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 (Modulo 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 Fabric 2004; Rothwell 2005).

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 (Rouveaux 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 (Mphatwe 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 Estatistica 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.
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).
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 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 consultations 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 (McGlynn 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 RHS 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, RHS 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 RHS time-series measures. Second, RHS 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, RHS data miss those without health care access, often the population most in need of interventions. Given these limitations, in the short term, RHS 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 RHS could be weighted to achieve population representativeness.

Moving forward, investments in, and use of, RHS 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’ (Victora et al. 2011; London School of Hygiene & Tropical Medicine 2012). The granular and continuous nature of RHS 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 RHS measurements. Nevertheless, given limited global resources, the high costs of intermittent community surveys have to be weighed against improving routine RHS 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, RHS 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 RHS 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 RHS capacity, and the use of RHS data for robust programme and policy evaluations across all levels of health systems.

Conflict of interest

None declared.

References


