Glossary

ROSSI ET. AL. EVALUATION

GLOSSARY

Accessibility:

The extent to which the structural and organizational arrangements facilitate participation in the program.

Accountability:

The responsibility of program staff to provide evidence to stakeholders and sponsors that a program is effective and in conformity with its coverage, service, legal, and fiscal requirements.

Accounting perspectives:

Perspectives underlying decisions on which categories of goods and services to include as costs or benefits in an economic efficiency analysis. Common accounting perspectives are those that take the perspective of program participants, program sponsors and managers, and the community or society in which the program operates.

Administrative data system:

A data system that routinely collects and reports information about the delivery of services to clients and, often, billing, costs, diagnostic and demographic information, and outcome status.

Administrative standards:

Stipulated achievement levels set by program administrators or other responsible parties, for example, intake for 90% of the referrals within 1 month. These levels may be set on the basis

of past experience, the performance of comparable programs, or professional judgment.

Articulated program theory:

An explicitly stated version of program theory that is spelled out in some detail as part of a program's documentation and identity or as a result of efforts by the evaluator and stakeholders to formulate the theory.

Assessment of program process:

An evaluative study that answers questions about program operations, implementation, and service delivery. Also known as a process evaluation or an implementation assessment.

Assessment of program theory and design:

An evaluative study that answers questions about the conceptualization, design, and theory of action of a program.

Assignment variable:

In regression discontinuity designs, the quantitative variable that provides values for each unit in the study sample that are used to assign them to intervention or control conditions depending on whether they are above or below a predetermined cut-point value. Also called a *forcing variable* or *cutting-point variable*.

Attrition:

The loss of outcome data measured on individuals or other units assigned to comparison or intervention

groups, usually because those individuals cannot be located or refuse to contribute data.

Benefits:

Positive program effects, usually translated into monetary terms in cost-benefit analysis or compared with costs in cost-effectiveness analysis. Benefits may include both direct and indirect effects.

Bias:

As applied to program coverage, the extent to which subgroups of a target population are reached unequally by a program.

Black-box evaluation:

Evaluation of program outcomes without the benefit of an articulated program theory or relevant program process data to provide insight into what is presumed to be causing those outcomes and why.

Case studies:

An approach to evaluations that focuses on a program site or small number of sites in which the program participants and program context, service delivery and implementation, and outcomes are described.

Causal designs:

Randomized designs, regression discontinuity designs, and all the varieties of comparison group designs that are implemented in evaluations assessing program impact and which provide the estimates of the program effects on the outcomes of interest.

Cluster randomized trial:

A randomized control design for impact evaluation in which aggregate units, such as communities, schools, or clinics, are randomly assigned to intervention and control conditions, with outcomes measured on individuals within those aggregate units.

Comparison group:

A group of individuals or other units not exposed to the intervention, or not yet exposed, and used to estimate the counterfactual outcomes for a group that is exposed to the program. Comparison groups are used in designs in which exposure to the intervention is not controlled as part of the design, as is done in randomized control designs in which the comparison group is typically referred to as a control group.

Confirmation bias:

A cognitive bias in which individuals gather, interpret, or remember information selectively in a way that confirms their preexisting beliefs or hypotheses.

Control group:

A group of individuals or other units assigned in an impact evaluation to the condition that is not provided with access or exposure to the intervention; used to estimate the counterfactual outcomes for a group assigned to receive access to the intervention. Control groups are used in randomized control and regression discontinuity designs in which access to the intervention is controlled as part of the design. Compare with *comparison group*.

Cost analysis:

An itemized description of the full costs of a program, including the value of in-kind contributions, volunteer labor, donated materials, and the like.

Cost-benefit analysis:

An analytical procedure for determining the economic efficiency of a program, expressed as the relationship between costs and outcomes, with the outcomes usually measured in monetary terms.

Cost-effectiveness analysis:

An analytical procedure for determining the economic efficiency of a program, expressed as the cost for achieving one unit of an outcome, often used to compare efficiency across different programs.

Costs:

The monetary value of the inputs, both direct and indirect and both paid or in-kind, required to operate a program.

Counterfactual:

The hypothetical condition in which the individuals (or other relevant units) exposed to a program are at the same time, contrary to fact, not exposed to the program. Can also refer to the counterfactual outcomes: the outcomes that would occur for those individuals in that counterfactual condition.

Covariate:

In the context of impact evaluations, a preintervention baseline descriptive variable characterizing the study sample (intervention and

comparison groups) that can be used, among other things, to reduce bias in the intervention effect estimates that is associated with baseline differences between the groups.

Coverage:

The extent to which a program reaches its intended target population.

Demonstration program:

Social intervention projects designed and implemented explicitly to test the value of an innovative program concept.

Descriptive designs:

Evaluation research designs that describe, depending on the purpose of the evaluation, the program participants and program context, service delivery and implementation, and outcomes.

Direct instrumental use:

Actions undertaken to improve program operations or other program modification by decision makers and other stakeholders on the basis of specific ideas and findings from an evaluation.

Discounting:

The treatment of time in valuing costs and benefits of a program in efficiency analyses. It involves adjusting future costs and benefits to their present values and requires choice of a discount rate and time frame.

Distributional effects:

Effects of programs that result in a redistribution of resources among the target population.

Dose-response analysis:

Examination of the relationship between the amount or quality of program exposure and the program outcomes.

Effect size statistic:

A statistical formulation of an estimate of a program effect that expresses its magnitude in a standardized form comparable across outcome measures using different units or scales. Two of the most commonly used effect size statistics are the *standardized mean difference* and the *odds ratio*.

Effective sample size:

The operative sample size in statistical power analysis for multilevel impact evaluation designs with assignment at the cluster level and outcomes measured on units within those clusters. Similarity among individuals within clusters makes their outcome data partially redundant (statistically dependent). The effective sample size, which is smaller than the actual total sample size, adjusts for that redundancy.

Effectiveness evaluation:

An impact evaluation of a program that is implemented and operated as routine practice at typical scale and serving a typical target population, that is, not set up as a research or demonstration program. Compare with *efficacy evaluation*.

Efficacy evaluation:

An impact evaluation of a program that is implemented and operated as a research or demonstration program, typically for purposes of determining the ability of the program to produce the intended effects under relatively favorable conditions. The program may be administered and/or evaluated by the program developer. Also known as a proof-of-concept study. Compare with *effectiveness evaluation*.

Efficiency assessment:

An evaluative study that answers questions about program costs in comparison to either the monetary value of its benefits or its effectiveness for bringing about changes in the social conditions it addresses. See also *cost-benefit analysis* and *cost-effectiveness analysis*.

Empowerment evaluation:

A participatory or collaborative evaluation in which the evaluator's role includes consultation and facilitation directed toward the development of the capabilities of the participating stakeholders to conduct evaluations on their own, to use the results effectively for advocacy and change, and to have influence on a program that affects their lives.

Evaluability assessment:

Negotiation and investigation undertaken jointly by the evaluator, the evaluation sponsor, and possibly other stakeholders to determine whether a program meets the preconditions for evaluation and, if so,

how the evaluation should be designed to ensure maximum utility.

Evaluation influence:

The direct or indirect effect of evaluation on the attitudes and actions of stakeholders and decision makers.

Evaluation questions:

Questions developed by the evaluator, evaluation sponsor, and/or other stakeholders that define the issues the evaluation will investigate. Evaluation questions should be stated in terms that can be answered using methods available to the evaluator and in a way useful to stakeholders.

Evaluation sponsor:

The person, group, or organization that requests or requires an evaluation and provides the resources to conduct it.

Ex ante efficiency analysis:

An efficiency (cost-benefit or cost-effectiveness) analysis undertaken before program implementation, usually as part of program planning, to estimate net effects in relation to costs.

Ex post efficiency analysis:

An efficiency (cost-benefit or cost-effectiveness) analysis undertaken after a program's effects are known.

External validity:

The extent to which an estimate of a program effect derived from a subset of the program's target population also characterizes the effect for the full target population, that is, generalizes to that population.

Focus group:

A small panel of persons selected for their knowledge or perspective on a topic of interest that is convened to discuss the topic with the assistance of a facilitator. The discussion is used to identify important themes or to construct descriptive summaries of views and experiences on the focal topic.

Formative evaluation:

An evaluative study undertaken to furnish information that will guide program improvement.

Fundamental problem of causal inference:

The outcome when exposed to the causal factor and the outcome when not exposed cannot both be observed at the same time for the same individuals, but it is the difference between those outcomes that defines the causal effect. See also *potential outcomes* and *program effect*.

Impact:

See *program effect*.

Impact evaluation:

An evaluative study that answers questions about program impact on the outcomes or social conditions the program is intended to ameliorate; that is, the change in outcomes attributable to the program. Also known as an impact assessment.

Impact theory:

A causal theory describing cause-and-effect sequences in which certain program activities are the instigating causes and certain changes in the individuals or other units exposed to the program are the effects they are expected to produce.

Implementation failure:

A situation in which a program does not adequately perform the activities and functions specified in the program design that are assumed to be necessary for bringing about the intended benefits.

Implementation fidelity:

The extent to which the program adheres to the program theory and design and usually includes measures of the amount of service received by the participants and the quality with which those services are delivered.

Implicit program theory:

Assumptions and expectations about how a program brings about its intended effects that are inherent in a program's services and practices but have not been fully articulated and recorded.

Incidence:

The number of new cases of a particular problem, condition, or event that arise in a specified area during a specified period of time. Compare *prevalence*.

Independent evaluation:

An evaluation in which the evaluator has the primary responsibility for developing the evaluation plan,

conducting the evaluation, and disseminating the results but has no role in developing or operating the program.

Influence:

A defining characteristic of evaluations is that they are conducted to influence attitudes and actions. Evaluations can influence individual attitudes or actions, interpersonal behaviors, or collective actions.

Intent-to-treat (ITT) effects:

The program effect estimates that result from a comparison of the outcomes of the intervention and control groups as they were originally assigned to those conditions. The intervention group in intent-to-treat comparisons thus includes those assigned to the intervention who did not receive the intervention along with those who did. Similarly, the control group includes those assigned to the control who did receive the intervention along with those who did not. Compare with *treatment-on-the-treated* effects.

Interfering event:

In the context of time series designs, an event that occurs at about the same time as the initiation of the intervention with potential to affect the outcome and thus bias the estimate of the intervention effect on that outcome.

Internal rate of return:

The calculated value for the discount rate necessary for total discounted program benefits to equal total discounted program costs. See *discounting*.

Internal validity:

The extent to which the direction and magnitude of an estimate of a causal effect on an outcome, such as a program effect, are an accurate representation of the unknowable true effect. Internal validity for program effects is presumed to be high when complete outcome data are available for individuals exposed to the program and counterfactual outcomes are estimated with little or no bias.

Interrupted time series:

In impact evaluation, a set of repeated measures of the outcome that begins before the initiation of an intervention and continues afterward, with the intervention thus intruding into the time series in a way that may allow its effects on the outcome to be estimated.

Intervention group:

A group of individuals or other units that are exposed to an intervention and whose outcome measures are compared with those of a *comparison* or *control* group. See also *program group*.

Key informants:

Persons whose personal or professional position gives them a knowledgeable perspective on the nature and scope of a social problem or a target population and whose views are obtained via interviews or surveys.

Matching:

A procedure for constructing a comparison group by selecting individuals or other relevant units not exposed to the program that are identical on specified

characteristics to those in an intervention group except for receipt of the intervention.

Maturation:

Natural changes in the individuals or units involved in an impact evaluation of a sort expected to influence the outcomes of interest, for example, the increased abilities of children as they age.

Mediator variable:

In an impact assessment, a proximal outcome that changes as a result of exposure to the program and then, in turn, influences a more distal outcome. The mediator is thus an intervening variable that provides a link in the causal sequence through which the program brings about change in the distal outcome.

Meta-analysis:

An analysis of effect size statistics derived from the quantitative results of multiple intervention studies for the purpose of summarizing and comparing the findings of that set of studies.

Milestones:

Major tasks and the dates when they are expected to be accomplished throughout the course of an evaluation.

Minimum detectable effect size (MDES):

The smallest effect size determined by some appropriate assessment to have practical significance in the context of a particular program and a given outcome; specified in the form of a standardized statistical effect size. Impact evaluations should be

designed to have adequate statistical power for detecting at a statistically significant level any program effect as large as or larger than the MDES.

Moderator variable:

In an impact assessment, a variable, such as gender or age, that characterizes subgroups of the target population for which program effects may differ.

Monitoring and evaluation:

The practice of ongoing collection and reporting of data on program activities, products, and outcomes along with resource utilization and staffing for managing the program combined with outcome or impact evaluation at appropriate points in the life cycle of the program.

Needs assessment:

An evaluative study that answers questions about the social conditions a program is intended to address, the appropriate target population, and the nature of the need for the program.

Negative side effect:

An unintended adverse effect of a program intended to produce beneficial effects; may accompany beneficial effects.

Net benefits:

The total discounted benefits minus the total discounted costs. Also called net rate of return.

Odds ratio:

An effect size statistic that expresses the odds of a successful outcome for the intervention group relative to that of the control group.

Opportunity costs:

The monetary value of opportunities forgone because of involvement of some sort in an intervention program.

Organizational plan:

Assumptions and expectations about what the program must do to bring about the interactions between the target population and the program that will produce the intended changes in social conditions. The program's organizational plan is articulated from the perspective of program management and encompasses both the functions and activities the program is expected to perform and the human, financial, and physical resources required for that performance.

Outcome:

The state of the target population or the social conditions a program is expected to change.

Outcome change:

The difference between outcome levels at different points in time. See also *outcome level*.

Outcome level:

The status of an outcome at some point in time. See also *outcome*.

Outcome monitoring:

Periodic measurement and reporting of indicators of the status of the social condition or outcomes for program participants the program is accountable for improving.

Participatory or collaborative evaluation:

An evaluation organized as a team project in which the evaluator and representatives of one or more stakeholder groups work collaboratively in developing the evaluation plan, conducting the evaluation, and disseminating or using the results.

Performance criterion:

The standard against which an indicator of program performance is compared so that the program performance can be evaluated.

Policy significance:

The significance of an evaluation's findings for policy and program decisions or assumptions (as opposed to their statistical significance).

Policy space:

The set of policy alternatives that are within the bounds of acceptability to policymakers at a given point in time.

Population at risk:

The individuals or units in a specified area with characteristics indicating that they have a significant probability of having or developing a particular condition or experience. Compare *population in need*.

Population in need:

The individuals or units in a specified area that currently manifest a particular problematic condition or experience. Compare *population at risk*.

Potential outcomes:

An outcome status that would become manifest under certain conditions. The potential outcomes framework for causal inference defines the effect of a known cause as the difference between the potential outcome that would appear with exposure to the cause (e.g., a program) and the potential outcome that would appear without exposure to that cause (e.g., no exposure to the program).

Prevalence:

The total number of existing cases with a particular condition in a specified area at a specified time. Compare *incidence*.

Primary data:

Data collected during the course of an evaluation specifically to address the research questions set forth for the evaluation.

Primary dissemination:

Dissemination of the detailed findings of an evaluation to sponsors and technical audiences.

Probability sample:

A sample from a population in which every member of that population has a known, nonzero chance of being selected for the sample. This means that selection into the sample is done randomly so that it

is a matter of chance without any systematic bias in the selection process.

Process evaluation:

Examination of what a program is, the activities undertaken, who receives services or other benefits, the consistency with which it is implemented in terms of its design and across sites, and other such aspects of the nature and operation of the program.

Process monitoring:

Process evaluation that is done repeatedly over time with a focus on selected key performance indicators.

Process theory:

The combination of the program's *organizational plan* and its *service utilization plan* into an overall description of the assumptions and expectations about how the program is supposed to operate.

Program effect:

That portion of an outcome change that can be attributed uniquely to a program, that is, with the influence of other sources controlled or removed; also termed the program's impact. See also *outcome change*.

Program evaluation:

The application of social research methods to systematically investigate the effectiveness of social intervention programs in ways that are adapted to their political and organizational environments and are designed to inform social action to improve social conditions.

Program group:

A group of individuals or other units that receive a program and whose outcome measures are compared with those of a comparison or control group. See also *intervention group*.

Program impact:

See *program effect*.

Program monitoring:

The periodic measurement or documentation of aspects of program performance that are indicative of whether the program is functioning as intended or according to an appropriate standard.

Propensity score:

A score that estimates the probability that an individual or other relevant unit is in the intervention group rather than the comparison group that can be used in various ways to try to reduce selection bias. Propensity scores are constructed from preintervention baseline covariates in a separate analysis before the estimation of the intervention effect.

Quantitative assignment variable:

A variable with at least four unique values on each side of a cutoff that assigns units to an intervention such that all units on one side of a cutoff receive access to the intervention and no units on the other side of the cutoff receive access.

Quasi-experiment:

An impact evaluation design in which intervention and comparison groups are formed by a procedure other than random assignment.

Random assignment:

Assignment of the units in the study sample for an impact evaluation to intervention and control groups on the basis of chance so that every unit in that sample has a known, nonzero probability of being assigned to each group. Also called *randomization*.

Randomized control design:

An impact evaluation design in which intervention and control groups are formed by random assignment and compared on outcome measures to estimate the effects of the intervention. Also called randomized designs, randomized control trials (RCTs), and randomized experiments. See *random assignment*.

Rate:

The occurrence or existence of a particular condition expressed as a proportion of units in the relevant population (e.g., deaths per 1,000 adults).

Regression discontinuity design:

An impact evaluation design in which intervention and control groups are formed on the basis of their scores on a quantitative assignment variable; the units on one side of a cut-point value on the assignment variable are assigned to the intervention condition, and those on the other side are assigned to the control condition. Also known as a cutting-point design. See *assignment variable*.

Regression to the mean:

A phenomenon that can bias estimates of intervention effects in which, with repeated measurement, an extreme value will tend to be followed by a more typical less extreme one (i.e., regress to the mean of the series). Regression to the mean can also occur when individuals are chosen on the basis of a measured variable from the tail of the distribution of scores for the sample from which they are drawn; the value of a subsequent measure will tend to be less extreme, regressing to the mean of the parent distribution.

Reliability:

The extent to which a measure produces the same results when used repeatedly to measure something that has not changed.

Sample survey:

A survey administered to a sample of units in the population. The results are extrapolated to the entire population of interest by statistical projections.

Sampling error:

The chance component introduced into an outcome measure because of the luck of the draw that produced the particular sample from the universe of samples that could have been selected to provide that outcome data. The primary determinant of sampling error is the size of the sample; larger samples are less likely to differ from one another than smaller samples.

Sampling frame:

A list of the units in a population from which a sample is drawn, typically used for a probability sample.

Secondary data:

Data collected before an evaluation, often for administrative purposes, which can be analyzed to address the research questions set forth for the evaluation.

Secondary dissemination:

Dissemination of summarized often simplified findings of evaluations to audiences composed of stakeholders.

Secondary effects:

Effects of a program that impose costs on persons or groups who are not the intended beneficiaries of the program.

Secular trends:

Natural trends in a population of individuals or other units that can bias intervention effect estimates, especially in time series designs. Examples of secular trends are demographic changes in the population resident in a geographical area, changes in economic conditions, increases or decreases in the prevalence of a health condition, and the like.

Selection bias:

Systematic misestimation of program effects that results from uncontrolled differences between a group of individuals exposed to the program and a comparison group not exposed that would result in

differences in the outcome even if neither group was exposed to the program. See *counterfactual*.

Sensitivity:

The extent to which the values on a measure change when there is a change or difference in the thing being measured.

Service utilization plan:

Assumptions and expectations about how the target population will make initial contact with the program and be engaged with it through the completion of the intended services. In its simplest form, a service utilization plan describes the sequence of events through which the intended clients are expected to interact with the intended services.

Shadow prices:

Imputed or estimated costs of goods and services not valued accurately in the marketplace. Shadow prices also are used when market prices are inappropriate because of regulation or externalities. Also known as *accounting prices*.

Snowball sampling:

A nonprobability sampling method in which each person who participates in an initial sample is asked to suggest additional people appropriate for the sample, who are then asked to make further suggestions. This process continues until no new names of appropriate persons are suggested.

Social indicator:

A series of periodic measurements designed to track the course of a social condition over time.

Social research methods:

Social science techniques of systematic observation, measurement, sampling, research design, data collection, and data analysis for producing valid, reliable, and precise characterizations of social behavior.

Stakeholders:

Individuals, groups, or organizations with a significant interest in how well a program functions, for example, those with decision-making authority over the program, funders and sponsors, administrators and personnel, and clients or intended beneficiaries.

Standardized mean difference:

A standardized effect size statistic that expresses the difference between the means for the intervention and control groups on an outcome variable in standard deviation units.

Standards:

The level of performance a program is expected to achieve to be judged adequate.

Statistical power:

The probability that an observed program effect will be statistically significant when, in fact, it represents a real effect. If a real effect is not found to be statistically significant, a Type II error results. Thus, statistical power is one minus the probability of a Type II error. See also *Type II error*.

Summative evaluation:

Evaluative activities undertaken to render a summary judgment on certain critical aspects of a program's performance, for example, to determine if specific goals and objectives were met.

Target population:

The population of units (individuals, families, communities, etc.) to which a program intervention is directed. All such units within the area served by a program constitute its target population.

Targeted program:

A program with a target population defined around specific characteristics or eligibility requirements that constrain who can receive services. Those constraints may relate to current conditions (e.g., low income, diagnosed mental illness) or indicated risk for an adverse outcome the program aims to prevent. Compare *universal program*.

Theory failure:

A situation in which a program is implemented as planned, but it does not produce the expected effects on the outcomes or the social benefits intended.

Treatment-on-the-treated (TOT) effects:

The program effect estimates that result from a comparison of the outcomes of the units that received the intervention and those that did not receive the intervention, irrespective of the condition to which they were originally assigned. The intervention group in treatment-on-the-treated comparisons thus includes only those who actually

received the intervention, and the control group includes only those who in fact did not receive the intervention. Compare with *intent-to-treat (ITT) effects*.

Type I error:

A statistical conclusion error in which an effect estimate is found to be statistically significant when, in fact, there was no actual effect on the respective outcome variable.

Type II error:

A statistical conclusion error in which an effect estimate is not found to be statistically significant when, in fact, there was an effect on the respective outcome variable. See *minimal detectable effect size*, *statistical power*.

Universal program:

A program with a target population that is defined with few or no constraints (e.g., programs in public parks open to all who wish to participate, afterschool programs that accept any child in the school district parents wish to enroll). Compare *targeted program*.

Validity:

When used to describe a measure, the extent to which it actually measures what it is intended to measure.