

Evaluating Public Health Interventions:

1. Examples, Definitions, and a Personal Note

In the first contribution to a new section in *AJPH* that will address critical methodological issues in evaluations of public health interventions, I will discuss topics in study design and analysis, covering the most innovative emerging methodologies and providing an overview of best practices. The methods considered are motivated by public health evaluations, both domestic and global. In this first contribution, I also define implementation science, program evaluation, impact evaluation, and cost-effectiveness research, disciplines that have tremendous methodological and substantive overlap with evaluation of public health interventions—the focus of this section. (*Am J Public Health*. 2016;106:70–73. doi:10.2105/AJPH.2015.302923)

Donna Spiegelman, MS, ScD

At the invitation of Alfredo Morabia, editor-in-chief of *AJPH*, I am pleased to offer the first piece in a new *Journal* section, “Evaluating Public Health Interventions,” that will address critical methodological issues that arise in the course of evaluating public health interventions. In this new section, I will consider topics in study design and analysis, describe the latest and most innovative emerging methodologies, and provide an overview of best practices. Although in many cases these best practices may be “tried and true,” others are promising recent innovations. I also will consider estimation of effectiveness as well as cost-effectiveness. The methods discussed are motivated by public health evaluations of interest, both domestic and global.

EXAMPLES

As an example, the Affordable Care Act (ACA) has much to offer in terms of support for primary prevention,^{1,2} and there is much to be learned about the extent to which it has achieved its potential, both in terms of process outcomes (i.e., whether newly and more widely covered preventive services have been used) and in terms of measurable, cost-effective health improvement. For instance, in assessing the impact of the ACA on cervical cancer prevention, Lipton and Decker estimated a 6-percentage-point increase in completion

of the human papillomavirus vaccine series by young women that could be attributed to enhanced ACA coverage.³ The extent to which this increased uptake leads to future reductions in the occurrence of cervical abnormalities and overt malignancies remains to be documented.

To consider another provocative example, the US government has spent nearly \$52 billion on the President’s Emergency Plan for AIDS Relief (PEPFAR) and related bilateral aid programs since 2004. At the time PEPFAR was reauthorized, Congress mandated that the Institute of Medicine assess its performance and health impact. This assessment resulted in a massive 814-page document that incorporated a mixed-method approach and included extensive health and economic data. The conclusion of the report was that although there were some areas for improvement, the program has performed well overall.⁴ During her confirmation hearings in March 2014, ambassador Deborah Birx, coordinator of United States Government Activities to Combat HIV/AIDS, stated:

As a physician and epidemiologist, I am strongly committed to ensuring that country-driven analysis steers efforts to accelerate action to rapidly scale up effective interventions for maximum impact and controlling the HIV epidemic. Science, epidemiology, and dynamic data systems are essential. We will work with partner countries toward scaling up the best models for facility- and community-based service delivery that ensures that our resources go to the right people at the right time. We will prioritize reduction of sexual transmission by driving programs using epidemiological data and intervention effectiveness. To achieve an AIDS-free generation, we must analyze the epidemic country-by-country and tailor our approach to those most at risk.⁵

Clearly, the PEPFAR enterprise recognizes the fundamental role of ongoing evaluation in focusing shrinking AIDS-related resources as effectively as possible as the program continues to unfold and as it appears that incidence and mortality decline throughout the world.⁶

What was the most important thing Chelsea Clinton, vice chairman of the Clinton Foundation, learned from the master of public health degree she

ABOUT THE AUTHOR

Donna Spiegelman is with the Departments of Epidemiology, Biostatistics, Nutrition, and Global Health, Harvard T. H. Chan School of Public Health, Boston, MA.

Correspondence should be sent to Donna Spiegelman, MS, ScD, Department of Epidemiology, Harvard T. H. Chan School of Public Health, 677 Huntington Ave, Boston, MA 02115 (e-mail: Donna_ContributingEditor@apha.org). Reprints can be ordered at <http://www.ajph.org> by clicking the “Reprints” link.

*This article was accepted September 23, 2015.
doi: 10.2105/AJPH.2015.302923*

TABLE 1—Sets of Disciplines for Which This Section Will Broadly Address Methodological Considerations

Discipline	Definition	Type of Knowledge
Implementation science	Assesses the extent to which efficacious health interventions can be effectively integrated within real-world public health and clinical service systems	Widely applicable
Impact evaluation	Assesses the efficacy and effectiveness of an intervention in terms of intended and unintended health, social, and economic outcomes; involves the explicit statement of a counterfactual	Widely applicable
Program evaluation	Assesses the processes and outcomes of a program with the intent of furthering its improvement	Program specific
Comparative effectiveness research	Assesses which treatment works best for whom, and under what circumstances, and considers health as well as economic outcomes	Widely applicable, clinical focus

earned at Columbia's Mailman School of Public Health? Her one-word reply (during the Voices in Leadership Series last spring at the Harvard T. H. Chan School of Public Health) when this question was posed to her by Atul Gawande, professor of surgery at Harvard Medical School and author of several best-selling health-related books,^{7,8} was "statistics." Using statistical programs such as Stata, she recounted to an enthralled audience, helps her "absorb information more quickly and mentally sift through and catalog it."⁹

Perhaps I have been "preaching to the converted" readers of the *Journal*, wherein quantitative evaluation is an essential feature of most of the material published; if not, however, I hope the examples I have highlighted will persuade any remaining skeptics as to the essential role of statistics, epidemiological methods, and quantitative methods in promotion of the public's health.

DEFINITIONS

In this section, I broadly address methodological considerations in an evolving set of disciplines that have variously

been labeled *implementation science*, *impact evaluation*, *program evaluation*, and *comparative effectiveness research* (Table 1). Although I offer definitions subsequently, credible alternative definitions exist for each one. It is my view that the similarities overwhelm the differences and that the distinctions that have been made obfuscate the underlying common ground. It seems that there is a great deal of interest currently in evaluation of public health programs, large and small, domestic and global, and that several professional communities are engaged in this overlapping work: health care policy analysts under the banner of comparative effectiveness research, health economists under the banner of impact evaluation, public health program implementers under the banner of program evaluation, and global health funders, especially at the Fogarty Institute of the National Institutes of Health (NIH), under the banner of implementation science.

Implementation science addresses the extent to which efficacious health interventions can be effectively integrated into real-world public health and clinical service systems.^{10,11} Implementation science

compares multiple evidence-based interventions, identifies strategies to encourage the provision and use of effective health services, promotes the integration of evidence into policy and program decisions with the goal of adapting interventions to a range of populations and settings, and identifies approaches for scaling up effective interventions to improve health care delivery. The growth of implementation science has been stimulated by US health agencies such as NIH, and the discipline has close links to the public health community of physicians, health care administrators, epidemiologists, statisticians, and others. The *Journal* has published numerous highly cited examples of implementation science studies, including the work of Wallerstein and Duran,¹² Glasgow et al.,¹³ Scheirer and Dearing,¹⁴ and Scheirer¹⁵ with many more expected given the high level of interest in this discipline.

In partial contrast, impact evaluation assesses how an intervention affects intended and unintended outcomes, and it involves the invocation of a counterfactual to compare what happened when the intervention was implemented with what would have happened had it not

been implemented. The latter is the counterfactual. An array of methods have been developed and are under development to eliminate or at least mitigate the biases that can occur because of the impossibility of turning back the clock and rerunning the alternative intervention with the same population during the same time.

It appears to me that much of the work conducted under the rubric of impact evaluation is focused on effectiveness. Briefly, efficacy denotes the ability of an intervention to produce its desired outcome under idealized, tightly controlled settings, whereas effectiveness refers to the ability of the intervention to produce the desired outcome under large-scale, relatively uncontrolled settings. Establishing strong proxies to counterfactuals—that is, eliminating or largely mitigating the biases to which effectiveness research is susceptible—is considerably more difficult than when efficacy is the focus, although in any case a counterfactual (the ideal comparison for obtaining an unbiased estimate of effect) can be conceptualized even when the researcher is unable to achieve it or even approximate it.

One organization promoting impact evaluation is 3ieimpact.

org, cofunded by the Bill and Melinda Gates Foundation, the UK Department for International Development, and others. 3ieimpact.org supports high-priority impact evaluations in low- and middle-income countries, disseminates methodology, and publishes a journal, the *Journal of Development Effectiveness*. As should be becoming apparent, the discipline of impact evaluation has arisen from the field of development economics, which itself has become increasingly focused on health outcomes related to alternative economic development strategies. The study by Trickett et al.¹⁶ is an example of a recent highly cited impact evaluation published in the *Journal*.

Program evaluation overlaps substantially with both implementation science and impact evaluation. Program evaluation has been defined as “the systematic assessment of the processes and/or outcomes of a program with the intent of furthering its development and improvement.”¹⁷ During program implementation, evaluators may provide findings to enable immediate, data-driven decisions for improving program delivery. At the completion of a program, evaluators provide findings—often required by funding agencies—that can be used to make decisions about program continuation or expansion.

In contrast to implementation science and impact evaluation, which aim to produce widely applicable knowledge about programs and interventions, program evaluation has the more modest goal of simply evaluating a given program in its given setting, time, and context, and it may in some instances lack the ability to provide a valid formal statistical hypothesis test owing to the continuous nature of the

evaluation process. Some recent highly cited program evaluations that have appeared in the *Journal* include those of Scheirer and Dearing,¹⁴ Pulos and Ling,¹⁸ Woodward-Lopez et al.,¹⁹ and Thrasher et al.²⁰

Comparative effectiveness research, which compares existing health care interventions to determine which are most effective for different groups of patients and which involve the greatest benefits and harms, overlaps substantially with the other disciplines as well.²¹ Comparative effectiveness research typically includes cost-effectiveness analyses incorporating incremental cost-effectiveness ratios²² and quality-adjusted life-year metrics,²³ with the pragmatic randomized controlled trial as a major design tool.²⁴ Although comparative effectiveness research shares much with the other three disciplines just discussed, it focuses more directly on the relative benefits and costs of alternative clinical treatment modalities. Brody and Light’s work²⁵ is an example of a highly cited study in the area of comparative effectiveness research that has appeared recently in the *Journal*.

PERSONAL NOTE

One may wonder, what qualifies me to be the *Journal’s* interlocutor of these topics? Well, I am one of the few people in the world with a joint doctorate in biostatistics and epidemiology.²⁶ As a result, I can freely speak the languages of both disciplines and switch between these two professional cultures. Until recently, my research has been motivated by problems arising in epidemiology that require biostatistical solutions. In particular, but by no means exclusively, I

have focused on study design and data analysis methods that reduce bias in estimation and inference due to measurement error or misclassification of exposure variables. However, my previous methodological work has also covered the development of improved meta-analysis methods, the study of gene-environment interactions, and estimations of population-attributable risk, among other areas.

My Web site is one of the most visited at the Harvard T. H. Chan School of Public Health, where I am a professor of epidemiologic methods in the Departments of Epidemiology, Biostatistics, Nutrition, and Global Health, because it contains much user-friendly and well-documented freeware (available at <https://www.hsph.harvard.edu/donna-spiegelman/software>) through which nonstandard methods useful in public health are implemented. I am the statistician for the Nurses’ Health Study 2; the Health Professionals Follow-Up Study; the MaxART Study, focusing on early access to antiretroviral therapy in Swaziland; the Harvard PEPFAR strategic information technical assistance initiative for HIV treatment and care in greater Dar es Salaam, Tanzania; and a new implementation science project assessing the effectiveness of a worksite intervention designed to reduce cardiovascular and diabetes risk in India. I am the author of more than 600 peer-reviewed publications.

Perhaps of greatest relevance to the launch of the current series is that I have received a Director’s Pioneer Award from NIH. One of 10 researchers so honored in 2014, to my knowledge I am the first epidemiologist and biostatistician—and the first faculty member from a school of public health—to receive this

award. The five-year, \$2.5 million prize, according to the NIH Web site, recognizes

individual scientists of exceptional creativity who propose pioneering, and possibly transforming, approaches to major challenges in biomedical and behavioral research.

I am using this opportunity to focus on the development of new methods needed to advance the field of implementation science and related disciplines. As part of this effort, I am developing a software and data platform for monitoring and evaluating large-scale disease prevention projects in real time. The methods incorporated into this toolkit will be general enough to be applicable to a variety of types of interventions, including those aimed at mitigating the global obesity epidemic, reducing maternal mortality, and increasing the use of cleaner-cooking stoves in developing countries. In addition to development of methodologies and identification of best practices, dissemination is an essential component, and my hope is that this bimonthly section will play a major role in getting the word out.

In the coming months, I will address topics of interest that cut across implementation science, impact evaluation, program evaluation, and comparative effectiveness research. The underlying unifying feature of all four disciplines is that each strives to provide causal estimates of the objects under study. Thus, the methodological commonalities across these four disciplines are vast and the differences quite small. I will point these out as they occur.

Topics of forthcoming sections will include some or all of the following: an exploration of stepped wedge designs; pros and

cons of cluster randomized and stepped wedge designs; two-stage designs for public health evaluations; an assessment of quasi-experimental designs and how they can be used in public health evaluations; the ways in which big data can be harnessed to improve public health; randomization, observation, and causal inference in public health evaluation; the many ways to control for confounding in nonrandomized public health evaluations; causal confusion in policy and program evaluations (what causal inference is and is not); process evaluation and mediation analysis; impact evaluation methods; best practices in post hoc policy evaluations; parsimony, modeling, and causal inference in nonrandomized public health evaluations; whether Bayesian statistics are useful in evaluations of public health interventions; the ethics of large-scale public health evaluations; and an examination of mixed methods and how they are used in public health evaluations.

Is further clarity needed on the distinctions or lack thereof between implementation science, impact evaluation, program evaluation, and comparative effectiveness research? Am I on track with the topics proposed? What have I neglected to mention that you would like to know about? Please write to me at Donna_ContributingEditor@apha.org with your questions, suggestions, and feedback. Don't be shy! Let's get this right. **AJPH**

ACKNOWLEDGMENTS

This work was supported by National Institutes of Health grant DP1ES025459.

REFERENCES

1. Koh HK, Sebelius KG. Promoting prevention through the Affordable Care Act. *N Engl J Med*. 2010;363(14):1296–1299.

2. Fox JB, Shaw FE. Clinical preventive services coverage and the Affordable Care Act. *Am J Public Health*. 2015;105(1):e7–e10.
3. Lipton BJ, Decker SL. ACA provisions associated with increase in percentage of young adult women initiating and completing the HPV vaccine. *Health Aff (Millwood)*. 2015;34(5):757–764.
4. *Evaluation of PEPFAR*. Washington, DC: National Academies Press; 2013.
5. Birx DL. Written testimony, Senate Foreign Relations Committee. Available at: http://www.foreign.senate.gov/imo/media/doc/030614PM_Testimony%20-%20Deborah%20Birx.pdf. Accessed October 12, 2015.
6. How AIDS changed everything—MDG 6: 15 years, 15 lessons of hope from the AIDS response. Available at: http://www.unaids.org/sites/default/files/media_asset/MDG6Report_en.pdf. Accessed October 12, 2015.
7. Gawande A. *Being Mortal: Medicine and What Matters in the End*. New York, NY: Metropolitan Books; 2014.
8. Gawande A, Lloyd JB. *The Checklist Manifesto: How to Get Things Right*. New York, NY: Metropolitan Books; 2010.
9. Clinton C. Voices in leadership. Available at: <http://www.hsph.harvard.edu/voices/events/chelsea>. Accessed October 12, 2015.
10. Madon T, Hofman KJ, Kupfer L, Glass RL. Public health. Implementation science. *Science*. 2007;318(5857):1728–1729.
11. Kruk ME. More health for the money—toward a more rigorous implementation science. *Sci Transl Med*. 2014;6(245):245ed17.
12. Wallerstein N, Duran B. Community-based participatory research contributions to intervention research: the intersection of science and practice to improve health equity. *Am J Public Health*. 2010;100(suppl 1):S40–S46.
13. Glasgow RE, Vinson C, Chambers D, Khoury MJ, Kaplan RM, Hunter C. National Institutes of Health approaches to dissemination and implementation science: current and future directions. *Am J Public Health*. 2012;102(7):1274–1281.
14. Scheirer MA, Dearing JW. An agenda for research on the sustainability of public health programs. *Am J Public Health*. 2011;101(11):2059–2067.
15. Scheirer MA. Linking sustainability research to intervention types. *Am J Public Health*. 2013;103(4):e73–e80.
16. Trickett EJ, Beehler S, Deutsch C, et al. Advancing the science of community-level interventions. *Am J Public Health*. 2011;101(8):1410–1419.
17. University of Washington, Office of Educational Assessment. Frequently asked questions. Available at: http://www.washington.edu/oea/services/research/program_eval/faq.html. Accessed October 12, 2015.
18. Pulos E, Ling K. Evaluation of a voluntary menu-labeling program in full-service restaurants. *Am J Public Health*. 2010;100(6):1035–1039.
19. Woodward-Lopez G, Gosliner W, Samuels SE, Craypo L, Kao J, Crawford PB. Lessons learned from evaluations of California's statewide school nutrition standards. *Am J Public Health*. 2010;100(11):2137–2145.
20. Thrasher JF, Huang L, Pérez-Hernández R, Niederdeppe J, Arillo-Santillán E, Alday J. Evaluation of a social marketing campaign to support Mexico City's comprehensive smoke-free law. *Am J Public Health*. 2011;101(2):328–335.
21. Wikipedia. Comparative effectiveness research. Available at: https://en.wikipedia.org/w/index.php?title=Comparative_effectiveness_research&oldid=669450050. Accessed October 12, 2015.
22. Drummond M, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. *Methods for the Economic Evaluation of Health Care Programmes*. 3rd ed. Oxford, England: Oxford University Press; 2005.
23. Klarman HE, Francis JOS, Rosenthal GD. Cost effectiveness analysis applied to the treatment of chronic renal disease. *Med Care*. 1968;6(1):48–54.
24. Roland M, Torgerson DJ. What are pragmatic trials? *BMJ*. 1998;316(7127):285.
25. Brody H, Light DW. The inverse benefit law: how drug marketing undermines patient safety and public health. *Am J Public Health*. 2011;101(3):399–404.
26. Spiegelman D. Donna Spiegelman home page. Available at: <http://www.hsph.harvard.edu/donna-spiegelman>. Accessed October 12, 2015.