HSS Evaluation Collaborative

Evaluating Health Systems Strengthening Interventions Using Routinely Collected Data

July 2022

Michael R. Law, David R. Hotchkiss, and Karen A. Grépin

Contents

Executive summary

Over the past two decades, routine health facility data collection and use have increased substantially. Electronic routine health information systems (RHIS) that collect data from health facilities at regular intervals, such as DHIS2, have enabled this growth in many low and middle-income countries (LMICs). While data from these systems have important potential application in evaluating health systems strengthening (HSS) interventions, to date, they have remained less widely used in such assessments.

In recent years, however, many strong studies have been conducted using RHIS data to answer key questions about programme effectiveness and impact. These data have been used in evaluations of a wide spectrum of HSS policies and interventions, including user fee policies, health financing schemes, governance interventions, supply chain interventions, vaccination campaigns, and others. In this report, we provide an overview of using RHIS data for HSS evaluations, including their strengths and weaknesses.

There are important considerations that analysts must take into account when planning such evaluations using RHIS data, including whether the desired programme outputs are measurable with the fields available in the RHIS. Further, while research indicates that the quality of RHIS data has been increasing over time in many countries, assessments must be made to ensure that the data are accurate, reliable, relevant, and complete. In comparison to household or facility surveys, RHIS data are limited to the fields that already exist in the dataset and are not well-suited to all types of impact questions, in particular questions about the equity impact of interventions.

Once you have decided to use RHIS data in an evaluation, there are several steps and considerations that must be made. After understanding the context and details of the programme in question, indicators must be selected from the available fields in the RHIS that reflect the desired impacts of the programme in question. Following this, a data quality assessment should be conducted, and issues of missing data, outliers, and denominators for the calculation of rates should be addressed. To facilitate these steps, there is a growing understanding and methods base for dealing with these issues for analysts to draw upon.

Based on the available indicators and data timeframes, an analyst must then make a decision about what research design best suits the evaluation. There are many methods for using RHIS in a way that can answer questions about programme impact in a rigorous fashion. These include very simple approaches that describe trends, all the way up to very powerful methods such as difference-in-differences and interrupted time series approaches. In the full report, we provide an overview of several methods and discuss a specific example of each from the health policy literature.

RHIS are a powerful addition to the data sources available for HSS programme evaluations. With careful use and strong methods, they can produce insights that are strong, often less expensive, and frequently not possible with other data sources. As RHIS data continue to improve and their use in HSS assessment proliferates, we believe they will become a standard component in future programme design and evaluation.

Introduction

The collection and use of routinely collected health facility data have increased in low- and middle-income country (LMICs) health systems over the past two decades. This has been facilitated by the development, deployment, and strengthening of routine health information systems (RHIS) in many countries. Over that time, these data systems have increased in scope to cover more services, improved in their completeness, and are increasingly used by health system policymakers and health system managers to monitor system outputs.[1,2]

To date, however, RHIS data remain less widely used in the evaluation of health system strengthening (HSS) interventions than data from population-based household and facility surveys. This situation exists despite the availability of research methods that can produce rigorous estimates of impact of interventions and other actionable insights. This report reviews the use of data sourced from RHIS in HSS evaluations, provides examples of studies using different evaluation methodologies, and provides guidance on best practices in using RHIS data. This will include a discussion of data availability, data quality issues, different research designs, and other methodological considerations. We will also provide an overview of the comparative strengths and weaknesses of RHIS data-based approaches versus other evaluation data sources, in particular the use of household and facility surveys.

This report will provide a high-level overview through four sections that answer the following questions:

1. What are RHIS, and how has RHIS data been used to evaluate HSS interventions?
2. What are the pros and cons of using RHIS data for policy and programme evaluation?
3. How does one get started using RHIS data to evaluate an HSS intervention?
4. What research designs should be used with RHIS data?

Routine Health Information Systems and Evaluation

What are Routine Health Information Systems?

RHIS are systems that collect information from health facilities for operational purposes at regular intervals.[3] These data are typically collected monthly, stored electronically, and can be used for system tracking and management, commodity tracking and procurement, disease surveillance and monitoring, and to improve quality assessment, among other uses. For example, RHIS data systems might include information on the number of antenatal care visits, the stock of a particular drug in each health facility, staffing levels, or the number of malaria cases detected monthly. Often these data entries are based on facility register books or paper forms. There are a number of features that are typical of RHISs: they contain repeated observations, the data are longitudinal in nature, and they contain information from many health facilities.[4] These data can often be disaggregated along a few dimensions, such as the level of the health facility, the age or gender category of the patient, or different geographic areas.

Examples of RHIS Indicators

RHIS typically collect information across a range of indicators. These can include service utilisation, diagnoses, equipment and supply stocks, and other data. While the indicators available varies across countries, below are some examples that are commonly found in these systems. We have also cited some examples of studies that have used these specific indicators.

Services

1. Number of outpatient visits[5]
2. Number of antenatal care visits[6]
3. Vaccinations administered[7]

Diagnoses

1. Confirmed malaria diagnoses[8]
2. Hypertension diagnoses[9]

Supplies

1. Stock-outs of medications, vaccines, and other supplies[10]
2. Bed nets received and dispensed[11]

Over the past decade, many countries and Global Health Institutions (GHIs) have been investing in expanding and improving these systems, which has led to substantial increases in their use and the quality of the data they contain.[12,13] The most prominent system is the open source District Health Information System 2 (DHIS2) platform, which is a cloud-based system used in more than 70 countries with an estimated combined population of 2.4 billion people.[2] Along with systems that collect information at the facility level, there is also an increasing number of countries and systems that use DHIS2 platforms to collect data from individual patients in electronic medical records. While this would also be considered routine data, they are comparatively new systems and thus will not be the focus of our discussion. Similarly, a few country health systems have also begun to collect and store additional data into their RHIS, including health payments and other data. In theory, most of the methods we describe below could be equally useful to study other types of outcomes or effects.

How have RHIS data been used to evaluate HSS interventions?

The increased collection and availability of RHIS data has led to increased use of these data for evaluation purposes.[14] But while the number of studies published using RHIS data for health system evaluations has been increasing in frequency, use remains lower than it could be: the majority of studies have been published within the last 10 years.[14] The breadth of topics covered in these prior evaluations is vast, and examples include evaluations of user fee exemption policies, health financing schemes, health governance interventions, and other community-level interventions.[14] They have also covered a wide range of different health programme areas, including malaria prevention and treatment, immunisation, maternal and child health, HIV treatment, and others.[14] RHIS data have also been used to evaluate the impact of other large-scale events, such as infectious disease outbreaks and the COVID-19 pandemic, on the use of health services in numerous international settings.[9]

In a broad sense, RHIS data can be used to answer questions about the impact of HSS interventions, and there are many published examples available.[5,15–17] These evaluations are best suited to detecting the

impact of reforms or interventions that influence the variables that are collected in the RHIS. For example, they are better suited to evaluate changes in the quantity of care services delivered than the quality of those interventions. In terms of evaluation, they are also well-suited to detecting the effects of large-scale programmatic changes. In contrast, determining the impact of small interventions, such as a set of training activities at one health facility, will be very difficult to ascertain. They are also not well-suited to studying topics where the data capture is not adequate. For example, the impact of programmes on wealth-based equity and health care quality is often difficult or impossible to capture with RHIS data.

When reviewing different study designs later in this report, we will use examples from different HSS interventions and programmatic areas of global health. Recent reviews have concluded that these studies employ rigorous study designs and analytic methods, and many have been published in high-impact peer-reviewed health policy journals.[14]

What about data quality?

Despite the possibility of using RHIS data for evaluation purposes, they remain less utilised than would be expected given their breadth and availability, likely due to concerns voiced by a number of researchers.[18,19] Most commonly, questions over the use of RHIS data in HSS evaluations have centred on data quality.[20–22] In addition, several other factors have been identified as hindering their use, including technical challenges in collecting and disseminating the data.[23] As such, data quality remains an important consideration for any evaluator using RHIS data and is an area of active methods and policy development.[23]

High-quality data are data that meet the needs of intended users. There are many definitions of data quality, but the features most discussed in the literature in the context of RHIS are described in the box below.[21]

Aspects of RHIS Data Quality

1. *Accuracy:* refers to the ability of the data to get as close as possible to the true outcome of interest and avoid systematic biases.
2. *Reliability:* refers to the idea that if repeated measures are conducted that they are collected the same way over time.
3. *Consistency:* refers to whether data are measured in the same way in different facilities or regions of the country that is the focus of the evaluation.
4. *Relevance:* refers to whether the data are those that are needed or understood by users.
5. *Timeliness:* refers to the temporal availability of data in that the data becomes available in a timeframe that is useful.
6. *Completeness:* refers to the breadth or scale of the data that are collected.

Several recent studies have aimed to better understand RHIS data quality in various international contexts and how data quality has been changing over time. A number of studies have found good data consistency and completeness in various countries for several indicators.[6,24] Similarly, a study of 14 countries in Africa found that completeness was generally good, but several had issues with outliers and the consistency of the data over time.[13] In our experience working with these datasets, data on maternal and child health services, malaria, and utilisation metrics such as total visits tend to be well populated and accurate, whereas other areas such as non-communicable diseases, mental health, and less

commonly-used services are often less-so. This, however, will depend on the specific country and topic under study.

In response to these concerns, there have also been active initiatives to improve the quality of data in RHISs over the past decade.[25] In addition, several interventions have been shown to improve the quality of data being collected.[26,27] Many of these interventions are simple and relatively low cost.[28,29] For example, capacity building interventions, including onsite training and mentoring, as well as technological enhancements, such as electronic data collection tools, have generally been found to improve data quality.[29] As a result, there is good evidence that RHIS data quality has been improving over time.[13,30] There are also emerging methods to deal with many data quality issues that we will review below, most notably missing data, inconsistent data, and outliers.[4,31] Improvements have also come from changes to data collection: for example, the availability of data on malaria burden has improved with the advent of rapid diagnostic testing and an increase in the use of RHIS for surveillance.[20]

The number of studies using RHIS data being published in leading peer-reviewed health policy journals suggests that data quality has reached a level where it can support evaluations in many countries.[14] Caution is still warranted, however, as implausible values and data quality remain a concern, even in countries where reporting is nearly complete.[32] Below, we will review many methods to address common data quality concerns that have shown good results. While such adjustments are important, it is also vital to understand that RHIS data do not need to be perfect to conduct a rigorous evaluation.[33] This is particularly true with more rigorous designs such as an interrupted time series analysis, which can address some of these concerns. In sum, the methods, breadth, and data quality in RHISs are continually improving, which we believe will continue to be an argument for expanding their use in HSS evaluation in the future.

What are the pros and cons of using RHIS data for policy and programme evaluation?

When planning an evaluation of an HSS intervention, programme managers, funders, and governments often have a choice of different data sources methods available to them. The two most commonly used data sources for evaluations of HSS interventions are RHIS data and household surveys. Importantly, it is important to note that neither of these data sources is perfect: each has distinct benefits and drawbacks that must be taken into consideration when considering an evaluation approach. For example, they have a wide variation in their cost, speed, the outcomes they can collect, and their ability to draw causal conclusions about programme impact. What is key is to understand these limitations and take them into consideration in the interpretation of the findings of any study.

RHIS data

Studies using RHIS data have several strengths that can make them an attractive option for evaluating HSS interventions. RHIS data tend to be made available relatively quickly, and studies based on RHIS data are less costly, as the necessary data have already been collected from health facilities. In fact, evaluation metrics can often be built into existing dashboards for real-time assessments of programme performance.[34] They can also enable analyses at the sub-national or individual facility level, which is important for HSS interventions that are not national in scope.[3] There are numerous options available for study design, several of which can produce very rigorous results that can be interpreted in a causal fashion. And finally, they are often possible to conduct after a programme has been put into place as the data will still be available for the pre-intervention period. RHIS also collect information on health system indicators that are not easy to capture in household surveys (e.g., data on stock-outs).

Household surveys

Evaluations of the impact of HSS interventions using household survey data have a long history. They have the distinct advantage of allowing the researcher to collect rich data from households and individuals that is tailored to answer the research questions being investigated.[19] This includes being able to study the impact of programmes based on household differences, sex and gender, and wealth or income levels. This contrasts with RHIS-based studies, which are limited to studying the variables and outcomes that are available in the data system, and where data are aggregated to the facility level and thus are limited in their ability to study individual-level characteristics. Household surveys can also capture the availability and use of services that are provided in private facilities, which are commonly missing from RHIS, such as DHIS2. In contrast to high-frequency RHIS, however, household surveys are infrequently conducted, very expensive, and are often unable to provide estimates at the sub-national level. The data collection process is also often slow, and the data are often not released until many months or years after they have been collected.[35] Finally, the lack of longitudinal data often leads to the use of sub-optimal study designs, such as cross-sectional comparisons or dose-response studies, which can be subject to serious biases.[14]

How does one get started using RHIS to evaluate an HSS intervention?

In planning and conducting an evaluation of an HSS intervention with RHIS, there are several considerations that need to be part of the process.

Understand your context

It is vital to understand the context in which the evaluation is taking place and the history of other associated interventions or programmes. The investigator should have a strong understanding of the different interventions that might have affected their outcomes, other policies that were put into place around the same time, their theory of change, and the pre-existing rates in the outcomes they are studying.

Another component of understanding your context is knowing the rules about accessing and using the RHIS in any given country. Rules vary widely between different jurisdictions, so knowing how to collaborate with data stewards is vital. Currently, many countries do not provide public access to data in the RHIS. In such contexts, data may only be accessed in partnership with particular institutions or through permission from government agencies. In our experience, we have found that working in close collaboration with Ministries of Health and in-country investigators greatly facilitates both access to data and the ability to understand, contextualise and disseminate results. Rules around research ethics also vary by country, but many countries require that investigators obtain approval from a local research ethics board as a prerequisite for using RHIS data.

Indicator selection

After getting an understanding of the context and given the theory of change of the intervention, it is important to evaluate the data available in the RHIS and choose indicators to use for the evaluation. These should be based on the desired impact of the HSS intervention and can reflect either process outcomes (e.g., drug availability), direct outcomes (e.g., the number of services delivered) or impact evaluation (e.g., cases of a disease averted). When selecting indicators, one is obviously limited to the indicators that are part of the RHIS, which varies by country.[36] Several countries make their data

collection forms public, and these can be used to initially inform indicator choice.

Care must be taken to ensure that the question being asked in the evaluation can actually be answered by the data that exist within the RHIS.[19] In other words, the outcomes that the HSS intervention was intended to target should be in the RHIS in some form. Further, the indicators must be an accurate reflection of the underlying concept of interest. For example, malaria mortality figures are often understated in RHIS data as most deaths do not occur in health facilities but rather in the community.[20] Therefore, it might not be appropriate to use them to evaluate an intervention targeting overall malaria death rates. In contrast, indicators such as facility delivery rates can be very well-captured in RHIS systems.

When selecting indicators, the following additional factors should be kept in mind:

1. The choice should be based on an assessment of data quality.[21] In some RHISs, there are some variables that are more complete, more consistent, or available for more facilities. This will obviously have implications for what indicators will be useful for an evaluation.
2. It is vital to understand the way in which the data are generated, their definition, and how they are aggregated.[18] For example, there may be variation in the data if records are generated based on paper forms that are inputted versus data that is directly entered into the RHIS. Even within-country variation is possible in this respect, which is important to understand. For example, smaller health facilities may use paper forms, and larger hospitals might enter data electronically. This may also lead to variation in data quality.
3. Investigators must be cautious of definitional changes in the RHIS over time. Issues with such changes have been highlighted in several studies. For example, prior reviews of HIV-related studies using RHIS have found issues with changing definitions in several countries.[28] Similarly, changes in malaria case counts from including both presumptive and confirmed cases to just confirmed cases can lead to measurement challenges.[33] Changes might also occur when collection methods are modified: in our work, changes in Rwanda's DHIS2 collection forms in 2012 made several indicators inconsistent across that date.[6] Plotting potential indicators longitudinally is an important first step in any analysis, examining for changes that are not policy- or intervention-related.[33]

Assessing and addressing data quality

Prior to analysis, an analyst must address any concerns with the quality of the data. This includes a series of considerations, including assessing and addressing missing data, inaccurate data, and the consistency of their data over time. Methods for assessing and potentially addressing completeness, internal consistency, and external consistency exist and can be implemented by an analyst prior to final modelling.[21] Below, we will discuss each of these considerations in turn.

Missing data

Missing data represent one of the most commonly cited reasons for poor data quality in RHISs, but it has been improving in DHIS2-based systems over time.[23] Prior to analysis, an investigator should examine the completeness of data at the indicator-specific level. There are many reasons why data generated at health facilities may be missing. This missingness is often not random (e.g. more remote or busier facilities), which may lead to bias if they are excluded from analyses.[22] More than 80% completeness is desirable, and the more complete the dataset, the better.[24]

In an effort to improve data completeness, several studies have gone back to the original registries to manually fill in missing values.[14] This might be an option when there are not many missing data, and the

original registries exist, are in fact complete, and can be easily accessed by the researcher. If this is infeasible, there are several different methods available to impute these values for analytic purposes.[22] In a review of seven different methods and using data from the Democratic Republic of the Congo (DRC), Feng et al. found that multiple imputation methods performed the best for these purposes with DHIS2 data.[22] In such methods, the missing values are generated several times and estimates recombined to reflect their uncertainty. These methods can be implemented using freely-available software packages.[37] Notably, just dropping the data points or facilities with missing data from a study is likely the worst approach to dealing with missingness and should be avoided if possible.[22]

There may be instances in which not all data are collected at all types of health facilities, and therefore this type of missing data may need to be treated differently. For example, it is not uncommon for hospitals to have very incomplete data when it comes to services such as immunisation, and it is not uncommon for lower-level health facilities to perform very few caesarean sections or other more specialised services. Plus, some health services may only be delivered at some times during the year. Therefore, it is also important for the user to take the time to try to understand the reasons why data may be missing in each context and to use this information in their assessment.

Outliers

Oftentimes, individual data points may appear to be unusually smaller or larger than one would expect. These "outliers" should be assessed, either through visual inspection or through one of several statistical methods that can address them.[14] This is particularly important when data verification is weak or not present so that implausible values do not unduly influence the analysis estimates. This often occurs when data are collected using paper forms, as the underlying RHIS software cannot validate the numbers given. While there are different bars for what constitutes an outlier, the WHO suggests that more than three standard deviations should be considered an extreme outlier in routine health information systems, and more than two standard deviations a moderate outlier.[21] These types of values should be investigated to see if they are true and corrected if they are not. If possible, these values could be compared to the original source registry. If that is not available or possible, another option is to simply drop them and then re-impute them using the missing data techniques discussed above.

Data stability

The analyst should also assess trends in the data over time and check that monthly values align with historical norms and show clear trends over time.[24] There should be reasonable explanations for deviations that you see in your data, such as the intervention itself or seasonality in outcomes (e.g. malaria rates). The level or trend in the outcome can also shift around the time of the intervention being studied if it is having an effect. Should you observe other abrupt changes that you can't explain, the investigator should go back to their contextual overview to look for other explanations (e.g., other policy interventions), look for definitional changes or changes in data collection methods, or determine if it's a different type of error in the data.

Denominators for comparisons

In any evaluation where values over time or between regions will be compared, it is important to have appropriate denominators for the calculation of population-level rates. For example, this might be the number of visits to a health facility divided by the size of their catchment population. This allows an analyst to compare values over time when there are shifts in the population, and to compare between facilities or regions with different underlying populations.

There are several different sources of denominators available depending on the level of aggregation (health facility, district, etc.) and what other data are available in the jurisdiction. For regional

populations, projections include those from the census or existing survey data.[32] As populations generally change slowly (outside of crisis situations with mass migration), surveys and/or census methods can be combined with RHIS measures to determine rates.[4] It is worth noting, however, that often these national data sources are used to estimate sub-national areas, which can introduce bias into an evaluation depending on the study design that is used.[18] If the key outcome of your study is service coverage or rates, it is important to consider that low-quality population estimates used as denominators can produce implausible values for some indicators.[35] If reasonable denominators are not available, it may be preferable to conduct the analysis using raw count data versus rates. This will, however, make comparisons between facilities or regions more difficult.

Beyond regional numbers, authors have argued that there is a need to produce better denominators to facilitate analyses using RHIS.[13] This is particularly the case at the facility level. There are many methods based on both area and road distance that have been used to estimate catchment populations.[20] One can also aggregate data to the district level if better population estimates are available, although this may not fit with how the intervention was actually put into place.[20] Researchers have also developed methods to derive and adjust indicators from the DHIS2 itself, which can be used for some services.[24,38]

Finally, it is worth noting that the precision of the denominator that is desirable for a study depends on the nature of the study design. For analyses where the goal is to determine the population level rate, obviously the calculation of a precise denominator is paramount. In contrast, longitudinal analyses of the rate of change in an indicator can still proceed with less precise indicators, as the relative change in the rate is the true outcome of interest.

What research designs should be used with RHIS data?

Based on the available indicators and timeframes, a decision must be made about what research design best suits the research question and available data. Study design encompasses the choice about how the data will be set up and analysed, what groups or areas will be included in the analysis, and what threats to validity might be a concern—in other words, how factors other than the HSS intervention might explain the results that you find.

A vital concept in research designs is that of the counterfactual. A counterfactual represents the assumptions that are made about what the world would look like absent the HSS intervention in question. For example, if a health financing intervention was put into place to increase case presentation rates, the counterfactual world would be the level of case presentation rates without the intervention. The issue with counterfactuals is that they can never actually be observed, as the world without the intervention will never exist. Different research designs make different assumptions to estimate the counterfactual, and each has its own strengths and weaknesses.

Along with the counterfactual, it is also vital to understand the major threats to validity that can impact each different study design. The major threats that are a concern include the following, with examples from a hypothetical intervention to reduce stock-outs of essential medicines in health facilities:

1. History: this represents something else in the world that impacts the outcome around the same time as the intervention being studied. For example, if there was a global shortage of a particular medicine that started around the same time.
2. Instrumentation: this represents a change in measurement that occurs around the same time as the intervention. For example, if the list of tracker medications changed around the same time as the intervention.

3. Maturation: this represents trends over time that are not related to the intervention under study. For example, if the use of a particular drug was trending up or down over time due to a changing burden of disease.
4. Selection: this represents differences between groups that are and are not selected to receive an intervention. For example, if regions with particularly high medicine use are chosen to receive the intervention, and these regions also have lower income levels.

Below we will review several different research designs, moving from designs that are often subject to serious biases to ones that are generally considered more rigorous. An overall summary table of the different methods is presented in **Table *1*** at the end of this section.

Descriptive trend analyses

Likely the simplest research design to use with RHIS data, descriptive trend analyses examine outcomes over time for a particular indicator. For example, one might track the facility delivery rate or the prevalence of non-communicable diseases over time. These types of assessments can be useful for monitoring subnational trends and performance.[24] However, they are not particularly helpful in evaluating specific HSS interventions, as many other factors can be at play in any particular jurisdiction. For example, due to maturation, the trend in an outcome may be a longstanding feature and not the result of any particular intervention. However, these methods can be suitable at the macro level when the key research question pertains to meeting targets or thresholds. For example, progress toward UHC targets in South Africa has been completed using DHIS2 data to assess several indicators over time.[39]

Dose-response studies

Another comparatively simple research design that can be used with RHIS data is a dose-response study, also known as a cross-sectional comparison. In this design, the analyst examines the outcomes from RHIS indicators in regions with different policy intensities to establish a dose-response relationship. The counterfactual is determined by assuming that the outcomes in low-dose regions represent what would have happened in high-dose regions had they themselves been low-dose. This design has been used in the study of malaria interventions in Eritrea and Zambia.[20]

While reasonably simple to undertake, the obvious concern in these studies is selection bias: something else could explain the differences found between regions. For example, factors that may have led implementers to target one region over another, and these factors could be unobserved by researchers. Given the longitudinal nature of RHIS data, it is often better to seek out a stronger research design such as those below.

Example dose-response Study

An example of a dose-response study by Bennett et al. studied the association between an insecticide-treated net distribution programme and malaria incidence in Zambia.[40] Using RHIS data at the district level between 2009 and 2011, they found that an increase in net distribution of one net per household was associated with a 27% reduction in malaria cases. While the study accounted for many potential confounders, such as climate and testing rates, there may have been other unobserved factors that could have resulted in selection bias. For example, the authors note that the programme was targeted at high-incidence and easily accessible areas, which means that other regions may not represent a good counterfactual in this case.

Pre-post studies

Among the more commonly used approaches with RHIS data, pre-post designs are another example of a comparatively simple approach to policy and programme evaluation. In this method, the outcomes of interest are compared between the period prior to the intervention with the outcomes afterwards. The counterfactual assumption in these studies is that the level of the outcome would have remained the same absent the intervention, as shown by the dotted line in Figure *1*. The estimated impact of the programme is then calculated as the difference in the actual observed outcome versus this counterfactual estimate. A pre-post approach has been used often with RHIS data: the majority of studies in a review of maternal and neonatal health intervention evaluations using RHIS were of this type (30 of 46 reviewed studies).[41]

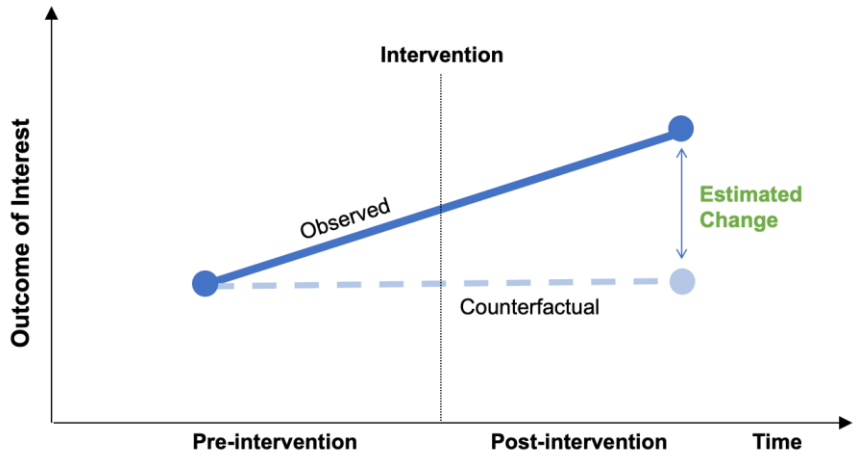

Figure 1. Setup and counterfactual in a pre-post study design

While pre-post studies are relatively simple to conduct, they are also highly subject to potential threats to validity. For example, history threats are quite common in these studies, as multiple interventions are commonly put into place at the same time, or other events might occur between pre- and post-periods. Similarly, maturation can often be a concern, as it is rare for indicators to have flat trajectories over time. Finally, instrumentation can also be a factor, as changes in measurement might occur, or the programme itself might cause changes, such as better detection of disease.

Example pre-post Study

An example of a pre-post study using RHIS data was published by Mphatswe and colleagues in 2011.[26] This study documented earlier efforts to improve the data contained in the DHIS-based system in in KwaZulu-Natal province, South Africa. The authors evaluated a 2008 audit and feedback intervention aimed at increasing data reporting into the RHIS. As shown in Figure *2*, the authors found an increase of 38% between pre- and post-intervention periods.

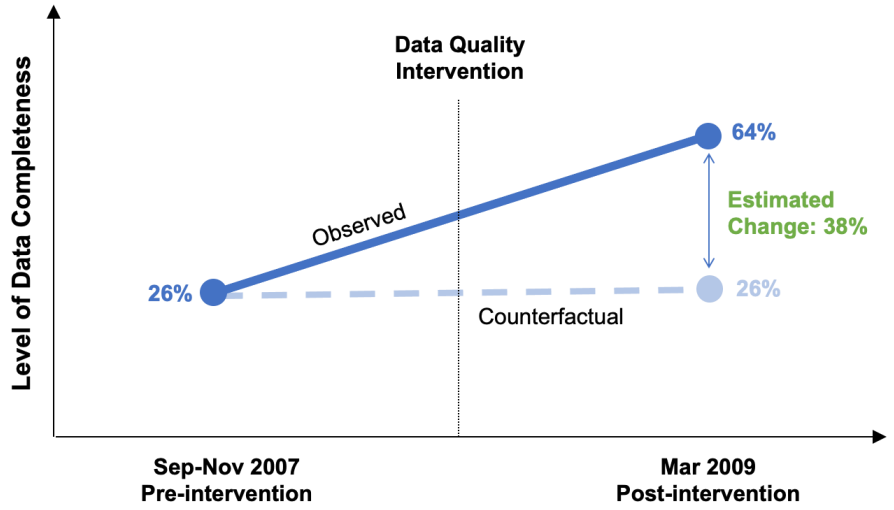

Figure 2. Example of a pre-post study conducted by Mphatswe et al.

Difference-in-differences designs

An extension on a pre-post design that offers more rigour is the so-called difference-in-differences design.[42] In this study approach, both an intervention group and a control group are included in the study and outcomes are assessed before and after the intervention begins. The counterfactual in this case is determined by the pre-post difference in the control group, as shown in Figure 3. This helps the analyst account for any changes between the pre and post periods that were not related to the programme. For example, if there were other interventions or underlying trends that were common to both the intervention group and control group.

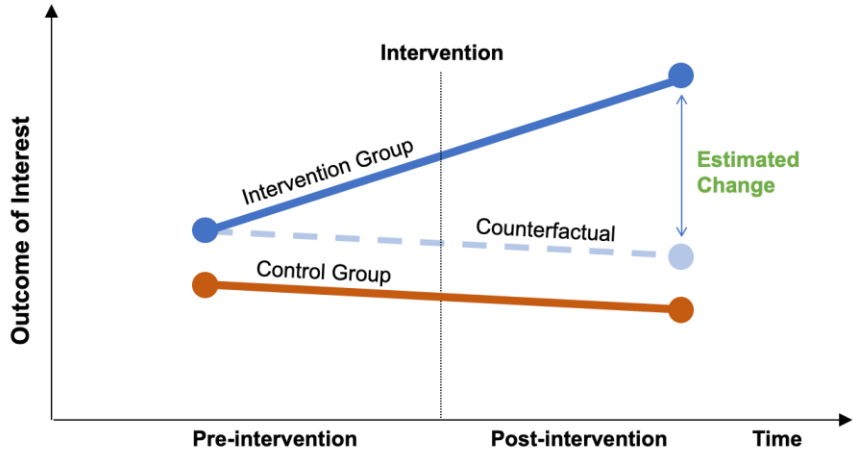

Figure 3. Setup and counterfactual in a difference-in-differences study design

There are several options for a control group in this type of study. The most common is another region or facility (or group of facilities) that did not implement the intervention in question. Often the control group is chosen based on similar characteristics or location. The selection of a control group is an important consideration as it has implications for potential threats to validity. In this design, the concern about bias would come from the counterfactual not accurately representing the trajectory the intervention group would have experienced absent the intervention. This might be due to history

(another event that happened in one region and not the other), maturation (different underlying rates of change in the outcome in one group and not the other), or instrumentation (changes in measurement in one group and not the other). It is, therefore, common to test for pre-existing trends in difference-in-difference study designs to determine if comparison areas had similar trends in the pre-intervention period and thus could represent a suitable comparison group.

Example differences-in-differences Study

A paper by Steenland et al. used a differences-in-differences approach to examine the impact of a performance-based financing (PBF) scheme that was piloted in three of Burkina Faso's health districts starting in 2011.[43] The aim of the programme was to increase the use of maternal health services. The outcomes in the three intervention districts were compared to three comparison districts from the same region with similar health systems and socio-economic characteristics.

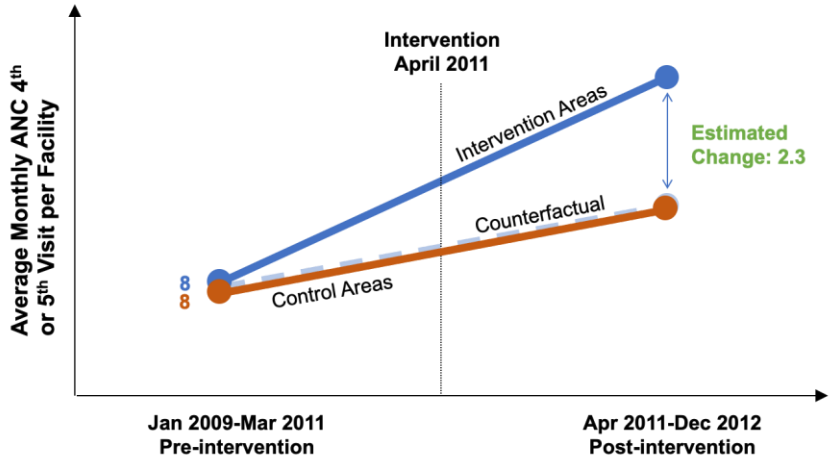

Figure 4. Example of a pre-post study conducted by Steenland et. al.

One of the key outcomes in this study was the average monthly number of 4th or 5th antenatal care (ANC) visits. A simplified version of their results is shown in Figure 4. It is important to note that while the pre-intervention visit rates were similar between the intervention and control districts, the authors also found an increase in the visit rate in the control areas. Therefore, their estimated change of 2.3 extra visits per facility per month is based on the additional increase that was observed in the intervention areas.

Interrupted time series analysis

One of the most rigorous quasi-experimental designs is an interrupted time series (ITS) analysis, which uses data over multiple time periods before and after the implementation of an intervention. The counterfactual in ITS is based on the pre-intervention trend in the outcome, which is projected forward to determine the counterfactual, as shown in Figure 5. The observed values are then compared to this projected trend, and the analyst can determine both an immediate level change and any long-term changes in the trend of the outcome being studied.

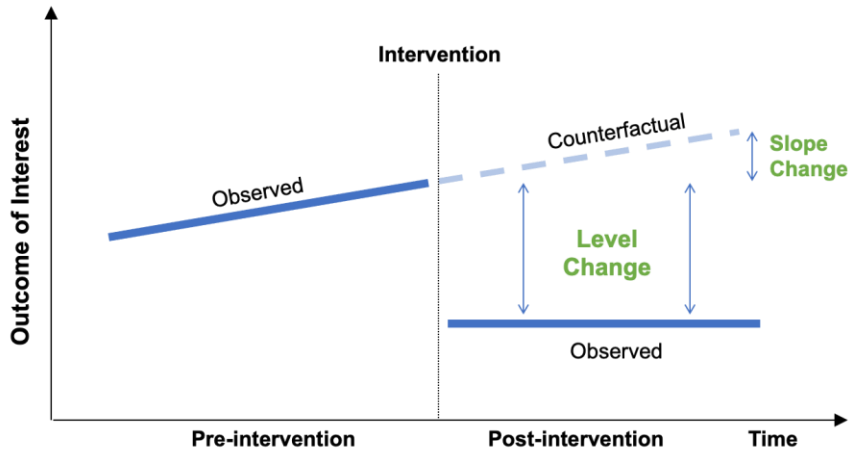

Figure 5. Setup and counterfactual in an interrupted time series design

The use of ITS has been recommended in reviews on evaluating HSS interventions using RHIS.[4,33] Due to these strengths, ITS is rapidly growing in use and has been used to evaluate quality improvement interventions in LMICs in several cases.[44] The methods for conducting ITS studies have been outlined in several prior articles.[36,45,46] It is also possible in these studies to add a control group, as in prior work investigating the impact of an HSS intervention in Rwanda.[5]

The major threat to validity in an ITS analysis are contemporaneous changes that impact the outcome of interest. The main concerns are history threats (something else that is implemented or changed at the same time) or instrumentation (a change in measurement that is closely timed to the intervention). Non-linearity in the outcome can also be a concern, but analysts can often account for this by using curved (quadratic) or seasonal (harmonic) terms in the model. Using a control group can help in limiting some of these threats to validity, particularly if the control had a similar trajectory in the outcome during a specified time period prior to the intervention being introduced.

Example ITS Study

A study of the impact of COVID-19 on the utilisation of other health services in the Democratic Republic of the Congo provides an example of an interrupted time series analysis using RHIS data.[9] This analysis focused on health facilities in the capital city of Kinshasa and examined numerous indicators of health services utilisation, including outpatient visits, infectious disease diagnoses, vaccinations, maternal health services, and diagnoses of two non-communicable diseases. One health zone within the city was put on strict lockdown during the first three months of the pandemic.

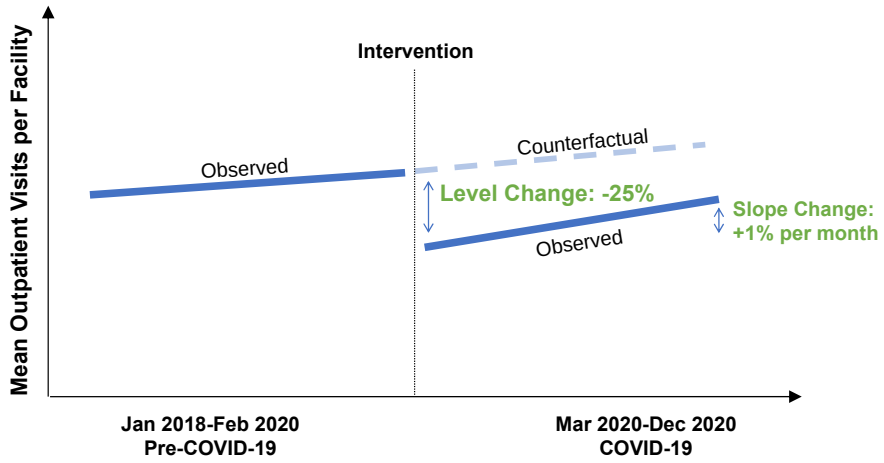

Figure 6. Example of an interrupted times series study conducted by Hategeka et al.

A simplified version of their model results is shown in Figure 6. As shown in the Figure, the average number of outpatient visits per facility per month was stable and slightly increased during the pre-COVID period. The counterfactual assumption in this study would have been that this increase would have continued, as shown by the dashed projection. Following the start of the pandemic, their analysis found an immediate level change in the rate of visits of -25%, accompanied by a trend increase of about 1% per month thereafter. This means that over time the difference between the observed visit rate and the counterfactual expectation was getting smaller.

Method	Description	Ability to Identify Causal Effects	Data Required	Example Study Reference	Example Study Question
Descriptive Trend Analysis	Examine outcomes over time for an indicator.	Weak, as it is difficult to determine what is causing changes in the trend over time.	Longitudinal data on the outcomes of interest.	Day et al. 2021.[39]	What progress is being made toward UHC targets?
Dose-response Study	Study whether outcomes are associated with the level of an intervention in different regions or facilities.	Weak, as selection bias can often be a factor. For example, facilities that receive an intervention can systematically differ from those that don't.	Data on who received the intervention and the outcome(s) of interest.	Bennett et al. 2014.[40]	What is the association between insecticide-treated net distribution and malaria incidence?
Pre-post Study	Compare outcomes before programme implementation to those after in the same region.	Low, as it is difficult to know if any change resulted from other factors or if there was a trend over time.	Data on the outcomes of interest from before and after the intervention is put into place.	Mphatswe et al. 2012.[26]	Can an audit and feedback intervention increase the data quality in a DHIS2-based RHIS system?
Difference-in-difference Design	Compare the before-after change in an intervention region to the before-after trend in a control region.	Medium, as the control group can aid in estimating what change would have occurred absent the intervention.	Data on the outcomes of interest from before and after the intervention is put into place in both intervention and control regions/facilities.	Steenland et al. 2017.[43]	What impact did a performance-based financing scheme have on the use of maternal health services?
Interrupted Time Series Analysis	Examine the level and trend of the outcome in the intervention area over time to assess if it changes when the intervention is introduced.	Strong, as it would only be subject to bias if other programmes or factors changed around the same time.	Data on the outcomes of interest in several time periods before and several time periods after the intervention is introduced.	Hategeka et al. 2021.[9]	What impact did COVID-19 lockdowns have on the use of health services?

Table 1. An overview of different evaluation methods that can be used with RHIS data to evaluate HSS interventions.

Conclusion

We hope that the above discussion and examples have shown the potential for RHIS to be a powerful addition to the repertoire of data available for HSS programme evaluations. While care is still required for their use, the advent of powerful methods, examples, and study designs should leave programme managers and evaluators ready to consider their use in the future. This should be accompanied by intentional moves from GHIs to involve themselves in improving RHIS data quality and making it more accessible for HSS evaluations. As RHIS data continue to improve and their use in HSS assessment proliferates, we believe they will become a standard component in future programme design and evaluation.

References

1. Nutley T, Reynolds HW. Improving the use of health data for health system strengthening. Glob Health Action. 2013 Feb 13;6:10.3402/gha.v6i0.20001.

2. Byrne E, Saebø JI. Routine Use of DHIS2 Data: A Scoping Review [Internet]. 2021 [cited 2021 Nov 6]. Available from: https://www.researchsquare.com/article/rs-965378/v1

3. Hotchkiss DR, Diana ML, Foreit KGF. How can routine health information systems improve health systems functioning in low- and middle-income countries? Assessing the evidence base. Adv Health Care Manag. 2012;12:25–58.

4. Wagenaar BH, Sherr K, Fernandes Q, Wagenaar AC. Using routine health information systems for well-designed health evaluations in low- and middle-income countries. Health Policy Plan. 2015 Apr 16;

5. Iyer HS, Hirschhorn LR, Nisingizwe MP, Kamanzi E, Drobac PC, Rwabukwisi FC, et al. Impact of a district-wide health center strengthening intervention on healthcare utilisation in rural Rwanda: Use of interrupted time series analysis. PloS One. 2017;12(8):e0182418.

6. Ruton H, Musabyimana A, Gaju E, Berhe A, Grépin KA, Ngenzi J, et al. The impact of an mHealth monitoring system on health care utilisation by mothers and children: an evaluation using routine health information in Rwanda. Health Policy Plan. 2018 Oct;33(8):920–7.

7. Dolan SB, Carnahan E, Shearer JC, Beylerian EN, Thompson J, Gilbert SS, et al. Redefining vaccination coverage and timeliness measures using electronic immunisation registry data in low- and middle-income countries. Vaccine. 2019 Mar 22;37(13):1859–67.

8. Ashton RA, Bennett A, Al-Mafazy A-W, Abass AK, Msellem MI, McElroy P, et al. Use of Routine Health Information System Data to Evaluate Impact of Malaria Control Interventions in Zanzibar, Tanzania from 2000 to 2015. eClinicalMedicine. 2019 Jul 1;12:11–9.

9. Hategeka C, Carter SE, Chenge FM, Katanga EN, Lurton G, Mayaka SM-N, et al. Impact of the COVID-19 pandemic and response on the utilisation of health services in public facilities during the first wave in Kinshasa, the Democratic Republic of the Congo. BMJ Glob Health. 2021 Jul 1;6(7):e005955.

10. Sato R, Thompson A, Sani I, Metiboba L, Giwa A, Femi-Ojo O, et al. Effect of Vaccine Direct Delivery (VDD) on vaccine stock-outs and number of vaccinations: Case study from Bauchi State, Nigeria. Vaccine. 2021 Mar 1;39(9):1445–51.

11. Suiyanka L, Alegana VA, Snow RW. Insecticide-treated net distribution in Western Kenya: impacts related to COVID-19 and health worker strikes. Int Health. 2021 Aug 16;1–3.

12. Jain S, Zorzi N. Investing for Impact: The Global Fund Approach to Measurement of AIDS Response. AIDS Behav. 2017;21(Suppl 1):91–100.

13. Maïga A, Jiwani SS, Mutua MK, Porth TA, Taylor CM, Asiki G, et al. Generating statistics from health

facility data: the state of routine health information systems in Eastern and Southern Africa. BMJ Glob Health. 2019 Sep 1;4(5):e001849.

14. Hung YW, Hoxha K, Irwin BR, Law MR, Grépin KA. Using routine health information data for research in low- and middle-income countries: a systematic review. BMC Health Serv Res. 2020 Aug 25;20(1):790.

15. Ir P, Horemans D, Souk N, Van Damme W. Using targeted vouchers and health equity funds to improve access to skilled birth attendants for poor women: a case study in three rural health districts in Cambodia. BMC Pregnancy Childbirth. 2010 Jan 7;10(1):1.

16. Bellows B, Kyobutungi C, Mutua MK, Warren C, Ezeh A. Increase in facility-based deliveries associated with a maternal health voucher programmeme in informal settlements in Nairobi, Kenya. Health Policy Plan. 2013 Mar;28(2):134–42.

17. Wagenaar BH, Gimbel S, Hoek R, Pfeiffer J, Michel C, Manuel JL, et al. Effects of a health information system data quality intervention on concordance in Mozambique: time-series analyses from 2009–2012. Popul Health Metr. 2015 Mar 26;13:9.

18. Watson-Grant S, Sutherland E, Xiong K, Thomas J. Beyond convenience: practical considerations with using routine health data for evaluations. Perspect Public Health. 2021 May 1;141(3):129–30.

19. Thomas JC, Doherty K, Watson-Grant S, Kumar M. Advances in monitoring and evaluation in low- and middle-income countries. Eval Programme Plann. 2021 Dec;89:101994.

20. Ashton RA, Bennett A, Yukich J, Bhattarai A, Keating J, Eisele TP. Methodological Considerations for Use of Routine Health Information System Data to Evaluate Malaria Programme Impact in an Era of Declining Malaria Transmission. Am J Trop Med Hyg. 2017 Sep 27;97(3_Suppl):46–57.

21. World Health Organization. Data quality review: module 1: framework and metrics [Internet]. World Health Organization; 2017 [cited 2021 Nov 21]. 30 p. Available from: https://apps.who.int/iris/handle/10665/259224

22. Feng S, Hategeka C, Grépin KA. Addressing missing values in routine health information system data: an evaluation of imputation methods using data from the Democratic Republic of the Congo during the COVID-19 pandemic. Popul Health Metr. 2021 Nov 4;19(1):44.

23. Hoxha K, Hung YW, Irwin BR, Grépin KA. Understanding the challenges associated with the use of data from routine health information systems in low- and middle-income countries: A systematic review. Health Inf Manag J Health Inf Manag Assoc Aust. 2020 Jun 30;1833358320928729.

24. Maina I, Wanjala P, Soti D, Kipruto H, Droti B, Boerma T. Using health-facility data to assess subnational coverage of maternal and child health indicators, Kenya. Bull World Health Organ. 2017 Oct 1;95(10):683–94.

25. Lee J, Lynch CA, Hashiguchi LO, Snow RW, Herz ND, Webster J, et al. Interventions to improve district-level routine health data in low-income and middle-income countries: a systematic review. BMJ Glob Health. 2021 Jun;6(6):e004223.

26. Mphatswe W, Mate K, Bennett B, Ngidi H, Reddy J, Barker P, et al. Improving public health information: a data quality intervention in KwaZulu-Natal, South Africa. Bull World Health Organ. 2012 Mar 1;90(3):176–82.

27. Gimbel S, Mwanza M, Nisingizwe MP, Michel C, Hirschhorn L, Hingora A, et al. Improving data quality across 3 sub-Saharan African countries using the Consolidated Framework for Implementation Research (CFIR): results from the African Health Initiative. BMC Health Serv Res. 2017 Dec 21;17(3):828.

28. Gloyd S, Wagenaar BH, Woelk GB, Kalibala S. Opportunities and challenges in conducting secondary analysis of HIV programmemes using data from routine health information systems and personal health information. J Int AIDS Soc. 2016;19(5 Suppl 4):20847.

29. Lemma S, Janson A, Persson L-Å, Wickremasinghe D, Källestål C. Improving quality and use of routine health information system data in low- and middle-income countries: A scoping review. PloS One. 2020;15(10):e0239683.

30. Nisingizwe MP, Iyer HS, Gashayija M, Hirschhorn LR, Amoroso C, Wilson R, et al. Toward utilisation of data for programme management and evaluation: quality assessment of five years of health management information system data in Rwanda. Glob Health Action. 2014 Nov 19;7:10.3402/gha.v7.25829.

31. Li M, Brodsky I, Geers E. Barriers to Use of Health Data in Low- and Middle-Income Countries: A Review of the Literature [Internet]. Chapel Hill, NC: MEASURE Evaluation; 2018 [cited 2021 Nov 8]. Available from: https://www.measureevaluation.org/resources/publications/wp-18-211.html

32. Agiraembabazi G, Ogwal J, Tashobya C, Kananura RM, Boerma T, Waiswa P. Can routine health facility data be used to monitor subnational coverage of maternal, newborn and child health services in Uganda? BMC Health Serv Res. 2021 Sep 13;21(Suppl 1):512.

33. Ashton RA, Prosnitz D, Andrada A, Herrera S, Yé Y. Evaluating malaria programmemes in moderate- and low-transmission settings: practical ways to generate robust evidence. Malar J. 2020 Feb 18;19(1):75.

34. Etamesor S, Ottih C, Salihu IN, Okpani AI. Data for decision making: using a dashboard to strengthen routine immunisation in Nigeria. BMJ Glob Health. 2018 Oct 1;3(5):e000807.

35. Adane A, Adege TM, Ahmed MM, Anteneh HA, Ayalew ES, Berhanu D, et al. Routine health management information system data in Ethiopia: consistency, trends, and challenges. Glob Health Action. 2021 Jan 1;14(1):1868961.

36. Lagarde M. How to do (or not to do) … Assessing the impact of a policy change with routine longitudinal data. Health Policy Plan. 2012 Jan 1;27(1):76–83.

37. Buuren S van, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. J Stat Softw. 2011 Dec 12;45:1–67.

38. Simmons EM, Singh K, Mpiima J, Kumar M, Weiss W. Assessing coverage of essential maternal and child health interventions using health-facility data in Uganda. Popul Health Metr. 2020 Oct

9;18(1):26.

39. Day C, Gray A, Cois A, Ndlovu N, Massyn N, Boerma T. Is South Africa closing the health gaps between districts? Monitoring progress towards universal health service coverage with routine facility data. BMC Health Serv Res. 2021 Sep 13;21(1):194.

40. Bennett A, Yukich J, Miller JM, Vounatsou P, Hamainza B, Ingwe MM, et al. A methodological framework for the improved use of routine health system data to evaluate national malaria control programmes: evidence from Zambia. Popul Health Metr. 2014 Nov 19;12(1):30.

41. Dossa NI, Philibert A, Dumont A. Using routine health data and intermittent community surveys to assess the impact of maternal and neonatal health interventions in low-income countries: A systematic review. Int J Gynaecol Obstet Off Organ Int Fed Gynaecol Obstet. 2016 Nov;135 Suppl 1:S64–71.

42. Saeed S, Moodie EEM, Strumpf EC, Klein MB. Evaluating the impact of health policies: using a difference-in-differences approach. Int J Public Health. 2019 May 1;64(4):637–42.

43. Steenland M, Robyn P, Compaore P, Kabore M, Tapsoba B, Zongo A, et al. Performance-based financing to increase utilisation of maternal health services: Evidence from Burkina Faso. SSM - Popul Health [Internet]. 2017 Oct 1 [cited 2021 Dec 20];3. Available from: https://pubmed.ncbi.nlm.nih.gov/29349214/

44. Hategeka C, Ruton H, Karamouzian M, Lynd LD, Law MR. Use of interrupted time series methods in the evaluation of health system quality improvement interventions: a methodological systematic review. BMJ Glob Health. 2020 Oct;5(10):e003567.

45. Wagner AK, Soumerai SB, Zhang F, Ross-Degnan D. Segmented regression analysis of interrupted time series studies in medication use research. J Clin Pharm Ther. 2002;27(4):299–309.

46. Penfold RB, Zhang F. Use of interrupted time series analysis in evaluating health care quality improvements. Acad Pediatr. 2013 Dec;13(6 Suppl).