

Considerations and Guidance for Using Routine and Program Monitoring Data for Social and Behavior Change Evaluation

Data collected routinely by governments and by program implementors can be leveraged to inform and evaluate social and behavior change (SBC) programs. What distinguishes routine data is that they are collected regularly within health information systems or within program monitoring systems. This brief is intended for global, regional, and country-level SBC program implementers, evaluators, and monitoring and evaluation teams who want to document whether their program is having an impact using routinely collected data. The brief provides an overview of the considerations of using routinely collected data for design and analysis, illustrates steps in undertaking an evaluation, and demonstrates how results can be applied to SBC programming.

Introduction

Understanding whether an intervention or activity is having its intended impact is considered program evaluation. Program evaluation can be defined as “a systematic method for collecting, analyzing, and using data to examine the effectiveness and efficiency of programs and, as importantly, to contribute to continuous program improvement.”¹ This brief will focus on the collection and analysis of routine and/or program data for evaluation in which the objective is to quantitatively assess the impact of a program or intervention, typically using statistical methods. To conduct program evaluation, data can come from primary or secondary sources.



KEY POINTS

The use of routine and program monitoring data for evaluation presents many opportunities for evaluating the impact of programs on priority health outcomes.

There are several advantages for using routine and program monitoring data for evaluation, including that they are collected more frequently, across a wider geography, over a longer period of time, and may cost less to acquire.

Routine and program monitoring data may not be under the direct control of evaluators and thus may not always be appropriate for evaluation. Careful reflection on the appropriateness of the routine data for program evaluation is needed.

Primary data are collected for purposes of informing and evaluating a specific program or intervention, with the evaluators determining the indicators of interest to be collected and their measurement (how variables are defined, among what populations and geographies, and how often). The indicators may include intermediate outcomes of interest (knowledge, attitudes, norms, and practices), as well as behaviors and health outcomes, but may also include a wide array of potential measures that may impact the accuracy and precision of the empirical results (sociodemographics, other programming, or investments). The collection of primary data may not be possible for all evaluations due to program or research timelines and resources available for evaluation.

Secondary data refers to existing data sources in which an evaluator may have little or no control of their structure and/or collection. Routine data are considered secondary data. Routine data are collected on a regular basis by national governments and ministries within their information systems and are aggregated from sub-national units, e.g., health facilities, wards, districts, and other health administrative units. These data may be complemented

by data collected routinely by an implementing program through its internal monitoring systems, e.g., collated from staff working in communities or program reports. Examples of such routinely collected data are noted in Table 1.

There are **several advantages** to using routine data over primary data for evaluation, including:

- Routine data are typically collected more frequently, across a wider geography, over a longer period of time.
- Routine data are less expensive for evaluators to obtain and potentially more rapidly available.
- As the data are typically not collected directly from individuals (human subjects) by the evaluators themselves, expedited review or exemptions by ethical review bodies may be possible and may facilitate data collection and analysis timelines.

However, as routine data may not be under the direct control of evaluators, they may not always be appropriate for evaluation. This brief outlines considerations for using routine data, and the design and analytical phases of program evaluation with routine data.

TABLE 1 EXAMPLES OF POTENTIALLY AVAILABLE ROUTINELY COLLECTED DATA BY HEALTH AREA AND SOURCE

Health area	Data source	Data examples	Recommended frequency
Family planning (FP)	Health facilities	Counts of individuals seeking FP services and adopting FP methods Number of FP methods distributed to a health facility	Monthly
	Program monitoring	Counts of individuals attending community mobilization activities who receive referrals for FP services	Monthly
Maternal, newborn and child health	Health facilities	Counts of pregnant women seeking antenatal care (ANC) services Counts of pregnant women delivering at facility Counts of immunized children	Monthly
	Program monitoring	Counts of pregnant women residing in households visited by program who receive referrals for ANC services Number of providers trained in respectful maternal and newborn care	Monthly
Malaria	Health facilities	Counts of pregnant women taking intermittent preventive treatment in pregnancy Counts of individuals seeking malaria diagnostic and treatment services Number of rapid diagnostic tests and artemisinin-based combination therapy administered	Monthly
	Program monitoring	Counts of individuals residing in households visited by program who sleep under insecticide treated nets (ITNs) Number of ITNs distributed during seasonal distribution campaigns	Monthly

Considerations for using routine data for evaluation

Before using routine data, evaluators need to ensure that these data are appropriate for the evaluation. Their appropriateness will be determined based on the specific research questions that have been developed, the resources available, and the evaluation timeline. Three considerations, and related questions to be answered, for using routine data for evaluation include:

- Usability:
 - ✓ Do the data include key indicators of interest that are expected to be changed in a meaningful way by the program, and can they be consistently measured over time?
 - ✓ Are the indicators available measured in a way that matches the program objectives, and is the measurement consistent over time?
 - ✓ Are the data collected at a level (e.g., individual, health facility, district) that is meaningful for the evaluation, and are there sufficient numbers of observations (across geographies and over time) at the level needed?
- Quality:
 - ✓ Are the data of sufficient quality to lead to valid and reliable results and conclusions, and does the quality of the data systematically vary over time or geographies?
 - ✓ Are the data sufficiently complete over time and across geographies to provide the necessary internal validity to draw conclusions about program impact?
- Accessibility:
 - ✓ Do the data require authorized approvals to obtain and use, and can the data be obtained and retrieved in time for the evaluation?
 - ✓ Does use of the data require review of results and conclusions by authorizing bodies who control the data?

If routine data are deemed to be appropriate for the evaluation, the next step is to develop the design of the evaluation.

Designing an evaluation with routine data

Steps in designing an evaluation with routine data

1. Collect key programmatic information and assess whether routine data are appropriate for evaluation
2. Develop an evaluation design
3. Develop a statistical analysis plan (SAP)
4. Establish data quality protocols
5. Create an integrated dataset for the evaluation
6. Conduct appropriate statistical analysis for routine data

STEP 1 Collect key programmatic information and assess whether routine data are appropriate for evaluation

To design a rigorous and effective evaluation, it is important to have a clear understanding of the program's:

- **Theory of change**, which explains the pathways through which the program is expected to have an impact. It also indicates why it is expected that the program will ultimately have an impact on the indicators that are measured in routine data systems.
- **Detailed implementation plans**, which allow you to decide if comparison (counterfactual) areas are available, the timing of program initiation, phases of scale-up, and the geographies where the program will be implemented.
- **Monitoring and tracking data** that can be useful for the evaluation. The program monitoring data may provide useful quantitative information about program coverage, intensity, and resources used to implement the program.

STEP 2 Develop an evaluation design

Evaluations can be experimental (includes randomized assignment of the program or parts of the program), quasi-experimental (has a non-randomized control group or

has pre-/post-program measurement) or non-experimental (no control group and a single post-program measurement). While randomized controlled trials are often considered the gold standard for evaluation, they are not always feasible for evaluating health interventions. Quasi-experimental designs that leverage routine data include, but are not limited to:

- Interrupted time series (ITS) that allows the evaluator to determine if there are shifts in the trend of an outcome indicator as a result of the program or a sudden interruption in programming (e.g., COVID-19). With ITS, the trend from the pre-program period is projected into the future and serves as the comparison (counterfactual) to the observed (actual) program trend.
- When program monitoring data are also available on reach (or exposure) and intensity of the program, a dose-response approach can be complementary to the ITS. Dose-response analyses help further assess the potential causal relationship between the program and outcomes.

Impact hypotheses: It is important in evaluation design to develop an a priori expectation, or hypothesis, as to the nature and magnitude of impact from the program, to distinguish it from an observed effect being found by chance, due to natural trends, or due to other factors unrelated to the program. This hypothesis may be informed by existing literature and/or the program’s theory of change. For instance, when conducting ITS, a hypothesis would be formulated regarding the magnitude and trend of the outcome if the program were truly effective. For example,

it may be expected that the program would cause an almost immediate change in the level of the indicator after initiation (Figure 1a) or, alternatively, a lag period with a more gradual change in the trend could be expected, as is illustrated in (Figure 1b). These hypotheses would inform the statistical evaluation of the impact.

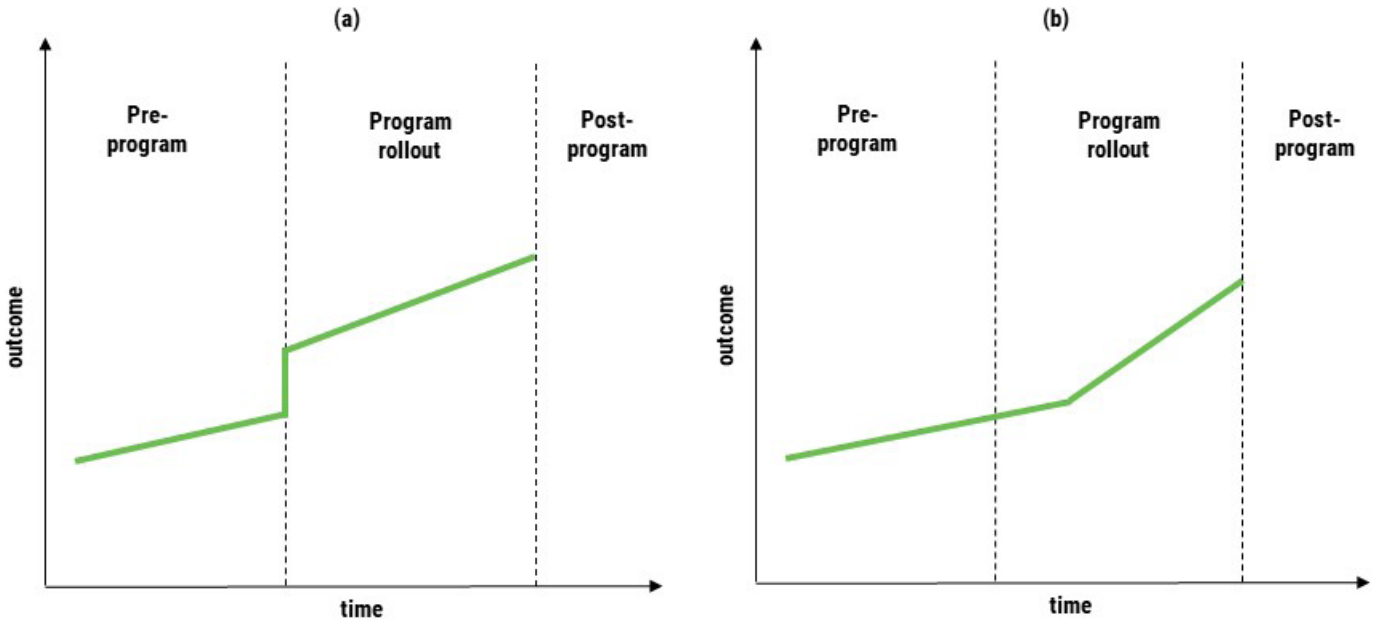
For **dose-response approaches**, the impact hypothesis should specify the relationship expected between the intensity of the program (quantified) and the outcome (quantified). The location, timing, and intensity of program implementation is often collected in the program monitoring and tracking data, and can be paired with other routine data. Depending on the nature of the program and monitoring data collected, the intensity may be measured dichotomously (existence of a program/no program) or continuously, by the degree of program implementation (e.g., number of households visited in a village per month).

STEP
3

Develop a statistical analysis plan (SAP)

A prospectively formulated SAP increases the transparency and credibility of findings regarding program impact. It represents good scientific practice in program evaluation. An SAP should contain enough details so that the analysis can be replicated by others, and any changes in initial hypotheses and methods should be documented. Components of an SAP typically include²⁻³:

FIGURE 1 HYPOTHETICAL PROGRAM IMPACT MODELS



- ✓ Administrative information: name of evaluation, current protocol version, SAP version, institutions, and principal investigators and roles.
- ✓ The evaluation background, rationale, and objectives.
- ✓ The evaluation methods: type of design, randomization (if applicable), statistical power considerations, null and alternative hypotheses, primary and secondary indicators, and timing of measurements.
- ✓ Statistical principles: significance levels, expected effect sizes, adjustments for multiplicity hypothesis testing.
- ✓ Study population: inclusion/exclusion criteria for data (e.g., based on quality), trends in indicators
- ✓ Outcome definitions, measurement of treatment, confounding covariates
- ✓ Statistical analysis methods.

STEP 4 Establish data quality protocols

Data quality protocols describe what metrics and benchmarks will be used to assess the quality of the data. Examples of metrics, such as those developed by the World Health Organization, center around four dimensions of data quality relevant for routine data⁴:

- Completeness and timeliness:
 - Completeness of reporting: percent of expected data available at [level] over [time period].
 - Timeliness of reporting: percent of expected data at [level] available on time.
 - Completeness of indicator data: percent of data elements that are non-zero values and % of data elements that are non-missing values.
- Internal consistency:
 - Outliers: percent of values that are above or below two standard deviations from the mean.
 - Consistency over time: graphic depiction of trends.
 - Consistency between related indicators: percent of cases where there are extreme differences between indicators that are expected to be roughly equal (e.g., first antenatal care visits and intermittent preventive therapy).
- External consistency: comparison of routine data with population-based survey values from similar period, if available.
- External consistency of population data used for denominators: for example, comparison of population

data used by program with official government statistics.

STEP 5 Create an integrated dataset for the evaluation

An evaluation using routine data will likely have to incorporate data from different sources for analysis. For example, routine data may be collected from health information systems, program monitoring data, and population data from censuses or other sources. Data from different sources should be carefully merged on matching identifiers, e.g., by time-period and geographical location and level. Documentation of the sources of data used to merge the data, their measurement, and how they were integrated is an important component of the SAP.

STEP 6 Conduct appropriate statistical analysis for routine data

- Data analysis should commence with descriptive analysis (data frequencies, geographic information system (GIS) mapping, dashboards, and visualizations). When using routine data, scatter plots are a useful tool to identify underlying trends, seasonal patterns, and outliers. GIS mapping by geographic area is a powerful tool for visualizing large-scale program implementation.
- Interrupted time-series analysis: ITS regressions, also known as segmented regressions, can be estimated on a single group, where the pre-intervention trend projected into the treatment period serves as the counterfactual. ITS regressions can also be estimated to compare intervention and control groups. When using time series data, it is important to consider the following methodological challenges:
 - Autocorrelation: consecutive observations tend to be more like one another than those further apart, which violates the assumption of standard regression models that observations are independent.
 - Seasonality: indicators can have a seasonal pattern which can lead to biased results and autocorrelation.
 - Time-varying confounders: ITS analysis^a is generally unaffected by typical confounding variables that change relatively slowly over time, such as population age distribution or socioeconomic status;

^aSee Lopez Bernal et al. (2017)⁵ for a general tutorial on the use of ITS regressions for evaluation.

however, it can be affected by confounders that change more rapidly such as weather events, civil unrest, or outbreaks of an infectious disease.

- Statistical software: Several statistical packages can be used to conduct ITS analysis. In a review of statistical methods used in ITS studies evaluating public health interventions, Turner et al. (2020) found that SAS and Stata are the most commonly used.⁶ The user-written “itsa” command in Stata performs ITS regressions for single and multiple group comparisons; it also allows the user to control for autocorrelation and estimate treatment effects over multiple periods.⁷

Limitations with analysis of routine data

There are important limitations associated with using routine data for program evaluation:

- Routine data are typically aggregated to administrative units and information on individuals is generally not available, so linking a program exposure directly to an outcome is not always possible. Prior to data collection, researchers may be able to work with health providers

and/or program implementers to enhance reporting systems so that in some circumstances, individual- or household level-data are available for analysis.

- Behavior change programs often want to know whether their interventions impact intermediate factors, as determined by the theory of change. However, routine data may not directly measure these intermediate outcomes (see box below). Therefore, proxies may be necessary to understand if the program is influencing behavior. Additionally, behaviors within the household are often not observed and thus data are not available (e.g., breastfeeding, use of insecticide treated bed net use, or dietary diversity).
- Researchers may need to control for potentially confounding factors that are not measured in routine data. Researchers may overcome this issue by identifying other secondary data sources to use to capture confounders. When important measured factors remain unavailable, researchers should consider how they may influence their findings and note them explicitly.
- Routine data are typically collected among only a subset of the population targeted by the program. To capture the proportion exposed to a program (a key indicator used in the assessment of impact), a

Feliciano Monti for PMI/Burma



Specific considerations for SBC program evaluation

While SBC programs ultimately want to change the health and behaviors of individuals, they often do so by more directly targeting immediate factors that are believed to influence behaviors, including knowledge, attitudes, beliefs, and norms. Often these factors are not measured in routine data, making program impact harder to assess for these outcomes. While some intermediate SBC indicators are captured in household surveys, e.g., the Demographic and Health Surveys,⁸ they may be limited by topic and too intermittently collected to be used to assess the impact of specific programs. Other surveys, such as the Malaria Behavior Survey,⁹ have greater coverage of SBC indicators. Additionally, as SBC programs may target certain populations or sub-national areas, household survey samples may not be representative of the populations or levels needed to observe sufficient change in indicators of interest. Advocacy efforts are underway for increasing the use of priority SBC indicators in program and routine data.¹⁰

denominator must be obtained from credible population data.

Conclusion

Evaluations of SBC programs using routinely collected health information and/or program data can be a useful tool for program evaluation. It is important to determine at the outset, when the evaluation and research questions are developed, whether routinely collected data are appropriate, specifically regarding the data's usability, quality, and accessibility. If the data are appropriate, such data has several advantages over primary data collection approaches. However, evaluations of routine data may not be feasible if health systems and/or program data are incomplete or of poor quality. Drawing together data from multiple sources into an integrated database may account for limitations and gaps in the data. To further improve upon the potential for using routinely collected data, it is preferable for evaluators to work with program implementers at the outset to ensure program data can be useful for program evaluation.¹¹ Donors, implementers, and researchers are currently working hand-in-hand with national ministries in improving data collected in national information systems.

References

1. U.S. Department of Health and Human Services Centers for Disease Control and Prevention. 2011. Office of the Director, Office of Strategy and Innovation. "Introduction to program evaluation for public health programs: A self-study guide." Atlanta, GA: Centers for Disease Control and Prevention.
2. Thomas, Laine, and Eric D. Peterson. 2012. "The value of statistical analysis plans in observational research defining high-quality research from the start," *Journal of the American Medical Association* 308(8): 773–774. doi: 10.1001/jama.2012.9502
3. Hiemstra, Bart et al. 2019. "DEBATE-statistical analysis plans for observational studies," *BMC Medical Research Methodology* 19: 233. doi: 10.1186/s12874-019-0879-5
4. World Health Organization. 2017. "Data quality review: a toolkit for facility data quality assessment—Module 1: Framework and metrics." Geneva: World Health Organization.
5. Lopez, Bernal James, Steven Cummins, and Antonio Gasparrini. 2017. "Interrupted time series regression for the evaluation of public health interventions: a tutorial," *International Journal of Epidemiology* 46(1): 348–355. doi: 10.1093/ije/dyw098
6. Turner, S.L. et al. 2020. "Design characteristics and statistical methods used in interrupted time series studies evaluating public health interventions: a review," *Journal of Clinical Epidemiology* 122: 1–11. doi: 10.1016/j.jclinepi.2020.02.006.
7. Linden, A. 2015. "Conducting interrupted time-series analysis for single- and multiple-group comparisons," *Stata Journal* 15: 480–500. doi: 10.1177/1536867X1501500208
8. ICF. "Demographic and Health Survey (DHS)" <https://dhsprogram.com/Methodology/Survey-Types/DHS.cfm>. [Accessed 1 September 2020].
9. Breakthrough ACTION. "Malaria behavior survey." <https://malaria-behaviorsurvey.org/> [Accessed 5 January 2020].
10. Breakthrough RESEARCH. 2020. "Twelve recommended SBC indicators for family planning," *Breakthrough RESEARCH Programmatic Research Brief*. Washington, DC: Population Council.
11. Centers for Disease Control and Prevention. 2019. "Program Performance and Evaluation Office," Webpage, 20 May 2019 [cited 8 October 2020]; Available from: <https://www.cdc.gov/eval/index.htm>.

Suggested citation

Breakthrough RESEARCH. 2021. "Considerations and guidance for using routine and program monitoring data for social behavior and change evaluation," *Programmatic Research Brief*. Washington, D.C.: Population Council.

Front cover photo credit: USAID

©2021 The Population Council. All rights reserved.

Email

BreakthroughResearch@popcouncil.org

Breakthrough RESEARCH | Population Council

4301 Connecticut Ave., NW, Suite 280 | Washington, DC 20008
+1 202 237 9400 | breakthroughactionandresearch.org



Breakthrough RESEARCH is made possible by the generous support of the American people through the United States Agency for International Development (USAID) under the terms of cooperative agreement no. AID-OAA-A-17-00018. The contents of this document are the sole responsibility of Breakthrough RESEARCH and Population Council and do not necessarily reflect the views of USAID or the United States Government.



Breakthrough RESEARCH catalyzes social and behavior change (SBC) by conducting state-of-the-art research and evaluation and promoting evidence-based solutions to improve health and development programs around the world. Breakthrough RESEARCH is a consortium led by the Population Council in partnership with Avenir Health, ideas42, Institute for Reproductive Health at Georgetown University, Population Reference Bureau, and Tulane University.