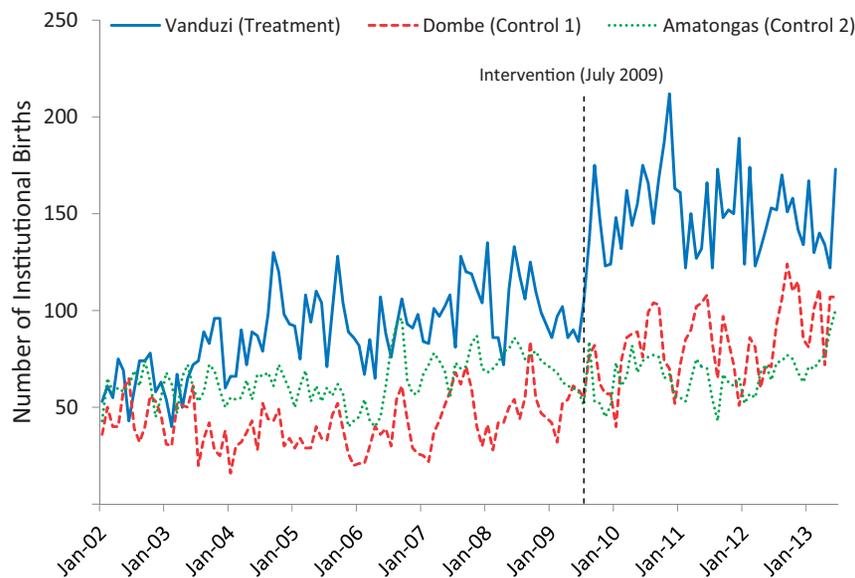


Figure 2. Time-series of number of institutional births from January 2002 to June 2013 for two control clinics and one clinic that underwent intervention to increase institutional births in July 2009. Data are from the national RHIS in Mozambique, the *Módulo Básico*.

Third, the structure of RHIS data allows for the selection of an array of no-treatment control groups, and the use of non-equivalent dependent variables (control outcomes) at any level of data aggregation. No-treatment control groups for a micro-level evaluation could match any public sector health facility in the country to treatment health facilities, whereas for macro-level evaluations can match any district or province allowing for extreme flexibility to make valid comparisons (Figure 1B). The availability of data on all possible comparison jurisdictions allows precise selection of optimal control groups to function as the research design counterfactuals, which optimally is a group where all determinants of health outcome are the

same as in the intervention group but for the programme/policy being evaluated. The time-series nature of RHIS data allows directly measuring the baseline correlations in outcomes across the intervention jurisdiction and each candidate comparison jurisdiction, and selecting the comparison with the highest correlation. If patient-level records are available, one could even construct an even more precise ‘synthetic’ comparison group using propensity scoring methods. To increase confidence in measuring the causal effects of the birth incentives intervention, we could compare changes in institutional births in Vanduzi to comparison clinics selected based on similar demographic characteristics (Figure 2).

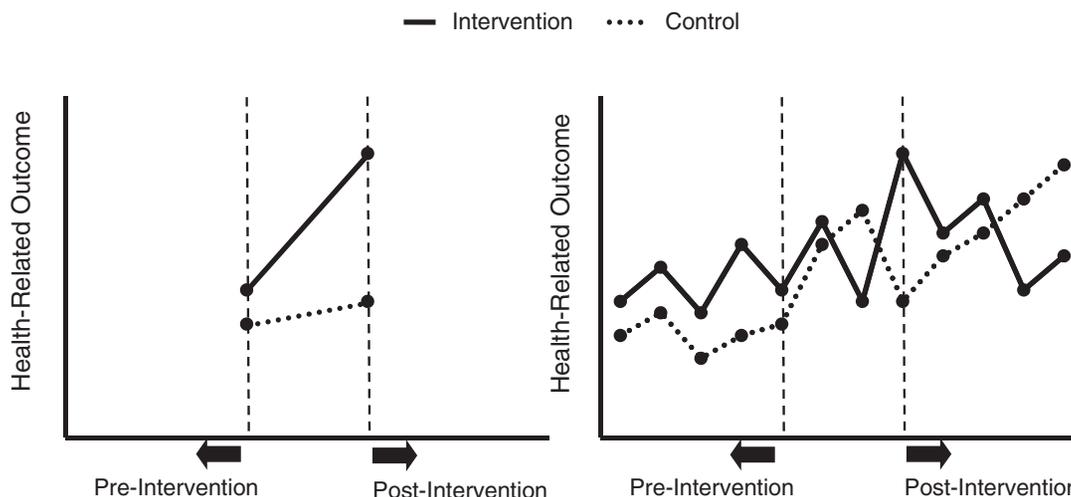


Figure 3. Simple controlled pre-post or difference-in-differences approach with linearity assumption (left) vs cITS design (right). Adapted from Wagenaar and Burris (2013).

Internal validity can be further strengthened by adding non-equivalent dependent variables not hypothesized to change in response to the intervention, but expected to respond to potential confounding factors in similar fashion as the dependent variable (Figure 1C) (Shadish *et al.* 2002). Community surveys can provide for multiple no-treatment control groups and non-equivalent dependent variables but are limited to applying this design at provincial or national levels given sample size considerations. Survey indicators rarely have the high time-resolution (weekly, monthly or quarterly estimates) to adequately measure and control for time trends. In our intervention example, we could select non-equivalent dependent variables of first antenatal care visits and outpatient consults because they are known to be of high quality, cover the same time period, and would be expected to change if there were any population changes or other confounding factors causing shifts in institutional birth counts at Vanduzi clinic.

Fourth, repeated measures at closely spaced intervals provided by RHIS are not only important for controlling secular trends due to extraneous factors, but also to understand typical patterns of gradual, S-curve, or decaying programme or policy effects (Wagenaar and Burris 2013). That is, depending on the level, type, outcome, or hypothesized mechanism that the health intervention acts through, one may anticipate a significant time lag before an effect is observed, a rapid shift in level with a slow regression to the mean, a slow gradual effect or even a change in outcome variability (McCleary and Hay 1980). In our Vanduzi example, one could a priori hypothesize that implementation of the birth kit incentive in a relatively small facility catchment area would result in an abrupt and sustained change in the number of institutional births, as dissemination is anticipated to occur quickly in a small, rural setting. Repeated measurements are a strong design improvement over most pre-post or difference-in-differences analyses that cannot observe intervention functional form and generally assume a linear trend between pre-intervention and post-intervention data points (Figure 3). Because linear trends are relatively rare when public health interventions are implemented in community settings, making this assumption can hinder proper understanding of intervention mechanisms and theories of change.

Fifth, the high time resolution of full-coverage RHIS data, along with the ability to observe and specify the functional form of the effect, allow for use of multiple replications (ABA or ABAB designs) and dose-response analyses to increase confidence that observed

intervention effects were not due to some idiosyncratic or uncontrolled factor (Biglan *et al.* 2000; Wagenaar and Burris 2013). If a health intervention or policy is amenable to introduction and removal in a relatively short time period, whether the observed intervention effect decays or returns to baseline after it has been removed can be assessed, which supports strong causal inference of a real intervention effect. Likewise, dose-response analyses can result from a natural experiment, where an intervention has haphazard differences in implementation fidelity or resources across jurisdictions, or could result from purposeful application of disparate intervention intensity and correlating these to effect trajectories. In our institutional birth example, to further increase strength of causal inference, one could replicate this intervention at different clinics (or remove and then reapply the incentive at Vanduzi clinic in an ABAB design), and examine dose-response effects by providing birth kits of differing values or for varying cut-offs of targeted recipients. Although dose-response can be potentially applied using community survey data, ABAB designs are not feasible with data gaps that often exceed multiple years.

Last, the design elements above allow for the use of advanced time-series analyses such as generalized estimating equations (GEEs) (Zorn 2001), segmented regression (Wagner *et al.* 2002), multi-level modelling (Goldstein *et al.* 1994) and autoregressive integrated moving average models (ARIMA) (Box and Jenkins 1970; Glass *et al.* 2008). One principal benefit of applying these analyses approaches to high resolution time-series data is their ability to forecast into the future, which could be useful for operations planning and resource allocation (McCleary and Hay 1980). ARIMA models have been applied to interrupted time-series quasi-experiments since the 1970s as a robust method to test the effects of, and advocate for, some of the most visible and effective public health interventions including mandatory seat belt laws (Wagenaar *et al.* 1988), increases in alcohol taxes and legal minimum drinking ages (O'Malley and Wagenaar 1991; Wagenaar *et al.* 2009), clean indoor air ordinances banning smoking (Khuder *et al.* 2007) and suicide prevention strategies (Morrell *et al.* 2007). Despite clear benefits, cITS designs have yet to be widely applied to research or programme evaluations in LMICs, where analyses are frequently situated within a complex nexus of ongoing Ministry of Health, non-governmental organization (NGO), bilateral/multilateral health interventions and programmes. We argue that cITS analyses are uniquely suited for these

contexts. The use of multiple control regions and outcomes, paired with many repeated measures over time, accurate specification of the theory-based functional form of the effect and multiple replications and/or dose-response analyses represent numerous design elements that are available to strengthen causal inference using RHIS data but are difficult or impossible to employ with available community survey data.

Conclusion

Data from RHISs enable the use of research designs that are superior to those relying on data from intermittent community surveys in their causal attribution of observed effects to an intervention or policy change. Notwithstanding these advantages, RHIS data present limitations for health evaluations. Accurate denominator estimates are often not clearly defined, especially for health facility catchment areas. Because populations typically change slowly, surveys and other methods could generate denominator estimates that can be combined with high time resolution RHIS time-series measures. Second, RHIS data only reflect what can be measured routinely at health facilities and are therefore best for evaluations of programmes or policies directly affecting care provided at facilities or outreach care and treatment, such as community immunization campaigns. Third, RHIS data miss those without health care access, often the population most in need of interventions. Given these limitations, in the short term, RHIS data will likely be most useful in estimating causal intervention effects on outputs and outcomes rather than population or community-level impact. Over time, health coverage concerns may diminish, especially if accurate denominator data are available to estimate those who are not currently reached by health services. If health care utilization and demographic patterns were clearly understood, rich clinical outcomes from RHIS could be weighted to achieve population representativeness.

Moving forward, investments in, and use of, RHIS for health evaluations should not only focus on causal effect estimation, but should also prioritize triangulating available data to develop implementation intensity measures of 'effective dose received' (Victoria *et al.* 2011; London School of Hygiene & Tropical Medicine 2012). The granular and continuous nature of RHIS makes them ideal for the development of 'effective dose' implementation metrics at all levels of the health system. For example, budget allocations, staff effort indicators and training, counts of visits by health problems/diseases, availability of essential commodities/equipment and supervision/monitoring information can interface with effect estimation to inform dose-response functions. As technology becomes less expensive and more available at lower levels of the health system, an increased number of implementation measures may be continuously available and of high quality.

The current reality of health data in Mozambique, like other LMICs, is that NGO-specific data collection, vertical programmes and intermittent community sample surveys have created fragmented data systems, led to inconsistent indicators, and have undermined in-country investment, commitment to and control of health data (Pfeiffer 2003). Governments increasingly rely on international partners to impute data based on models developed from other locations, or develop complex statistical models to address problems inherent in the structure of community surveys, while their own national information systems are neglected.

We recognize that community surveys will continue to be essential in the short to medium-term to allow accurate measurement of health practices and coverage on a population level, as well as to

compare with and potentially adjust or validate RHIS measurements. Nevertheless, given limited global resources, the high costs of intermittent community surveys have to be weighed against improving routine RHIS for evaluations, detailed measurement of implementation strength, primary health care organization, operations research, continuous quality improvement and resource allocation.

Increased attention to, and use of, RHIS data by the research and health policy communities will help build sustainable long-term capacity for effective management and evaluation of health improvement efforts in LMICs. Targeted improvement of RHIS data systems will support the movement away from population 'estimates' derived using complex modelling approaches and towards complete-coverage population 'measurements' (Byass 2010). We urge investment in data quality interventions to assess current RHIS capacity, and the use of RHIS data for robust programme and policy evaluations across all levels of health systems.

Conflict of interest

None declared.

References

- AbouZahr C, Boerma T. 2005. Health information systems: the foundations of public health. *Bulletin of the World Health Organization* 83: 578–83.
- Amouzou A, Kachaka W, Banda B *et al.* 2013. Monitoring child survival in 'real time' using routine health facility records: results from Malawi. *Tropical Medicine and International Health* 18(10). doi: 10.1111/tmi.12167.
- Biglan A, Ary D, Wagenaar AC. 2000. The value of interrupted time-series experiments for community intervention research. *Prevention Science* 1: 31–49.
- Bosch-Capblanch X, Ronveau O, Doyle V, Remedios V, Bchir A. 2009. Accuracy and quality of immunization information systems in forty-one low income countries. *Tropical Medicine and International Health* 14: 2–10.
- Box GEP, Jenkins GM. 1970. *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day.
- Braa J, Heywood A, Sahay S. 2012. Improving quality and use of data through data-use workshops: Zanzibar, United Republic of Tanzania. *Bulletin of the World Health Organization* 90: 379–84.
- Byass P. 2010. The imperfect world of global health estimates. *PLoS Medicine* 7: e1001006.
- Chanda E, Coleman M, Kleinschmidt I *et al.* 2012. Impact assessment of malaria vector control using routine surveillance data in Zambia: implications for monitoring and evaluation. *Malaria Journal* 11: 437.
- Chilundo B, Sundby J, Aanestad M. 2004. Analysing the quality of routine malaria data in Mozambique. *Malaria Journal* 3: 3.
- Cibulskis RE, Hiawalyer G. 2002. Information systems for health sector monitoring in Papua New Guinea. *Bulletin of the World Health Organization* 80: 752–8.
- Fernandes QF, Wagenaar BH, Anselmi L *et al.* 2014. Effects of health-system strengthening on under-5, infant, and neonatal mortality: 11-year provincial-level time-series analyses in Mozambique. *Lancet Global Health* 2(8) e468–e477.
- Gimbel S, Micek M, Lambdin B *et al.* 2011. An assessment of routine primary care health information system data quality in Sofala Province, Mozambique. *Population Health Metrics* 13:9:12.
- Glass VG, Willson VL, Gottman JM. 2008. *Design and Analysis of Time-Series Experiments*. Charlotte, NC: Information Age Publishing.
- Goldstein H, Healy MJR, Rasbash J. 1994. Multilevel time series models with applications to repeated measures data. *Statistics in Medicine* 13: 1643–55.
- Hazel E, Requeko J, David J, Bryce J. 2013. Measuring coverage in MNCH: evaluation of community-based treatment of childhood illness through household surveys. *PLoS Medicine* 10: e1001384.
- Instituto Nacional de Estatística. 2013. *Mozambique 2011 DHS Final Report*. Measure DHS. <http://dhsprogram.com/publications/publication-fr266-dhs-final-reports.cfm> (accessed 3 March 2015).

- Khuder S, Milz S, Jordan T *et al.* 2007. The impact of a smoking ban on hospital admissions for coronary heart disease. *Preventive Medicine* 45: 3–8.
- Lambdin BH, Micek MA, Koepsell TD *et al.* 2012. An assessment of the accuracy and availability of data in electronic patient tracking systems for patients receiving HIV treatment in central Mozambique. *BMC Health Services Research* 12: 30.
- Lim SS, David SB, Charrow A, Murray CJL. 2008. Tracking progress towards universal childhood immunisation and the impact of global initiatives: a systematic analysis of three-dose diphtheria, tetanus, and pertussis immunisation coverage. *The Lancet* 372: 2031–46.
- London School of Hygiene & Tropical Medicine. 2012. *Measuring Implementation Strength: Literature Review Draft Report 2012*. http://ideas.lshtm.ac.uk/sites/ideas.lshtm.ac.uk/files/Report_implementation_strength_Final_0.pdf, accessed 29 October 2014.
- Mate KS, Bennett B, Mphatswe W, Barker P, Rollins N. 2009. Challenges for routine health system data management in a large public programme to prevent mother-to-child HIV transmission in South Africa. *PLoS One* 4: e5483.
- McCarney R, Warner J, Illiffe S *et al.* 2007. The Hawthorne Effect: a randomized, controlled trial. *BMC Medical Research Methodology* 7: 30.
- McCleary R, Hay R. 1980. *Applied Time Series Analysis for the Social Sciences*. Thousand Oaks: Sage.
- MEASURE DHS. 2014. Demographic and Health Surveys. <http://www.dhsprogram.com/>
- Morrell S, Page AN, Taylor RJ. 2007. The decline in Australian young male suicide. *Social Science and Medicine* 64: 747–54.
- Mphatswe W, Mate KS, Bennett B *et al.* 2012. Improving public health information: a data quality intervention in KwaZulu-Natal, South Africa. *Bulletin of the World Health Organization* 90: 176–182.
- Mutale W, Chintu N, Amoroso C *et al.* 2013. Improving health information systems for decision making across five sub-Saharan African countries: implementation strategies from the African health initiative. *BMC Health Services Research* 13: 59.
- O'Malley PM, Wagenaar AC. 1991. Effects of minimum drinking age laws on alcohol use, related behaviors and traffic crash involvement among american youth: 1976–1987. *Journal of Studies on Alcohol* 52: 478–491.
- Pfeiffer J. 2003. International NGOs and primary health care in Mozambique: the need for a new model of collaboration. *Social Science and Medicine* 56: 725–38.
- Ronveaux O, Rickert D, Hadler S *et al.* 2005. The immunization data quality audit: verifying the quality and consistency of immunization monitoring systems. *Bulletin of the World Health Organization* 83: 503–510.
- Rothwell PM. 2005. External validity of randomised controlled trials: “to whom do the results of this trial apply?” *The Lancet* 365: 82–93.
- Rowe AK. 2009. Potential of integrated continuous surveys and quality management to support monitoring, evaluation, and the scale-up of health interventions in developing countries. *American Journal of Tropical Medicine and Hygiene* 80: 971–9.
- Rowe AK, Kachur PS, Yoon SS *et al.* 2009. Caution is required when using health facility-based data to evaluate the health impact of malaria control efforts in Africa. *Malaria Journal* 8: 209.
- Rowe AK, Steketee RW, Arnold F *et al.* 2007. Evaluating the impact of malaria control efforts on mortality in sub-Saharan Africa. *Tropical Medicine and International Health* 12: 1524–39.
- Shadish WR, Cook TD, Campbell DT. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Belmont, Wadsworth.
- Sherr K, Cuembelo F, Michel C *et al.* 2013. Strengthening integrated primary health care in Sofala, Mozambique. *BMC Health Services Research* 13(Suppl 2): S4.
- Short Fabic M, Choi Y, Bird S. 2013. A systematic review of demographic and health surveys: data availability and utilization for research. *Bulletin of the World Health Organization* 90: 604–612.
- The Global Fund to Fight AIDS, Tuberculosis, and Malaria. 2009. *Global Fund Five-Year Evaluation: Study Area 3—The Impact of Collective Efforts on the Reduction of the Disease Burden of AIDS, Tuberculosis, and Malaria*. http://www.theglobalfund.org/documents/terg/TERG_SA3_Report_en/, accessed 23 October 2014.
- UNICEF. 2014. Statistics and Monitoring: Multiple Indicator Cluster Survey.
- Verguet S, Jassat W, Bertram M *et al.* 2013. Impact of supplemental immunization activity (SIA) campaigns on health systems: findings from South Africa. *Journal of Epidemiology and Community Health* 67: 947–52.
- Victora CG, Black RE, Boerma TJ, Bryce J. 2011. Measuring impact in the millennium development goal era and beyond: a new approach to large-scale effectiveness evaluations. *The Lancet* 377: 85–95.
- Victora CG, Habicht J-P, Bryce J. 2004. Evidence-based public health: moving beyond randomized trials. *American Journal of Public Health* 94: 400–405.
- Wagenaar AC, Burris SC. 2013. *Public Health Law Research: Theory and Methods*. San Francisco: Jossey-Bass.
- Wagenaar AC, Maldonado-Molina MM, Wagenaar BH. 2009. Effects of alcohol tax increases on alcohol-related disease mortality in Alaska: time-series analyses from 1976–2004. *American Journal of Public Health* 99: 1464–70.
- Wagenaar AC, Maybee RG, Sullivan KP. 1988. Mandatory seat belt laws in eight states: a time-series evaluation. *Journal of Safety Research* 19: 31–70.
- Wagner AK, Soumerai SB, Zhang F, Ross-Degnan D. 2002. Segmented regression analysis of interrupted time series studies in medication use research. *Journal of Clinical Pharmacy and Therapeutics* 27: 299–309.
- Wagner BH, Gimbel S, Hoek R, *et al.* 2015. Effects of a health information system data quality intervention on concordance in Mozambique: time-series analyses from 2009–2012. *Population Health Metrics* 13, doi:10.1186/s12963-015-0043-3.
- World Health Organization. 2008. *Framework and standards for country health information systems. 2nd edition. Published by the Health metrics network and the world health organization*, Geneva, Switzerland.
- Zorn CJW. 2001. Generalized estimating equation models for correlated data: a review with applications. *American Journal of Political Science* 45: 470–490.