# Evidence-Based Public Health: Moving Beyond Randomized Trials

Cesar G. Victora, MD, PhD, Jean-Pierre Habicht, MD, PhD, and Jennifer Bryce, EdD

Randomized controlled trials (RCTs) are essential for evaluating the efficacy of clinical interventions, where the causal chain between the agent and the outcome is relatively short and simple and where results may be safely extrapolated to other settings.

However, causal chains in public health interventions are complex, making RCT results subject to effect modification in different populations. Both the internal and external validity of RCT findings can be greatly enhanced by observational studies using adequacy or plausibility designs. For evaluating large-scale interventions, studies with plausibility designs are often the only feasible option and may provide valid evidence of impact.

There is an urgent need to develop evaluation standards and protocols for use in circumstances where RCTs are not appropriate. (*Am J Public Health.* 2004;94:400–405)

Public health has moved forward in recent years to improve the scientific standards for evidence underlying interventions and actions. “Evidence-based public health”[1] calls for a solid knowledge base on disease frequency and distribution, on the determinants and consequences of disease, and on the safety, efficacy, and effectiveness of interventions and their costs. The efficacy of an intervention is defined as its effect under “ideal conditions.”[2] The effectiveness of an intervention is defined as its effect under normal conditions in field settings. In this report, we question common assumptions about the types of evidence needed to demonstrate the efficacy and effectiveness of public health interventions and suggest that the guidelines for such evidence be updated.

Designs for large-scale impact evaluations of health and nutrition interventions are often based on the principles that have guided “gold standard” trials of new medicines and preventive agents in the past.[3,4] Over time, more and more medical scientists turned to randomized controlled trials (RCTs) in an effort to increase the internal validity of their designs. More recently, this increased attention to quality standards in clinical research has led to the Movement for Evidence-Based Medicine[5] and the establishment of the Cochrane Collaboration,[6] resulting in important improvements in methods and the quality of available evidence.

The success of these efforts encouraged the extension of RCT designs to the fields of public health and health policy.[7,8] The Cochrane Collaboration now includes meta-analyses of many public health topics,[6] and the on-line *Journal of Evidence-based Healthcare* has recently been established to provide an outlet for work in this area.[9] RCTs have increasingly been promoted for the evaluation of public health interventions.

In an earlier report,[10] 2 of the authors (C.G.V. and J.-P.H.) described 3 types of scientific inference that are often used for making policy decisions in the fields of health and nutrition. *Probability* statements are based strictly on RCT results. *Plausibility* statements are derived from evaluations that, despite not being randomized, are aimed at making causal statements using observational designs with a comparison group. *Adequacy* statements result from demonstrations that trends in process indicators, impact indicators, or both show substantial progress, suggesting that the intervention is having an important effect.

Although the evaluation literature has dealt with nonexperimental or quasi-experimental designs for several decades,[11] most examples of these methods arise from fields such as education, law enforcement, and economics. We are unaware of a systematic discussion of their application to public health.

In this article, we argue that the probability approach, and specifically RCTs, are often inappropriate for the scientific assessment of the performance and impact of large-scale interventions. Although evidence-based public health is both possible and desirable, it must go well beyond RCTs. We describe the limitations of using RCTs alone as a source of data on the performance of public health interventions and suggest complementary and alternative approaches that will yield valid and generalizable evidence.

## INTERNAL VALIDITY: RCTS AND BEYOND

The strength of the scientific inference supported by a study depends on its internal validity. Traditional epidemiological thinking commonly assumes that a methodologically sound design is sufficient to maximize internal validity. The 3 objectives of a sound design are to minimize selection and information bias, to control confounding, and to attempt to rule out chance.[2] RCTs are believed to be successful in achieving these objectives and are thus considered the gold standard of design validity. There are clearly defined standards for conducting and reporting on RCTs, all intended to increase the validity of their results and interpretation.[2,5]

An additional assumption, less often recognized, is that the intervention delivered through RCTs can be replicated under routine conditions. This will be discussed in a later section of this article.

Issues related to feasibility and ethics often preclude the use of RCTs for testing potential interventions.[12] Less frequently recognized is that probability results alone often fail to provide answers to many of the relevant questions posed by evaluations of large-scale public health interventions. A perfectly conducted RCT provides an unbiased probability statement of causality between the intervention and the impact indicator. This probabilistic statement, however, requires further evidence to be biologically and conceptually plausible.

Plausibility arguments can strengthen a probability statement made by an RCT. First, plausibility thinking is required to correct the inevitable shortcomings of any RCT, even if perfectly designed and conducted. For example, randomization does not exclude confounding—the possibility that variables other than the intervention may be independently associated both with exposure to the intervention and with the outcome. However, confounding is very likely if information is collected—as it should be—on a sufficient number of baseline characteristics of the intervention and comparison groups. In such cases, when amending the statement of probability to adjust the results for this confounder, researchers are in fact using observational findings to improve on RCT results. Even the most stringent RCT proponents are willing to modify a probability statement if it will enhance the plausibility of their findings.

The second way that plausibility can strengthen a probability statement is by accounting for diversions from the RCT protocol in the analysis and interpretation of the results. Losses to follow-up, lack of perfect blinding, and other problems that often affect RCTs, particularly in the field of public health, must be discarded by drawing on plausibility arguments. The implementation of interventions in large-scale studies is often imperfect—poor compliance or crossover between groups results in some individuals in the intervention group not receiving the intervention, and in some of those in the control group receiving it. "Intent to treat" analysis[2] ignores these problems, but it is essential from a probability standpoint and should always be presented. However, this type of analysis may underestimate or even miss the impact if compliance is inadequate. One way to address this issue is to also present conventional plausibility analyses, comparing subjects who received the intervention and those who did not, and adjusting for possible confounders.[13] Another way to address crossover is to conduct dose–response analyses within both groups. If the dose–response slopes are similar in both groups, and these groups differ only in terms of the distribution of the intervention, the plausibility of a causal effect increases.

The third way that plausibility arguments can strengthen a probability statement is by providing additional evidence that the association between the intervention and the outcome was causal. In traditional RCTs, evidence that the biological agent was reliably delivered to participants, with supporting evidence from animal or human studies demonstrating a possible physiological pathway, is often sufficient. In contrast, the causal pathways for public health interventions involve not just biological but also behavioral steps that need to be understood and measured, to demonstrate a logical sequence between intervention and outcome.

An example of a public health intervention with a long and complex causal pathway is the immunization of children against vaccine-preventable disease. Successful immunization minimally requires that health workers are trained to deliver the correct dose to children within specific age ranges; that health workers have syringes, needles, and viable vaccines available at the delivery site; and that mothers know when and where to take their child for vaccination and have the means and motivation to get there. Only after the successful completion of these steps can the biological agent be delivered to the target population. After delivery, the vaccine leads to an immune response that produces an intermediate biological outcome (diminished disease incidence) and then finally the ultimate outcome of fewer child deaths from vaccine-preventable disease. Again, findings demonstrating changes in the various links in the causal pathway can provide strong plausibility support that program impact has occurred. In settings with poor vaccine distribution systems or low immunization coverage, probability statements attributing mortality declines to an immunization program do not make sense.

A recent evaluation of a program designed to improve child growth through the training of health workers in nutrition counseling provides a good example of how measurement of intermediary behavioral steps enhances the plausibility of RCT findings.[14] As shown in Figure 1, even a simplified impact model for this intervention includes at least 6 levels. A universal response to such an intervention cannot be assumed because characteristics of the public health system (e.g., capacity to mount training programs, opportunities for contact between trained health workers and mothers) and behaviors of the population (e.g., local feeding patterns) must be taken into account in addition to the biological impact of the intervention. The large-scale impact of the program will also depend on factors outside the health system, such as the availability of foods with adequate nutritional value.

From a strict probability viewpoint, the multilevel analysis performed by Santos et al.[14] showed that 1-year-old children attending 14 health facilities randomized to a health worker training program on the Integrated Management of Childhood Illnesses (IMCI)[15] had significantly ($P<.05$) greater weight gain over a 6-month period than those attending 14 matched facilities with standard care. This result had limited internal validity, and it would have convinced few readers in spite of strict adherence to RCT principles had it not been demonstrated (Figure 1) that (a) it was possible to train a large proportion of health workers in IMCI, (b) IMCI-trained workers had better counseling performance than untrained workers, (c) mothers were receptive and understood the messages they received, (d) mothers in the IMCI intervention group changed their child-feeding behavior while mothers in the comparison group did not, (e) children in the IMCI intervention group ate more nutritious foods than children in the comparison group, and only then that (f) children in the IMCI intervention group had better growth rates than those in the comparison group. In such interventions with many complex steps, information on intermediate causal steps is essential for attributing the observed outcomes to the intervention because a *P* level alone will not convince a critical reader.[10]

Further plausibility can be obtained by demonstrating that the expected changes in the pathway that leads from the intervention to the outcome were of sufficient magnitude and occurred in a temporal sequence consistent with the hypothesized impact. It is important to document the *adequacy* of the observed changes along the causal pathway by using terminology proposed previously.[10] In traditional RCTs, this is often referred to as "clinical significance," because the investigators must discuss whether the observed change was clinically meaningful.

In summary, RCTs depend on plausibility and adequacy arguments to make their causal inferences believable, even in studies that meet the most rigorous standards of design and conduct. This is particularly true in the field of public health, where causal pathways between the intervention and health impact

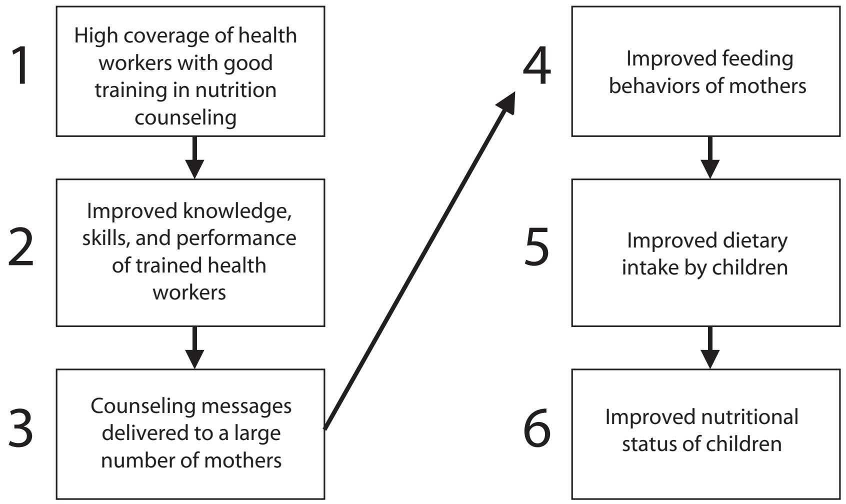
1
High coverage of health workers with good training in nutrition counseling
2
Improved knowledge, skills, and performance of trained health workers
3
Counseling messages delivered to a large number of mothers
4
Improved feeding behaviors of mothers
5
Improved dietary intake by children
6
Improved nutritional status of children

**FIGURE 1—Simplified impact model of how health worker training in nutrition counseling will lead to increased weight gain among young children in a developing country.**

are often long and complex. In these studies, evidence of plausibility and adequacy is as important as *P* levels or confidence intervals.

There are also occasions when evaluations based solely on adequacy criteria, or on a combination of adequacy and plausibility, may have sufficiently high internal validity for some outcomes to lead to correct decision-making. These issues are discussed at the end of this article.

## RCTS: EXTERNAL VALIDITY

Evidence from RCTs has often been challenged on the grounds of limited generalizability.[12,16] Such studies and their meta-analyses are based on the assumption of universal biological response. A natural corollary of this assumption suggests that if a study is internally valid (i.e., is capable of supporting a causal inference), this will ensure its external validity (generalizability). These assumptions may be appropriate for evaluations of interventions with short causal pathways and relatively simple impact models. Individual-level studies of vaccines or nutrition supplements are examples of such interventions, in which the administration of an agent leads directly to a defined biological response. These assumptions are unlikely to be appropriate, however, in evaluations of interventions involving long, complex causal pathways, or in large-scale evaluations where these pathways can be affected by numerous characteristics of the population, health system, or environment.

Epidemiologists refer to “effect modification” when the intervention–outcome association varies according to the presence of external characteristics. We propose 2 potential types of effect modification that must be considered when the generalizability of RCT results is assessed:

- Differences in the actual dose of the intervention delivered to the target population, referred to here as “behavioral effect modification,” which includes institutional, provider, and recipient behaviors.
- Differences in the dose–response relationship between the intervention and the impact indicator, to be referred to as “biological effect modification.”

### Behavioral Effect Modification Affects Dose of Intervention That Reaches Recipients

Table 1 proposes a typology of evaluation studies, with an emphasis on their expected success in delivering the intended dose of the intervention (e.g., biological agent, educational message). The studies differ by the unit of study, by how intensively the intervention is delivered to the recipients, and by whether or not measures are taken to promote high compliance among recipients.

Clinical efficacy trials (Table 1, row A) follow the standard clinical trial model in which study participants are individually selected and randomized and the dose is ensured at the individual level. It is in this type of study that RCTs have initially shown their merit. To ensure ideal compliance, staff in clinical efficacy studies are intensively trained, supervision is strong, subjects are intensively counseled and may be reimbursed for any expenses associated with the intervention (e.g., transportation to clinic), dosages are strictly controlled, side effects are monitored and managed, and noncompliers are actively sought. The dose of the intervention, therefore, is often considerably greater than can be achieved in routine circumstances, and its impact will tend to be larger.

Public health regimen efficacy studies (row B) are similar to clinical efficacy trials, but the intervention is applied to groups rather than individuals. The optimal dose of the intervention is delivered to every subject and compliance is ensured. The vitamin A efficacy trials of the 1970s provide a classic example.[17]

Public health delivery efficacy studies (row C), like regimen efficacy studies, ensure that an optimal dose of the intervention is delivered to the individual or family. However, there is no resource-intensive effort to promote compliance, although compliance is likely to be somewhat above that observed in routine circumstances (and is thus described as “best practice”). Any differences between rows B and C are due to compliance at the recipient level. For example, in a recently conducted iron supplementation trial in Bangladesh, 50 health centers were randomized to deliver either daily or weekly iron supplements to pregnant women.[18] The delivery schedule was ideal, but no special efforts were made to improve compliance among recipients.

Public health program efficacy studies (row D) entail making the intervention available to the health services but not promoting any resource-intensive efforts to ensure optimal delivery or compliance. Thus, behavioral factors pertaining to health systems and individuals are

**TABLE 1—Different Types of Studies Aimed at Evaluating the Impact of an Intervention, With Emphasis on the Dose of the Intervention That Reaches Program Recipients**

| Type of Study | Units of Treatment | Delivery Mechanism for Intervention | Compliance With Agent by Recipients | Comments |
|---|---|---|---|---|
| A. Clinical efficacy trials | Individuals | Ideal | Ideal | Classical clinical trials of drugs, vaccines, etc. |
| B. Public health regimen efficacy | Clusters of individuals[a] | Ideal | Ideal | Same as above, but delivered to clusters rather than individuals |
| C. Public health delivery efficacy | Clusters of individuals | Ideal | Best practice | Ideal delivery is ensured, and compliance is actively promoted according to best practice |
| D. Public health program efficacy | Clusters of individuals | Best practice | Best practice | Randomized allocation of geographical areas to best practice implementation |
| E. Public health program effectiveness | Clusters of individuals | Routine | Routine | Randomized allocation of geographic areas to routine implementation |

[a] For example, geographically defined administrative units, or catchment areas of health facilities.

allowed to affect the dose of the intervention. Given the presence of the study team, delivery and compliance are likely to be somewhat above routine levels, described here as "best practice." Differences between rows C and D are due to variability in health services behavior affecting the implementation of the intervention, such as poor health worker performance or drug shortages. The above-mentioned trial of nutrition counseling delivered to mothers by health workers in Brazil[14] provides an example of this design: no special incentives were given to health workers to improve their performance, nor to mothers to improve compliance.

Public health program effectiveness studies (row E) entail allocating geographic units to receive or not receive the intervention but making no additional efforts to improve delivery or compliance above routine levels. The difference between rows D and E reflects the contrast between "best practice" and "routine" levels. This type of trial is listed here because of its relevance for the external validity discussion, but to our knowledge few if any such studies exist. The mere presence of an evaluation team and participation in the evaluation process will encourage health systems managers to attempt to achieve "best practice" rather than "routine" delivery and compliance, thus moving along the continuum from effectiveness trials toward efficacy trials. Therefore, the public health program efficacy trial (row D) is the most relevant comparison (among the above trials) with program effectiveness.

Efficacy—defined as an intervention's effect under "ideal conditions"[2]—moves from total control of provider and recipient behaviors in studies type A and B to control of provider behaviors only in study type C, and finally to less control in study type D. However, the standard definition of "ideal" is irrelevant in public health practice, because the questions asked by health managers are more specific. For example, a type C study asks whether having total control up to the point of delivery, but investing relatively little effort in improving recipients' behavior—a "best practice" that could be generalized to other settings—would have an impact on health. A type D study, on the other hand, tests the impact when there are relatively small changes at both health facility and community levels.

This section has addressed factors affecting the dose of the intervention delivered to the subjects in different types of trials. This must be considered when projecting the impact of an intervention, because in most real-life situations the dose received by the population is likely to be smaller than in any type of trial. For this reason, it is important for public health evaluations to provide detailed information on both delivery and compliance.

## Biological Effect Modification Affects Dose–Response Association Between Intervention and Outcome

In addition to differences in the dose of the intervention, the dose–response relationship may also vary from site to site. Table 2 shows 6 categories of effect modification. Antagonism (row A) and synergism (row B) are well-known, but the presence of curvilinear dose–response associations (row C) is not as well recognized. For example, RCTs of nutritional interventions are often carried out in populations where both the outcome and nutrient deficiency are highly prevalent; when applied in better-nourished populations, their effects are often smaller.

The presence of competing interventions (row D) also often contributes to a smaller than expected impact, as does the absence of a critical cofactor (row E). Finally, other determinants that are not affected by the intervention may account for most of the disease burden in the population of interest (row F).

The above discussion highlights the importance of the length of the causal pathway between the implementation of the intervention and the final biological outcome. Drug trials have short pathways, surgical studies somewhat longer ones,[16] and public health pathways are usually the longest by far because they include (a) operational changes in provider behaviors that are required to deliver the intervention, (b) compliance by recipients, and (c) biological effects. Although effect modification can occur even for short causal pathways (Table 2), it will be more likely if there are several steps in the causal chain. For example, results of an intervention that requires improvements in health worker performance through training, assurance of a regular drug supply, and high compliance among recipients will be inherently difficult to generalize because each of these 3 components may vary—often in opposite directions—from one setting to another.

In summary, there are important restrictions to the external validity of RCTs for complex public health interventions. The likelihood of effect modification implies that one cannot take for granted that interventions that are proven efficacious in controlled trials can be generalized to other settings. This is particularly true in international health, where it

**TABLE 2—Types of Biological Effect Modification Affecting the Generalizability of Findings From Randomized Controlled Trials**

| Category of Effect Modification | Description | Examples |
|---|---|---|
| A. Presence of other factors reduces the dose–response slope (antagonism) | Other factors are present in the target population that reduce the extent to which the intervention affects the outcome. | Iron and zinc supplementation will be less effective in places where the local diet contains substances that reduce their absorption (e.g., phytates and polyphenols). |
| B. Presence of other factors increases the dose–response slope (synergism) | Other factors are present in the target population that enhance the extent to which the intervention affects the outcome. | Iron supplementation will be more effective if the local diet is rich in meat and ascorbic acid, which enhance iron absorption. |
| C. Curvilinear dose–response association | Many biological responses are curvilinear; the same dose will have less effect if there is less need for it. | Iron supplementation will have different effects on hemoglobin according to baseline iron stores. Also, iron absorption is inversely related to iron status. |
| D. Limited scope for improvement in the impact (outcome) indicator because other interventions already provide protection | The intervention that is already in place acts on another link in the causal chain. | Use of insecticide-treated bed nets will have a limited effect on malaria mortality if case-management is already appropriate. |
| | The intervention acts on the same causal link. | Improved breastfeeding will have less effect if water supply and sanitation are adequate. |
| E. Intervention is inappropriate because a critical cofactor is missing | The intervention only works in the presence of another factor that is absent in the population in question. | Improving water quality will have an impact on diarrheal diseases only if water quantity is adequate. |
| F. Intervention is addressing a determinant that is not important | The intervention is being applied in a setting where it is not needed because the outcome it addresses has other causes. | Energy supplementation in pregnancy will have limited impact on low birthweight if the latter is mostly due to maternal smoking and to preterm deliveries caused by infections.<br>The impact of improved breastfeeding on infant mortality will be lower in populations where infectious diseases account for a small proportion of deaths. |

will never be possible to carry out RCTs in all countries where the interventions will be applied. For instance, results of a meta-analysis of randomized trials of large-scale integrated programs[19] are uninterpretable. The effectiveness of new interventions, therefore, must be monitored when implemented on a large scale. New randomized trials are not required or appropriate for this purpose. Evaluations with adequacy and plausibility designs, carried out in several settings under routine implementation conditions, are essential.

## ROLE OF PLAUSIBILITY AND ADEQUACY EVALUATIONS

For the reasons discussed above, evidence-based public health must draw on studies with designs other than RCTs. *Plausibility* evaluations attempt to document impact and to rule out alternative explanations by including a comparison group—historical, geographic, or internal—and by addressing confounding variables.[10] They are particularly useful when (a) an intervention is so complex that RCT results will be unacceptably artificial; (b) when an intervention is known to be efficacious or effective in small-scale studies, but its effectiveness on a large scale must be demonstrated; and (c) when ethical concerns preclude the use of an RCT. In these 3 scenarios, plausibility evaluations are not just a "second best" alternative to RCTs, they are indeed the only feasible alternative.

*Adequacy* evaluations—documentation of time trends in the expected direction, following introduction of an intervention[10]—may also stand on their own. Evaluations of the impact of motorcycle helmet legislation[20–22] and of *Haemophilus influenzae* type B vaccine in Uruguay[23] were highly persuasive. We propose 3 prerequisites for valid adequacy evaluations: (a) the causal pathway must be relatively short and simple, (b) the expected impact must be large, and (c) confounding must be unlikely. Regardless of the length of the causal pathway, adequacy evaluations are particularly useful when there is no impact. If an assessment of intermediate steps in the causal pathway reveals that changes have not occurred, or that they are not of sufficient magnitude or have occurred in an illogical temporal sequence, expensive and time-consuming plausibility and probability evaluations are unnecessary.

## CONCLUSIONS

RCTs are rightly regarded as the gold standard for clinical decisionmaking purposes. However, we argue that in the evaluation of public health interventions, RCTs are never sufficient by themselves. Randomization, without further analyses for adequacy and plausibility, is never sufficient to support public health decisionmaking, regardless of the level of statistical significance achieved.

Evaluating the impact of large-scale public health programs is difficult because the interventions are usually multiple and their pathways to impact are complex and subject to effect modification. An intervention that works well in a given setting may be ineffective elsewhere, presenting a huge challenge to international health recommendations. True evidence-based public health must rely on a variety of types of evidence, often in combination. Current trends toward acceptance of RCTs as the gold standard source of evidence may limit the knowledge base needed to make sound decisions about public health priorities and policies. This limitation both results in making poor recommendations by failing to account for effect modification when generalizing from RCTs and prevents making useful recommendations on the basis of sound plausibility inferences.

Depending on the circumstances, adequacy or plausibility evaluations may be sufficient to support sound decisionmaking in public health. Resources for public health research and evaluation are scarce. More attention must therefore be given to assessing the cost and feasibility of various study designs relative to their effectiveness in producing data sufficient for sound decisionmaking.

Evidence-based public health requires the further refinement of protocols for the conduct and reporting of plausibility designs. Over time, methods for interpreting results across plausibility studies and an organized system to facilitate access (similar to the Cochrane collection for RCTs) will need to be developed. These resources should be specifically designed to address the challenges of evaluating large-scale public health interventions with complex causal chains. The urgency of this need has already become clear to public health policymakers facing major decisions. For example, the development of policies on the prevention of mother-to-child transmission of HIV was hindered by the absence of adequate data on the efficacy of various potential delivery strategies. Another example is the recent revision of recommen-dations for the optimal duration of exclusive breastfeeding, which was heavily dependent on observations from plausibility studies[24] because ethical and logistical limitations precluded the implementation of probability trials.

Evidence-based public health must continue to draw on RCTs as an important source of information. At the same time, existing standards and methods must be adapted to meet the methodological challenges of evaluating large-scale public health interventions. Although some progress can be made through extensions and adaptations of the RCT model, this incremental approach provides only a partial answer. New designs that incorporate adequacy and plausibility approaches must also be developed, tried, and taught. ■

## About the Authors

*Cesar G. Victora is with the Post-Graduate Program in Epidemiology, Federal University of Pelotas, Pelotas, Brazil. Jean-Pierre Habicht is with the Division of Nutritional Sciences, Cornell University, Ithaca, NY. At the time this report was written, Jennifer Bryce was with the Department of Child and Adolescent Health and Development of the World Health Organization, Geneva, Switzerland.*

*Requests for reprints should be sent to Cesar G. Victora, MD, PhD, Post-Graduate Programme in Epidemiology, Universidade Federal de Pelotas, CP 464, 96001–970, Pelotas, RS, Brazil (e-mail: dcdesjarla@aol.com).*

*This article was accepted April 23, 2003.*

## Contributors

C. G. Victora conceived this report and produced the first draft. J.-P. Habicht contributed insights based on his public health experience and devised the classification scheme for evaluations. J. Bryce provided additional insights and was in charge of the final version. All authors helped to conceptualize ideas, interpret findings, and review the article.

## Acknowledgments

This report was prepared as a background document for the Multi-Country Evaluation of IMCI (Integrated Management of Childhood Illness) Impact, Costs and Effectiveness, supported by the Division of Child and Adolescent Health and Development, World Health Organization, and by the Bill and Melinda Gates Foundation.

This report was originally presented as the keynote address at the closing conference of the 5th Brazilian Epidemiology Congress in Curitiba, Brazil, March 2002.

## References

1. Eriksson C. Learning and knowledge-production for public health: a review of approaches to evidence-based public health. *Scand J Public Health.* 2000;28:298–308.

2. Last JM, ed. *A Dictionary of Epidemiology.* 3rd ed. New York, NY: Oxford University Press; 1995.

3. Moher DL, Schulz KF, Altman DG for the CONSORT Group. The CONSORT statement: revised recommendations for improving the quality of reports of parallel group randomized trials. *Lancet.* 2001;357:1191–1194.

4. Lambert SM, Markel H. Making history: Thomas Francis, Jr, MD, and the 1954 Salk Poliomyelitis Vaccine Field Trial. *Arch Pediatr Adolesc Med.* 2000;154:512–517.

5. Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ.* 1996;312:71–72.

6. The Cochrane Collaboration: preparing, maintaining and promoting the accessibility of systematic reviews of the effects of health care interventions. Available at: http://www.update-software.com/Cochrane. Accessed December 17, 2003.

7. Ham C, Hunter D, Robinson R. Evidence based policymaking. *BMJ.* 1995;310:71–72.

8. Macintyre S, Chalmers I, Horton R, Smith R. Using evidence to inform health policy: case study. *BMJ.* 2001;322:222–225.

9. *Journal of Evidence Based Healthcare* (formerly known as *Evidence-Based Health Policy and Management*). Available at: http://www.harcourt-international.com/journals/ebhc. Accessed December 17, 2003.

10. Habicht J-P, Victora CG, Vaughan JP. Evaluation designs for adequacy, plausibility and probability of public health programme performance and impact. *Int J Epidemiol.* 1999;28:10–18.

11. Cook TD, Campbell DT. *Quasi-Experimentation: Design and Analysis Issues.* Boston, Mass: Houghton Mifflin; 1979.

12. Black N. Why we need observational studies to evaluate the effectiveness of health care. *BMJ.* 1996;312:1215–1218.

13. Kramer MS, Chalmers B, Hodnett ED, et al. Promotion of Breastfeeding Intervention Trial (PROBIT): a randomized trial in the Republic of Belarus. *JAMA.* 2001;285:413–420.

14. Santos IS, Victora CG, Martines JC, et al. Nutrition counseling increases weight gain among Brazilian children. *J Nutr.* 2001;131:2866–2873.

15. Tulloch J. Integrated approach to child health in developing countries. *Lancet.* 1999;354(suppl 2):SII16–SII20.

16. Black N. Evidence based policy: proceed with care. *BMJ.* 2001;323:275–279.

17. Sommer A, Tarwotjo I, Dunaedi E, et al. Impact of vitamin A supplementation on childhood mortality: a randomized controlled community trial. *Lancet.* 1986;1:1169–1173.

18. Ekström EC, Hyder SMZ, Chowdhury MA, et al. Efficacy and trial effectiveness of weekly and daily iron supplementation among pregnant women in rural Bangladesh: disentangling the issues. *Am J Clin Nutr.* 2002;76:1392–1400.

19. Briggs CJ, Capdegelle P, Garner P. Strategies for integrating primary health services in middle- and low-income countries: effects on performance, costs and patient outcomes. *Cochrane Database Syst Rev.* 2001;(4):CD003318.

20. Fleming NS, Becker ER. The impact of the Texas 1989 motorcycle helmet law on total and head-related fatalities, severe injuries, and overall injuries. *Med Care.* 1992;30:832–845.

21. Kraus JF, Peek C, McArthur DL, Williams A. The effect of the 1992 California motorcycle helmet use law on motorcycle crash fatalities and injuries. *JAMA.* 1994;272:1506–1511.

22. Panichaphongse V, Watanakajorn T, Kasantikul V. Effects of law promulgation for compulsory use of protective helmets on death following motorcycle accidents. *J Med Assoc Thai.* 1995;78:521–525.

23. Pan American Health Organization. Impact of Uruguay's introduction of the Haemophilus influenzae type b (Hib) vaccine. *EPI Newsl.* 1996;18:6.

24. Kramer MS, Kakuma R. Optimal duration of exclusive breastfeeding. *Cochrane Database Syst Rev.* 2002;(1):CD003517.