External Validity: We Need to Do More

Russell E. Glasgow, Ph.D.
Kaiser Permanente Colorado

Lawrence W. Green, Dr.PH
University of California at San Francisco

Lisa M. Klesges, Ph.D.
Mayo Clinic

David B. Abrams, Ph.D.
Office of Behavioral and Social Sciences Research, NIH

Edwin B. Fisher, Ph.D.
University of North Carolina

Michael G. Goldstein, M.D.
Institute for Healthcare Communication

Laura L. Hayman, Ph.D., R.N., F.A.A.N.
New York University

Judith K. Ockene, Ph.D.
University of Massachusetts

C. Tracy Orleans, Ph.D.
Robert Wood Johnson Foundation

This editorial discusses the implications of two important trends in health research for enhancing the design, evaluation, and reporting of applied research and evaluation studies. The first trend is the gap between research findings and application in practice. In the context of doubling of the National Institutes of Health (NIH) budget, as well as increased attention from the Agency for Healthcare Research and Quality, the Centers for Disease Control, the Institute of Medicine, and the NIH Road Map, the percentage of evidence-based findings that have been translated into practice is discouragingly small (1–4). Evidence-based guidelines, such as those for behavioral counseling strategies from the U.S. Preventive Services Task Force (5), are very important, but more effort is needed to increase translation into practice.

There are multiple, complex reasons for this gap or chasm between research findings and their application in practice (1–4). Prominent among these reasons are insufficient training in translational research; systems barriers including brief, acute-care oriented medical visits; insufficient time and competing resources; and nonexistent or perverse incentive and reimbursement systems. Promising approaches to help cross this chasm include guidance and training of more practitioners on the subtleties of evidence-based medicine; practice-based research networks and partnerships among transdisciplinary collaborators, including practitioners; participatory research that involves decision makers and recipients; tools and guidelines for policy- and decision makers on blending research evidence with the other determinants of policy and practice; and a greater focus on external validity. In this article we address only the last approach. Specifically, we are concerned with the scientific issue of what types of indicators and criteria are considered relevant to report. We propose a specific step that *Annals* (and we hope other health and health care journals in the future) might take as an initial action to help improve the current situation.

The second trend is the increased focus on the methodological quality of research reports. The primary exemplar of this trend is the establishment and wide-scale adoption of the CONSORT reporting criteria for randomized clinical trials (6). The *Annals* and most other health journals have adopted these criteria. Related methodological quality rating scales, such as TREND (7), have been developed for nonrandomized trials, and quality scales such as that of Jadad and colleagues (8) are also widely used in evidence-based reviews. Recent reports document the enhanced internal validity and analytic reporting quality associated with the use of such criteria (9). Although these are minimal standards, they have increased the consistency and scientific rigor of trials. The widespread adoption of these standards and the extent to which current criteria have enhanced internal validity have been impressive and raise the question of whether adoption of reporting standards for external validity could similarly improve the evidence on external validity, and thus further improve the quality and relevance of the evidence-base.

How are these trends related and how might the bridge between science and practice be strengthened? More specifically,

We thank Dr. Bonnie Spring for her helpful comments and feedback on earlier versions of this editorial.

Reprint Address: R. E. Glasgow, Ph.D., Kaiser Permanente Colorado, Clinical Research Unit, 335 Road Runner Lane, Penrose, CO 81240. E-mail: russg@ris.net

what can we learn from the success of the quality reporting experiences to date to help us reduce the gap between research and practice? From our perspective, one commonality is the relative lack of emphasis and reporting on external validity (10,11). CONSORT and related criteria, such as the Jadad scale (8), are concerned predominantly with internal validity criteria such as randomization, double blinding, and other controls over potential confounders, so that results can be attributed to highly specified interventions. Although these criteria are important and necessary to improve the quality and status of applied health research, they do not provide needed information on the context and potential applicability of research to local conditions and settings.

The questions and concerns of clinicians, administrators, and policymakers are related more to *external validity*, generalization, and applicability of findings (12,13). A recent report by leading Canadian thinkers about knowledge translation and integration lamented the often unmet need for relevant, usable research for knowledge synthesis and cites data that 76% of decision makers rated the relevance of research questions as a barrier to research uptake (14). Current quality rating approaches do not address the needs of decision makers (12), with the exception of one or two general items on target population. As discussed in detail by Green and Glasgow (15), two key external validity issues require increased attention to advance the field and to help close the gap between research and practice. First, practitioners are concerned about whether research findings apply to *their* setting and practice. Often because of exclusion criteria and the selection or establishment of research settings to maximize experimental control and intervention expertise, the settings, patients, and intervention staff in controlled studies are quite disparate from the conditions that practitioners face. All too often, studies do not report sufficient contextual information for readers to judge applicability.

Policymakers and research leaders committed to research-to-practice application are concerned about the breadth or generality of findings and their efficiency and cost-effectiveness when delivered in real-world settings and with hard-to-reach, diverse populations. Conducting a study with excellent internal validity provides important but limited information (typically on the sample of convenience who participated in the trial and using an intervention delivered with optimal training and quality controls) about the external validity of these findings. Most reports do not include any information from which to judge the feasibility and cost-efficiency of replication in diverse settings and populations (11,15). As recent reviews have demonstrated, the level and quality of reporting on external validity factors in behavioral medicine studies is substantially worse than on internal validity issues (11).

The common implication of these two trends is the need to go beyond CONSORT, TREND, and current rating scales to also promote and adopt reporting standards that include external validity and are related to translation. Just as CONSORT has improved the evidence base, our hope is that an emphasis on external validity will improve "the science of implementation and dissemination." There is urgent need to also improve the evidence base to help identify programs that can have an impact on population-level disease burden (16) and help improve the health of the nation (1). As a partial solution to the challenge of translation, Green and Glasgow (15) suggested a specific set of reporting criteria, the adoption of which would address the concerns of both clinicians and policy makers. We refer readers to that article for details, but the key criteria involve reporting on participation and representativeness at the levels of patients, clinicians, or interventionists, and settings. Specific issues that are critical to report so that policymakers and practitioners can assess relevance are categorized under the headings of Reach (e.g., participation rate and representativeness), Program or Policy Implementation (e.g., levels of interventionist expertise and training; consistency of delivery; degree of adaptation to local circumstances), Outcomes (e.g., impact on costs, quality of life and adverse consequences), and Maintenance and Sustainability (e.g., which components are institutionalized or modified over time) (15).

The majority of intervention studies conducted and reported in *Annals* and other health journals have been "efficacy" studies that maximize internal validity (and are usually silent on external validity) so that cause and effect attributions can be made for specific intervention elements (10,17). Such studies are also the type of research that the NIH Roadmap refers to as "bench to bedside" translation. These trials have been extremely valuable in establishing the scientific basis of behavioral medicine. In our perspective, two additional steps are now needed:

1. Controlled efficacy studies are needed that are somewhat less "decontextualized" and report at least a minimum amount of key information on how the setting, patients, and interventionists were selected and how they compare to those in practice. Such studies do not need to employ large, diverse samples; to be representative; or to be conducted in typical real-world conditions. However, we do think that researchers developing and evaluating interventions intended for future application to practice must begin to think about and report on minimal criteria affecting external validity, in addition to internal validity, even in the early stages of research (18).
2. Greater emphasis needs to be placed on so-called practical clinical trials (12,13) and what are increasingly called implementation, effectiveness, or dissemination research methods. Such studies need to place a high priority on generalization and feasibility. They should report important information on contextual variables such as representativeness, reach, implementation and adaptation, and outcomes important to decision makers. These studies should include comprehensive reporting on external validity criteria.

Both steps need to be taken. Otherwise, it is likely that the past will predict the future. Whatever more efficacious interventions may come out of the front end of the "research pipeline" envisioned in the NIH Roadmap (e.g., bench to bedside) either will likely be slow or will never reach the community settings and populations that need them most. Even just reporting criteria related to study context—without necessarily changing other

design issues—would help a great deal in making judgments about the applicability of study results. Given the pressing nature of behavioral and health care issues such as the twin obesity and diabetes epidemics, we cannot wait for interventions from the traditional types of early-stage research to diffuse passively down the imaginary research pipeline to later stages of application (10,12,15). The criteria we recommend are not radical or even really new: Most scientists learned about external validity in graduate training. However, unlike internal validity indices, external validity criteria are seldom reported or requested by journals or study sections that review health intervention research.

We are not suggesting that all studies need to be strong on all of these external validity elements but that, in the spirit of transparent reporting (7) and of evidence–based medicine (19), this information should be provided so that readers can make better judgments as to the applicability of a study or review. Neither are we calling for a moratorium on basic or efficacy research: Such studies have made, and will continue to make, fundamental contributions to scientific knowledge. We applaud these accomplishments and the advances that have been made in widespread application of internal validity criteria. We now urge that, for maximum impact of behavioral medicine science, we also need to improve reporting on criteria for external validity. We have not seen the desired progression from efficacy to effectiveness to dissemination research to widespread adoption of evidence-based practices and policies (10,12). As just discussed, there are many reasons for this lack of transfer from one "stage of research to the next"; research quality and relevance are only two of them. More consistent reporting of information affecting both internal and external validity is however, a contributing factor that is under the control of behavioral medicine researchers and a step that we can take now to improve the situation.

Researchers, funders, grant reviewers, and publishers all need to consider dimensions of generalizability and external validity more carefully. The field also needs to explicitly attend more to the appropriate "balance" of internal and external validity issues that are commensurate with the question being asked and the context of a particular study. Enhanced reporting on external validity criteria will allow practitioners to better judge the applicability of research to local situations, reviewers to have information to abstract and synthesize to draw conclusions about generalization, and policymakers to have a more relevant body of evidence on which to rely.

This article extends the call for greater relevance of behavioral medicine research in the recent *Annals* editorial by Klesges, Dzewaltowski, and Christensen (20) by suggesting a specific set of external validity criteria and concrete actions that SBM's flagship journal can take. We encourage *Annals* and other health journals to adopt external validity reporting criteria such as those suggested by Green and Glasgow (15). As a field and an organization, we can and need to do more. Such action would both improve the quality of health research and help achieve the part of the SBM mission statement that involves "the application of knowledge to improve the health and well being of individuals, families, communities, and populations."

In closing, we are reminded of Green and Ottosen's observation that "if we want more evidence-based practice, then we need more practice-based evidence" (21). We encourage *Annals* to take a proactive stance and to seize the opportunity to lead the field of evidence-based health research, not just behavioral medicine, on these issues by promoting evidence-based external validity criteria.

REFERENCES

(1) Zerhouni EA: Translational and clinical science—time for a new vision. *New England Journal of Medicine.* 2005, *353:*1621–1623.

(2) Institute of Medicine, Committee on Quality of Health Care in America: *Crossing the Quality Chasm: A New Health System for the 21st Century.* Washington, DC: National Academy Press, 2001.

(3) McGlynn EA, Asch SM, Adams J, et al.: The quality of health care delivered to adults in the United States. *The New England Journal of Medicine*. 2003, *348:*2635–2645.

(4) Kerner J, Rimer B, Emmons K: Introduction to the special section on dissemination: Dissemination research and research dissemination: How can we close the gap? *Health Psychology.* 2005, *24:*443–446.

(5) Whitlock EP, Orleans CT, Pender N, Allan J: Evaluating primary care behavioral counseling interventions: An evidence-based approach. *American Journal of Preventive Medicine.* 2002, *22*:267–284.

(6) Moher D, Schulz KF, Altman DG, for the CONSORT Group: The CONSORT statement: Revised recommendations for improving the quality of reports. *The Journal of the American Medical Association.* 2001, *285*:1987–1991.

(7) Des Jarlais DC, Lyles C, Crepaz N, TREND Group: Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: The TREND statement. *American Journal of Public Health.* 2004, *94*:361–366.

(8) Jadad AR, Moore RA, Carroll D, et al.: Assessing the quality of reports of randomized clinical trials: Is blinding necessary? *Controlled Clinical Trials.* 1996, *17:*1–12.

(9) Spring B, Pagoto S, Knatterud G, Kozak A, Hedeker D: *Will Behavioral Medicine Randomized Clinical Trials (RCTs) Earn an A?* Boston: Annual Meeting of the Society for Behavioral Medicine, 2005.

(10) Glasgow RE, Lichtenstein E, Marcus AC: Why don't we see more translation of health promotion research to practice? Rethinking the efficacy to effectiveness transition. *American Journal of Public Health.* 2003, *93:*1261–1267.

(11) Glasgow RE, Klesges LM, Dzewaltowski DA, Bull SS, Estabrooks P: The future of health behavior change research: What is needed to improve translation of research into health promotion practice? *Annals of Behavioral Medicine.* 2004, *27*:3–12.

(12) Tunis SR, Stryer DB, Clancey CM: Practical clinical trials. Increasing the value of clinical research for decision making in clinical and health policy. *The Journal of the American Medical Association.* 2003, *290*:1624–1632.

(13) Glasgow RE, Davidson KW, Dobkin PL, Ockene J, Spring B: Practical behavioral trials to advance evidence-based behavioral medicine. *Annals of Behavioral Medicine.* 2006, *31*: 5–13.

(14) Kiefer L, Frank J, Ruggiero ED, et al.: Fostering evidence-based decision-making in Canada: Examining the need for a Canadian Population and Public Health Evidence Centre and

Research Network. *Canadian Journal of Public Health.* 2005, *96(3):*I1–40 following 200.

(15) Green LW, Glasgow RE: Evaluating the relevance, generalization, and applicability of research: Issues in external validity and translation methodology. *Evaluation and the Health Professions* (in press).

(16) Abrams DB, Orleans CT, Niaura RS, et al.: Integrating individual and public health perspectives for treatment of tobacco dependence under managed health care: A combined stepped care and matching model. *Annals of Behavioral Medicine.* 1996, *18:*290–304.

(17) Flay BR: Efficacy and effectiveness trials (and other phases of research) in the development of health promotion programs. *Preventive Medicine.* 1986, *15*:451–474.

(18) Klesges LM, Estabrooks PA, Glasgow RE, Dzewaltowski D: Beginning with the application in mind: Designing and planning health behavior change interventions to enhance dissemination [abstract]. *Annals of Behavioral Medicine.* 2005, *29*(Suppl.):66S–75S.

(19) Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes R: *Evidence-Based Medicine: How to Practice and Teach EBM* (2nd Ed.). Edinburgh, Scotland: Churchill Livingstone, 2000.

(20) Klesges LM, Dzewaltowski DA, Christensen AJ: Are we creating relevant behavioral medicine research? Show me the evidence! *Annals of Behavioral Medicine.* 2006, *31:*3–4.

(21) Green LW, Ottosen JM: From efficacy to effectiveness to community and back: Evidence-based practice vs. practice-based evidence. *Proceedings from Conference: From Clinical Trials to Community: The Science of Translating Diabetes and Obesity Research. National Institutes of Diabetes, Digestive and Kidney Diseases,* 2004.