

IMPORTANCE OF AND CRITERIA FOR EVALUATING EXTERNAL VALIDITY

Russell E. Glasgow, PhD
Lawrence W. Green, DrPH



OVERVIEW

- **Rationale for and Importance of External Validity (EV) for Translation**
- **Proposed EV Reporting Criteria**
- **Editors' Meeting on EV Criteria**

CURRENT SITUATION

- **CONSORT widely used for reporting trials (only 2 of 23 items address EV for non-pharmacologic trials)***
- **Methodological quality ratings used in reviews almost solely internal validity focused**
- **No such widely used criteria for EV, which are critical for translation and dissemination**

*1 of 22 criteria for pharmacologic RCTs

RELATED EFFORTS

- **TREND criteria for non-randomized trials**
- **SQUIRE quality improvement reporting guidelines include number of items on contextual issues**

Des Jarlais DC, et al. *Am J Pub Health* 2004;94:361-366

Davidoff F, Batalden P. *Qual Saf Health Care* 2005;14(5):319-325

DEFINITION AND DIMENSIONS OF EV

External Validity – “Inferences about the extent to which a causal relationship holds over variations in persons, settings, treatments and outcomes.”
(Shadish et al, 2002)

External Validity – “To what populations, settings, treatment variables and measurement variables can this effect be generalized?” (Campbell & Stanley, 1963)

Shadish WR, Cook TD, Campbell DT. 2002. Experimental and quasi-experimental design...Boston: Houghton Mifflin

Campbell DT, Stanley JC. Experimental and quasi-experimental designs for Research. Chicago, IL: Rand McNally. 1966.

Perspective and Context

**“To different degrees, all causal relationships are context dependent, so the generalization of experimental effects is always at issue.”
(Shadish et al, 2002)**



**“Everything is Contextual”
(We made this up)**

USES OF EXTERNAL VALIDITY INFORMATION

Local Practitioners:

To determine if a program or study is relevant to their particular setting (patients, resources, staff, measures, etc.)

Decision and Policy Makers:

To determine the range of conditions and settings across which a given program/policy/product will apply (generalizability)

“Lack of consideration of external validity is the most frequent criticism by clinicians of RCTs, systematic reviews, and guidelines.”

Rothwell PM, *Lancet* 2005(365):82-93

“Where did the field get the idea that evidence of an intervention’s efficacy from carefully controlled trials could be generalized as THE best practice for widely varied populations and settings?”

L.W. Green

**Green LW. From research to "best practices" in other settings and populations
Am J Health Behav 2001; 25:165-78**

DESIRED CHARACTERISTICS OF EV REPORTING CRITERIA

- **Address key conceptual dimensions of EV**
- **Address issues of concern to practitioners and policy makers**
- **Able to be reliably coded**
- **Feasible to report**

Criteria on next slides adapted from Green & Glasgow, *Evaluation and the Health Professions*, 2006;29(1):126-153

EV REPORTING CRITERIA

1. Settings and Populations

- A. **Target Audience:** Are the intended “end users” identified for: 1) adoption (at the setting level, such as worksites, medical offices, etc.) and 2) application (at the individual level)?
- B. **Inclusion and Exclusion Criteria:** Are both 1) inclusion criteria and 2) exclusion criteria (e.g., run-in period, language, comorbid conditions, other treatments, demographic characteristics) reported?
- C. **Participation:** Are there analyses of the participation rate among potential a) settings, b) delivery staff, and c) patients (consumers).
- D. **Representativeness:** Are comparisons reported on the similarity of settings participating to the intended target audience of program settings--or to those settings that decline to participate?
- E. **Representativeness:** Are analyses reported on the similarity and differences between patients, consumers, or individuals who participate vs. either those who decline, or the intended target audience?

EV REPORTING CRITERIA (cont.)

2. Program or Policy Implementation and Adaptation

- A. Consistent Implementation (“Fidelity” or well-delineated scope of adaptations): Are data presented on the range of implementation variations of different program components during the evaluation/study?
- B. Staff Expertise: Are data presented on 1) the level of training or experience required to deliver the program and 2) quality and extent of implementation by different staff?
- C. Program Customization or Adaptation: Is information reported on the ways different settings modified or customized the program to fit their setting (or that no variation was observed)?

EV REPORTING CRITERIA (cont.)

3. Outcomes for Decision Making

- A. **Significance:** Are the outcomes compared to either clinical guidelines (and their intended outcomes) or community preventive services guidelines or other standards of practice for best practices and their associated public health goals?
- B. **Adverse Consequences:** Do the outcomes reported potentially negative effects on quality of life or other outcomes?
- C. **Moderators:** Are there analyses of moderator effects-- including 1) different subgroups of participants and 2) types of intervention staff or settings--to assess robustness vs. specificity of effects?
- D. **Program Intensity:** Are data reported on either or both the total amount of staff time or patient/consumer contact time required?
- E. **Costs:** 1) Are data on the costs presented? If so, 2) are the assumptions made and perspective adopted (e.g., societal, health care payer, patient) and both physical and person costs reported?

EV REPORTING CRITERIA (cont.)

4. Time: Maintenance and Institutionalization

- A. Long-term Effects: Are data reported on longer-term effects, at least 12 months following treatment / intervention?
- B. Institutionalization: Are data reported on the sustainability (or re-invention or evolution) of program implementation at least 12 months after the formal evaluation / study?
- C. Attrition: 1) Are data on attrition by condition reported, and 2) are analyses conducted of a) representativeness of those who drop-out or b) imputation?

EDITORS' MEETING ON EV

- **Editors from 13 leading health and behavior journals**
- **Met in 2006 to discuss above issues and criteria**
- **Meeting sponsored by RWJF, OBSSR, CDC, AHRQ, and held at UNC**

OUTCOMES OF EDITORS' MEETING

- Agreement all EV criteria were important
- Felt all EV criteria, except cost and institutionalization were feasible
- Majority felt that guidelines or principles, rather than mandatory reporting criteria, were best approach



ACTIONS TAKEN BY JOURNALS

- Editorial and recommendations on EV reporting (n = ≥ 6)
- EV criteria for authors (n = ≥ 1)
- EV Criteria for reviewers (n = ≥ 1)

OTHER IDEAS

Highlight article(s) with exemplary EV reporting

Special issue on EV topic

WHAT IS NOT PROPOSED?

- ***NOT*** saying all articles have to be strong on EV criteria
- ***NOT*** saying all articles have to report (but more should than is presently the case)
- ***NOT*** saying that internal validity (or RCTs) are not important (just that we need more of a balance)

FUTURE DIRECTIONS AND NEEDS

- **Document reliability of EV coding criteria**
- **Consider “summary metrics”, composite, or overall EV quality scores**
- **Assistance to practitioners on how to combine with theory and local experience**
- **Evaluate which criteria most strongly related to long-term dissemination success**
- **Revise criteria based on lessons learned**