

Part II.G
Operations
Research

14 Process indicators

11 Impact indicators

6 Contextual and other factors

What is Operations Research?

Operations Research (OR) is an approach used to improve service delivery or to strengthen other aspects of programs. Although OR can include diagnostic or evaluative studies, the most common use of OR is the intervention study, consisting of five steps:

1. Identifying problems related to service delivery;
2. Identifying possible strategies to address these problems;
3. Testing these strategies under quasi-experimental conditions;
4. Disseminating the findings to program managers and policymakers; and
5. Using the information to improve service delivery programs (Fisher et al., 1991).

This approach is particularly useful in testing new and potentially controversial strategies to service delivery. The implementing organization can experiment with the new approach on a limited scale, without having to adopt it throughout the organization. If the strategy (intervention) proves ineffective or creates unwanted political backlash, then the organization can decide to discontinue it and pursue alternative approaches, at relatively little political cost. If the intervention proves effective and acceptable to the population in question, then the organization can use these results to justify the adoption/expansion of the intervention within the organization. Moreover, the results of a successful OR project may prompt other organizations to adopt the same intervention in their own programs.

The evaluation of OR should address both process and impact. Until recently, there has been relatively little evaluation of operations research, in part because most OR projects are designed to evaluate an intervention. Should one then “evaluate an evaluation?” To the extent evaluation of OR projects occurred, it tended to measure outputs (e.g., how many studies were conducted, how many reports were distributed). With a

few notable exceptions (Solo et al., 1998), there was little systematic assessment of impact: the extent to which the OR study resulted in changes in service delivery procedures or policy.

In 1992-93, an OR Working Group, convened under The EVALUATION Project, proposed a set of indicators to evaluate OR studies (later published in Bertrand and Brown, 1997). This work paved the way for the development of a more complete set of indicators under the FRONTIERS Program; evaluators tested these indicators in various countries between 1999 – 2001. The new set of indicators, presented here, measures both how well a study is carried out (“process”) and the extent to which a study results in changes in service delivery procedures or policy (referred to in this section as “impact”). In addition, the set includes indicators of context, which describe factors that facilitate or hinder the conduct of OR and the utilization of results; they are useful in explaining what has (or has not) happened, but – in contrast to the indicators of process or impact – they are not scored.

To those interested in more systematically tracking “what happens” as a result of OR studies, this list of indicators should prove beneficial. For others, the exercise may seem too academic and the list of indicators too extensive. Whereas we present the full list of indicators developed to evaluate OR projects, we encourage users to select a subset of these indicators most relevant to their own needs (e.g., relevance, technical soundness, credibility of the findings, and demonstration of change).

Methodological Challenges of Evaluating Operations Research

- **“Impact” is generally defined as change attributable to the project, but OR is generally only one of many influences in decision-making.**

Operations research has been a great catalyst in the field of family planning, and it is now playing an increas-

ingly large role in the area of HIV/AIDS, safe motherhood, postabortion care, prevention of FGC, and related areas of reproductive health. Notwithstanding, an OR study alone rarely results in a major change in service delivery or policy, and demonstrating cause and effect is virtually impossible when evaluating the impact of an OR study on the service delivery environment.¹ Other agencies are often involved in the provision of technical assistance essential to the successful implementation of the intervention. Moreover, a shifting political climate (positive or negative) can influence the utilization of OR results. A more realistic goal is to attempt to demonstrate plausible attribution. To meet the conditions of plausible attribution, the change in service delivery or policy must:

- Be instigated by persons familiar with the OR results;
 - Take place after the OR study; and
 - Be consistent with the results and recommendations of the OR study.
- **Decision-making is a complex and not necessarily rational process.**

Studies on the role of research findings in decision-making have shown that many other competing factors influence decision-making (Trostle, Bronfman, and Langer, 1999; Anderson et al., 1999; Iskandar and Indrawati, 1996). Program managers and other key decision-makers will only consider implementing recommendations from research they consider to be of high quality, conducted by reputable researchers, consistent with organizational values and needs as well as with social and political context, and able to provide an adequate solution to a recognized problem with available resources. Other less concrete factors such as personal relationships with researchers (Trostle, Bronfman and Langer, 1999) or job security also affect decisions. Context is not easy to measure; yet evaluators must consider it because of its important role in the translation of research recommendations into program and policy change.

- **The term “policy change” covers a large range of actions that differ substantially in their potential impact.**

Policy includes formal government declarations, laws, and statutes which those working in policy refer to as “Policy with a capital P.” In addition, policy can refer

to the regulations, guidelines, norms, and standards of a given organization (which some label as “policy with a small p”). Within the same country, policies can be enacted at different levels of the program and by different processes. Decisions that are voted into law by parliamentarians or other elected officials or made by the executive branch (e.g., such as the conditions under which induced abortion is legal in a given country) will subsequently affect the entire nation. The policies of a given organization may or may not have widespread ramifications, depending on the size and importance of that organization. For example, a decision by the MOH to introduce adolescent reproductive health services nation-wide should have substantial consequences for the service delivery environment. By contrast, a change in operational policies at a small NGO would have much less potential impact. Because operations research may be undertaken with organizations of varying sizes, from national ministries of health to small, local NGOs, the methodological challenge for evaluating the effects of OR projects on policy is to establish a working definition of the type of policy that will be considered relevant in making this judgment. One possible criterion for defining a “policy change” is that the change in regulations, guidelines, norms, or standards be implemented system-wide within the organization conducting the research (e.g., throughout all the service delivery points operated by the organization).

¹ To clarify, it is possible – with an experimental design – to demonstrate that an intervention caused a certain change in behavior among the intended audience. However, the difficulty lies in determining that a specific OR project was uniquely responsible for a change in service delivery procedures or in policy. Often other events are taking place at the same time that influence the decision-making processes. For example, in Guatemala, different organizations conducted OR studies on the acceptability of DepoProvera distributed through different channels (e.g., community-based distribution and clinics). Subsequently, the MOH adopted Depo as part of its contraceptive method mix. Whereas the studies most likely contributed to the inclusion of Depo in the MOH method mix, one can not rule out the possibility that the MOH might have decided to provide Depo to its clients for other reasons as well (e.g., the approval of Depo by the Food Drug Administration (FDA) in the United States, reports of its popularity in neighboring countries, the willingness of international donor agencies to supply the method). For this reason, we state above that “demonstrating cause and effect when evaluating an OR study is virtually impossible.”

- **Usually evaluators cannot measure impact until two to three years after the intervention is completed; however, in the course of this delay, other factors may intervene.**

While there is no golden rule for how long to wait to evaluate the impact of an OR study, at least two or three years are usually needed to allow adequate time for an organization to adopt and institutionalize changes based on the research. An alternative is to wait for all OR studies in a given program to end, and then evaluate them as a group. However, a time lapse of much more than three years may allow too many other changes to take place that might further complicate an evaluation. Due to high staff turnover, evaluators may find it difficult or impossible to contact and interview important informants. Many contextual changes may occur in this time and may further complicate the question of attribution.

- **The responses of key informants are by definition subjective.**

The indicators presented in this section rely on three primary data sources: key informant interviews, project documents, and site visits to observe innovations adopted as a result of an OR project. While key informants attempt to be objective, by definition their answers come from their own perspective. To minimize the bias of subjectivity, the evaluator should interview several individuals regarding the given study to increase the credibility to the information. Where disagreement occurs, evaluators may seek more information from other sources, but ultimately must use their best judgment, because they lack a systematic way to “weight” the opinions of two key informants.

- **The checklist of indicators does not adequately measure or reflect the importance of the dissemination of results.**

Operations research is conducted with a purpose: to use the results to improve programs. Thus, a necessary (though not sufficient) condition is that appropriate audiences learn about the results and use them in designing their own programs. For example, in 1997, a small Guatemalan NGO tested the use of a necklace to help Mayan couples correctly practice the rhythm method. Use failure rates were low, and most couples (who had wanted to use rhythm) found the method very satisfactory. The researchers widely disseminated the results

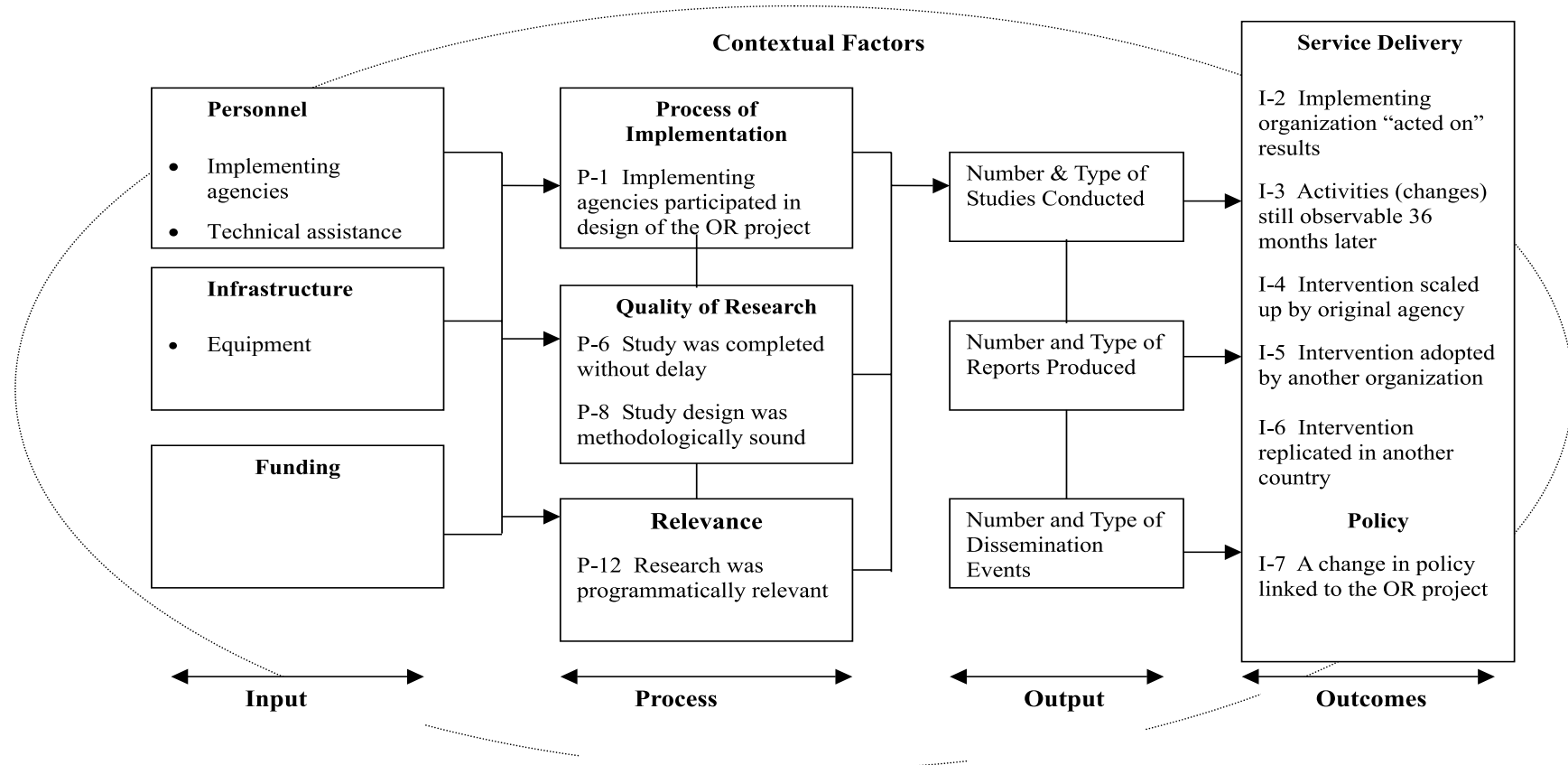
of this study to other groups working with both Mayans and non-Mayan (or ladino) populations, with the result that the Ministry of Health has subsequently included this method in the range of contraceptives it provides. In contrast to this example of successful dissemination of OR results, many studies in the past have not reached the decision-makers in a position to use the results, with the result that they had minimal impact on the service delivery system.

In short, dissemination is a crucial part of the OR process. Without effective dissemination, OR cannot influence service delivery or policy as it was designed to. However, finding the **right** indicators to measure dissemination has proven elusive for several reasons. First, no standard format for dissemination applies to all situations. Although the end-of-project dissemination seminar is now fairly standard, there are multiple other channels, including face-to-face presentations to high-level officials, technical assistance to the implementing agency, and presentations at international conferences (that lend credibility to the results in the local context). Second, the number of persons they must reach to ensure effective dissemination is not fixed. In some cases, reaching a single individual in a key decision-making role may be sufficient to launch an idea within an organization. In others, reaching a large number of persons through multiple channels over time is necessary before the message penetrates. A recent OR evaluation in Guatemala tested three additional dissemination questions, but none emerged as entirely satisfactory. Thus, we wish to emphasize the importance of dissemination and underscore the challenge that faces evaluators in improving the indicators to measure it.

The Operations Research Conceptual Framework

The conceptual framework used to develop the OR indicators in this section is presented in Figure II.G.1. The framework illustrates that events occurring as part of the process of conducting the OR can affect the utilization of results. In addition, factors beyond the control of the OR project (e.g., staff turnover, changing economic conditions in the country) also determine the extent to which OR findings will be translated to changes in service delivery or policy. Figure II.G.1 includes illustrative indicators of process and impact, on which the evaluation methodology is based. The full list of indicators appears in Box II.G.1

Figure II.G.1 Conceptual Framework for the Impact of Operations Research (with illustrative indicators)



Definition

This set of 25 indicators serves to collectively evaluate OR studies in terms of (a) the process of conducting the study, and (b) its impact (i.e., utilizing the study results to change service delivery or to influence policy).

An additional six indicators measuring context and other factors are listed in Box II.G.1, but they provide background only; evaluators do not score them in measuring performance of the OR team.

Data Requirements

Assessment of an external evaluator based on available information. Evaluators score each of the 25 items on a scale of 1 (not at all) to 3 (very much so). If indicator I-1 is negative (the intervention was not effective), then I-3, I-4, and I-5 are non-applicable. (See example of the data collection form in Appendix F.)

Data Source(s)

Project documents, in particular the final report of the project

Interviews with key informants, including researchers (especially the Lead Investigator), program managers, and other providers in the service delivery organizations who stand to benefit from the OR, donor agency staff, policymakers and other key decision-makers.

Purpose and Issues

The set of OR indicators allows the evaluator to arrive at a set of numerical scores supported by qualitative justifications for the scores for each OR study under review. The 3-point scale for each item distinguishes among those studies that performed well (3), those that performed satisfactorily but with notable problems (2), and those that did not perform satisfactorily on the relevant indicator (1).

Although other formats are possible, the OR indicators developed under the USAID-funded FRONTIERS Program use a grid format, which doubles as a data collection tool and a reporting format. (See Appendix F) Evaluators can use the blank grid as an interview guide to ensure consistency among key informants, evaluators, and projects. Evaluators can present results for each project using the same grid format. In addition, they can summarize the numerical scores of multiple projects in a table so that one can easily compare performance of studies overall (comparing columns) or can compare specific indicators across studies (comparing rows). This enables evaluators to identify areas of consistent strength and those requiring improvement. (For an example, see Bertrand and Marin, 2001.)

The OR indicators fall into three categories: process, impact, and context/other. Process indicators relate to the conduct of the study; evaluators can assess them immediately upon completion of the study. By contrast, they should assess impact indicators three years later, although impact occurring sooner certainly “counts.” In addition, the instrument contains six contextual and other indicators that provide insight into the process but are not included in the scoring because they do not reflect “performance.”

While each indicator has a score, the set of indicators does not lend itself to a summary score. Indicators measure different aspects of process and impact that are not necessarily of equal importance, nor is there sufficient experience in evaluating OR to reliably weight them. Rather, evaluators can compare scores individually or in groups of indicators that measure similar aspects, such as participation of the implementing agency at various stages of the study or conduct of subsequent research.

Box II.G.1 Indicators for Evaluating OR Projects

Process indicators (note: each indicator in this set begins with a “P”)

P-1 The implementing/collaborating organization(s) actively participated in the design of the OR project

The design of the OR project is the formulation of the study, which includes identifying the problem, establishing the objectives, designing the intervention, and selecting a research methodology. “Active” participation involves contributing original ideas to the work, not simply attending meetings.

P-2 The implementing/collaborating organization(s) actively participated in the implementation of the OR study

“Active participation” indicates that the organization was involved in decision-making and played a technical role in the implementation of the study, for example hiring new staff, conducting training, or analyzing and interpreting results.

P-3 The implementing/collaborating organization(s) participated in developing programmatic recommendations

This indicator asks whether these organizations participated, as well as how, for example, collaboration in report preparation, through formal meetings, and in working groups at dissemination conferences.

P-4 The study accomplished its research objectives

Each study is designed with one or more objectives. This indicator determines whether the study achieved each of its objectives.

P-5 The intervention was implemented as planned (or with some modifications)

Changes between the proposal and implementation of the intervention frequently occur and often are for the better. This indicator seeks to determine whether the organization carried out all of the activities specified in the intervention, allowing for some change in response to local realities. If not, the reviewer should identify any changes between the design and actual realization of these activities.

This indicator is not intended to penalize an organization for making modifications. Rather, it ascertains that the organization made some meaningful change in service delivery (that there was “something to evaluate”). An intervention study fails to show any change in the desired outcome for two plausible reasons: (1) the organization never implemented the intervention or implemented it so weakly that the study hardly constituted a fair test of its potential effectiveness, or (2) the organization fully implemented the intervention but it failed to show the expected results. This indicator attempts to eliminate the first possibility by determining that the intervention was in fact implemented.

P-6 The researcher(s) completed the study without delays (or other adjustments to the timeline) that would compromise the validity of the research design

Study activities are often delayed. This indicator seeks to identify delays that affected the timing of the intervention or that could have reduced the effectiveness of certain activities (e.g., a delay in training resulted

in diluting the effects of the activity; the period between intervention and final data collection had to be cut short, and thus the desired change had insufficient time to take place).

P-7 Key personnel remained constant over the life of the OR project

“Key personnel” are any personnel with a decision-making role in the design or implementation of the sub-project. Such personnel include the Principal Investigator, the study coordinator, and counterparts in the collaborating agencies, including key service personnel or government officials actively participating in implementation.

P-8 The study design was methodologically sound (free of flaws that could have affected the final results)

Evaluators should assess this item based on the methodology section of the report and (if appropriate) on discussions with the researchers. Generally, the external evaluator (not a staff member of any of the participating organizations) makes an “informed decision” on this point; key informants may have less knowledge or experience to make this judgment.

P-9 The research design was feasible in the local context

“Feasible” here means “reasonable” or “manageable,” a design that could be repeated without unduly draining financial or human resources. “Local context” includes not only program-related factors but also socio-cultural or political factors, among others.

P-10 The implementing/collaborating organization(s) judged the OR technical assistance to be useful and provided in a collegial manner

To qualify for a full score, both elements must be positive. If, for example, the advice was technically sound, but counterparts reacted negatively to the manner in which the OR team provided assistance (e.g., in an offensive or condescending way, “imposed upon them”), then the study should receive a lower score on this indicator.

P-11 Stakeholders judge results of the OR study to be credible/valid in the local context

This indicator refers to the judgment of stakeholders (policymakers, researchers, donors, program managers). Utilization of results would be likely limited if stakeholders seriously questioned the validity of the results.

P-12 Research was programmatically relevant

The perceptions of the same stakeholders listed above determine relevance. Relevant research addresses a priority problem of the program, whether a national program of the MOH or a more local program of an NGO.

P-13 Results were disseminated to key audiences, including policymakers, program managers, service providers, and donors

All studies involve dissemination of results. This indicator seeks to determine whether the dissemination strategies used effectively reached the target audience. “Key audiences” are those in a position to act on the results (e.g., policymakers, key decision-makers or service providers in implementing/collaborating agencies, donor agency staff). In addition, dissemination efforts may reach other interested parties (e.g., students

at the local university, members of the international RH community), but the indicator refers only to those in a position to act upon the results.

P-14 Results are readily available in written form

This indicator verifies the existence of a document on the key findings of the study that is well presented (of professional quality) and is locally available in sufficient quantity. This document may appear in a variety of media (e.g. website, CD-ROM) in addition to print. Ideally, results should be available in various formats appropriate to the intended audience: final reports and journal articles for donors and the academic RH community, summaries or research briefs for decisionmakers and program managers.

Impact indicators (note: each indicator in this set begins with an “I”)

I-1 The results indicate that the intervention was effective (i.e., that it improved service delivery in the areas identified by the study)

OR studies generally either test one or more interventions or they evaluate changes resulting from interventions already implemented. If all studies found the intervention under study to be effective, then research would be unnecessary. This indicator asks whether the intervention tested successfully improved front-line service delivery (e.g., increase in utilization of services, improved quality of services). Negative results can also be instructive, but they would not influence service delivery except to discontinue an ineffective strategy (see I-2).

I-2 The implementing/collaborating organization(s) “acted on” the results

“Acting on the results” consists of implementing the actual services of the intervention or the activities to support those services (e.g., training courses, development of service delivery guidelines, changes in allocation of personnel, production and testing of IEC materials, supervision, monitoring) if the intervention was effective, or not implementing or discontinuing these services and activities if the intervention was ineffective.

I-3 (If the intervention was effective and continued after the study) The activities tested under the intervention were still observable 36 months post-implementation

“Activities tested under the intervention” are those specific items mentioned in connection with the previous indicator. Where only some of the original activities are observable, the study should receive only a partial score on this indicator. In the case of an improvement that has lasted fewer than 36 months, this indicator does not apply.

I-4 (If the intervention was effective and continued after the study) The original implementing/collaborating organization scaled up the intervention in the same country

Most OR studies are conducted in a specific geographical area. “Scaling up” refers to implementing the intervention activities in additional geographical areas. It can but does not necessarily refer to expansion to the national level.

I-5 (If the intervention was effective and continued after the study) Another organization within the same country adopted the intervention

An organization that did not participate in the OR study “adopts” the intervention by implementing its primary components (see I-1).

I-6 Another country replicated the intervention

Some evidence must exist that links the original intervention to the activities carried out in the other country (e.g., program managers from other countries visited the project site and subsequently adopted similar strategies).

I-7 A change in policy can be linked to the OR project

This indicator measures legislation or other official changes that potentially affect service delivery, for example, authorization for the sale of the pill by non-medical personnel. For further discussion of policy changes, see the introduction to this section.

I-8 The implementing/collaborating organization conducted subsequent OR studies

“Subsequent OR studies” refers specifically to research activities that test interventions. OR studies do NOT include actions or program activities, such as training and materials production (described in I-2) or research for other purposes (e.g., the DHS, epidemiological research).

I-9 The implementing/collaborating organization conducted subsequent OR studies without external technical assistance (TA)

This indicator is included to reflect whether the organization has sufficient capacity to conduct these types of activities as a result of the previous OR experience and has the opportunity to do so.

I-10 The original donor funded new program activities based on the results of the OR study

New program activities are those activities tested in the intervention that the donor had not already funded.

I-11 Other donors provided new or expanded funding based on results of the OR studies

“Other donors” are those donor agencies that did not contribute financial support to the original OR project but subsequently funded the initiation or expansion of program activities – specifically, service delivery or support activities, including training, production of IEC materials, construction or renovations of facilities, and purchase of supplies and equipment. The indicator does not include funding of additional research only.

Context and other factors (note: each item begins with a “C”)

C-1 Other factors (not mentioned above) that facilitated the conduct of the research project

Situations that helped complete the activity might be: strong research capacity in the counterpart organization, powerful local person or donor intent on getting answers, good relationship between researchers and program staff, or others.

C-2 Other factors (not mentioned above) that facilitated the utilization of results

This indicator refers to situations that encourage the translation of the results into programmatic actions at the field level. For example, dissemination of results coincides with the planning cycle for a new program initiative or strategy; the intervention is a good match with the organization; and a committed individual continues to provide TA or to promote the intervention beyond the project end.

C-3 Other factors (not mentioned above) that hindered the conduct of the research project

Events beyond the control of the researchers and their collaborators in local organizations may impede a study's implementation. Such factors range from contraceptive stockouts, inter-organizational or interpersonal conflicts, or financial difficulties to political changes, civil unrest, or natural disasters.

C-4 Other factors (not mentioned above) that hindered the utilization of results

Situations that limit the incorporation of research results into policy or programs may also be beyond the control of researchers and their collaborators. Some examples are the intended population opposes the intervention (e.g., believes that contraceptives cause sterility); local authorities for political reasons veto a proposed initiative or change in service delivery; or the health system is restructured.

C-5 The donor used the data from the OR study for a specific purpose

Does any evidence exist that the donor used the data for its own purposes? Donors may use results for resource allocation, funding decisions, or miscellaneous strategic planning, among other possibilities.

C-6 The study included an assessment of costs of the intervention

Evaluators should mention any data collected on the cost of the intervention, primarily for the purpose of cost-effectiveness analysis. This indicator serves informational purposes only, since all OR studies do not necessarily need an assessment of cost.