

Annual Review of Public Health

Making Health Research Matter: A Call to Increase Attention to External Validity

Amy G. Huebschmann,^{1,2} Ian M. Leavitt,³
and Russell E. Glasgow^{2,4}

¹Division of General Internal Medicine, Center for Women's Health Research, School of Medicine, University of Colorado, Aurora, Colorado 80045, USA; email: amy.huebschmann@ucdenver.edu

²Dissemination and Implementation Science Program of Adult and Child Consortium for Outcomes Research and Delivery Science (ACCORDS), School of Medicine, University of Colorado, Aurora, Colorado 80045, USA

³Department of Social and Behavioral Sciences, T.H. Chan School of Public Health, Harvard University, Boston, Massachusetts 02115, USA; email: ianleavitt@g.harvard.edu

⁴Department of Family Medicine, School of Medicine, University of Colorado, Aurora, Colorado 80045, USA; email: russell.glasgow@ucdenver.edu

**ANNUAL
REVIEWS CONNECT**

www.annualreviews.org

- Download figures
- Navigate cited references
- Keyword search
- Explore related articles
- Share via email or social media

Annu. Rev. Public Health 2019. 40:45–63

First published as a Review in Advance on
January 21, 2019

The *Annual Review of Public Health* is online at
publhealth.annualreviews.org

<https://doi.org/10.1146/annurev-publhealth-040218-043945>

Copyright © 2019 by Annual Reviews.
All rights reserved

This article is part of a symposium on Causal Inference and Public Health. For a list of other articles in this symposium, see <http://www.annualreviews.org/toc/publhealth/40/1>.

Keywords

external validity, internal validity, replication, translational research, dissemination and implementation science, D&I, designing for dissemination, D4D

Abstract

Most of the clinical research conducted with the goal of improving health is not generalizable to nonresearch settings. In addition, scientists often fail to replicate each other's findings due, in part, to lack of attention to contextual factors accounting for their relative effectiveness or failure. To address these problems, we review the literature on assessment of external validity and summarize approaches to designing for generalizability. When investigators conduct systematic reviews, a critical need is often unmet: to evaluate the pragmatism and context of interventions, as well as their effectiveness. Researchers, editors, and grant reviewers can implement key changes in how they consider and report on external validity issues. For example, the recently published expanded CONSORT figure may aid scientists and potential

program adopters in summarizing participation in and representativeness of a program across different settings, staff, and patients. Greater attention to external validity is needed to increase reporting transparency, improve program dissemination, and reduce failures to replicate research.

INTRODUCTION: WHAT IS THE SIGNIFICANCE OF EXTERNAL VALIDITY?

Public health and clinical practitioners are “drowning in information but starved for relevance” (39, p. 301). This disconnect between the immense quantity of research data available and its limited relevance to practitioners is a concerning indictment of the research establishment’s lack of attention to external validity. Numerous reviews have documented that criteria and information related to external validity are underreported in most, if not all, areas of science (27, 38, 39, 43, 50, 55, 65). It is even more unequivocally documented that few evidence-based interventions translate into practice or policy (2, 3, 9, 12, 21, 39, 40). Despite recognition of this problem, and exhortations to address it for more than a decade, external validity is still underemphasized relative to internal validity for most journals, funders, reviewers, and researchers (2, 39). In contrast, perceived lack of relevance is the primary stated reason that practitioners do not adopt research-based interventions (39, 55, 69). Review of compendia of evidence-based programs such as the Research-Tested Intervention Programs (RTIPS) and the National Registry of Evidence-Based Programs and Practices (NREPS) reveals that very few programs report sufficient information for investigators in potential future settings to judge the applicability to their sites and staff (13, 71). The problem includes a lack of clarification of stakeholder perspectives on what constitutes return on investment, but it even goes beyond this issue because sites cannot identify the breadth and conditions of application, the impact of context, or the cost in terms of staff training, delivery time, and burden (57, 72).

We propose that the lack of attention to external validity is a key contributor to another major problem within health research: the failure to replicate findings (43, 45, 46). Failure to replicate is reaching crisis level and is arguably even more problematic in applied T3/T4 behavioral and public health research than in basic research, owing, at least in part, to the greater complexity in T3/T4 research designs that embrace multilevel settings and participants with more comorbidities that lead to greater contextual variability in intervention response than in highly controlled efficacy studies (43, 45, 46). Without adequate reporting on the context, conditions, settings, inclusion/exclusion criteria, and participation data at all socioecological levels of a study (39, 55, 59), it is impossible to determine if inconsistent results across studies are due to true failure to replicate or if they occur because effects are conditioned on important contextual factors (19, 23, 55, 80). We review and recommend reporting criteria that can feasibly address this research replication crisis.

Persisting with science-as-usual practices will continue to hamper the translation of research to practice. However, several recent efforts to use contextual perspectives to help plan, adapt, report, evaluate, review, and disseminate research results are intended to produce a noticeable public health impact (4, 33, 43, 65, 80). Our goal is not to denigrate efficacy research or valued approaches and methods to enhance internal validity, such as randomized controlled efficacy trials (37, 39). However, as detailed below, research that maximizes internal validity at the expense of external validity is not sufficient to create programs, policies, or products that will translate into practice or that will generalize to diverse or complex, low-resource settings. Throughout this review, we provide suggestions for research design and reporting that are relevant to all phases of health services research. We explicitly distinguish how external validity should be considered

differently for (a) explanatory (international term for efficacy in the United States) (56) research studies focused on proving proof-of-concept, determining causality, or testing the early efficacy of programs—or specific interventions within programs—that could ultimately be translated into practice and (b) pragmatic (56) dissemination and implementation (D&I) studies that evaluate the implementation of an evidence-based program or policy into practice.

For the latter category of pragmatic D&I studies/trials, we provide some additional recommendations that go beyond the types of reporting needed for explanatory trials, as external validity factors are particularly critical for these real-world tests. In summary, this article provides a combination of (a) discussion of key conceptual and methodological issues related to external validity; (b) review of the underreporting on external validity, and likely reasons for this state of affairs; (c) identification of approaches to enhance planning and reporting on external validity; and (d) concrete recommendations for all those who influence the conduct of research and its later translation, including grant funders, researchers, journal editors, clinical and public health practitioners, and policy makers.

EXTERNAL VALIDITY

What It Is and the Current State of the Literature on External Validity Reporting

External validity was defined early by Cronbach & Shapiro as the degree to which an observed causal relationship between a treatment and an outcome may generalize to four separate domains of variation: units/people, treatments, outcomes, or settings (18). Cronbach & Shapiro's work chiefly addressed educational interventions as the treatments delivered by trained teachers in classroom settings (18). In contrast, public health studies typically utilize intervention staff members to deliver a specific behavioral intervention in various community settings, including homes, schools, community health centers, recreation centers, and churches; despite these variations, both educational research and public health research embrace the complexities of implementing programs in real-world settings (70). A more in-depth review of external validity is outside the scope of this article, but may be referenced elsewhere (1, 17, 18, 39). To the extent that an experiment is less representative of the real-world of participants, practitioners, interventions, outcome measures, and settings where it is expected to apply, the external validity is diminished (17, 39).

A broad gap currently exists between standard recommendations for external validity reporting (2, 3) and current research practice; this section reviews the theoretical and historical reasons for underreporting on external validity. The existing literature provides numerous examples of the failures of researchers to document key areas of external validity (14, 33, 42, 50, 57, 72, 78). For example, several studies have documented the lack of reporting on representativeness of participants and settings for research interventions delivered in public health and clinical health care settings (1, 27, 33, 42, 50, 53, 78). Furthermore, two recent reviews found that most randomized controlled studies of interventions across several diverse fields, such as mental health and diabetes, were not representative of the general population to which the intervention would be applied, and those enrolled were selected to be typically younger, healthier (i.e., fewer comorbidities and less severe disease), and of a higher socioeconomic status than the general population (47, 73). One possible reason for the limited attention to reporting data related to external validity is that these data were not part of the standard reporting requirements from CONSORT and other sources. Two 2017 publications (4, 65) added new reporting requirements for key areas of external validity, including eligibility criteria for intervention settings/staff, and the generalizability of the study findings according to several elements: the participants studied, the intervention characteristics, the intervention staff, and the intervention settings.

Other studies have demonstrated a lack of reporting on numerous other factors that relate to future dissemination, including the relative paucity of data on implementation factors such as cost and other resources necessary to implement the program (42, 57, 72, 78). Perhaps most importantly, data on the maintenance or sustainability of the program are quite lacking: Two reviews found that fewer than 10% of studies reported on factors related to program sustainability (50, 57). Although recommendations on the importance of prioritizing these elements of external validity for translation-focused studies have been made for nearly 40 years, a prioritization of internal validity factors at the expense of external validity persists (25, 39, 75).

More recently, several publications have highlighted the need for applied research to pay greater attention to external validity from a slightly different perspective: to address problems of research replicability (4, 39, 43, 45, 46, 65). Indeed, part of the failure to disseminate evidence-based research programs is because these programs are often not replicable when translated into real-world, nonacademic settings (21, 37, 39, 40). Green & Nasser (39) have highlighted three key areas where “science as usual” is failing public health settings:

- Research is too seldom tested with realistic expectations of the staff resources and other organizational infrastructure of future adopters: The research staff training, time, and/or resources are often greater than what resource-constrained public health practice settings can muster. (We would also frame this issue as a failure to test programs with representative staff and delivery conditions, including the use of unrepresentative levels of supervision, feedback, and measurement.)
- Research is too seldom tested with representative participants. Participants are often less diverse than in the real world, in terms of both the spectrum of healthy and diseased individuals as well as the cultural, demographic, and health literacy differences that may influence program effectiveness.
- Research interventions are highly controlled and inflexible to local site adaptation.

Why Has This Happened?

The three areas listed above may result from a historical research methods bias to prioritize internal validity over external validity for research experiments (39, 75). For example, in their classic text on research design, Campbell & Stanley (8) emphasized the importance of prioritizing internal validity over external validity, noting that internal validity is the “basic minimum without which any experiment is uninterpretable” (p. 5). Campbell & Stanley also highlighted the tension between internal validity and external validity, noting that emphasizing internal validity may jeopardize external validity, and vice versa. For explanatory trials that seek to identify mechanistic predictors of an outcome, it is often considered acceptable by clinical researchers to emphasize internal validity at some expense to external validity, as the fundamental mechanisms of physiology or other biological behavior may be “diluted” in populations with differential gene expression owing to disease or other factors (25). However, for pragmatic D&I studies that seek both to prove an intervention’s effectiveness and to identify implementation strategies necessary for interventions to work in the real world, it is critical to avoid the three pitfalls described above (39). Avoiding these pitfalls is particularly important for pragmatic D&I studies, as compared with explanatory studies that are seeking to demonstrate the effectiveness of a program for the first time (Table 1). In addition, pragmatic D&I studies should be designed and should report data in a way that addresses the following critical areas of external validity: (a) context of the participating units,

Table 1 Contrast of research design methods commonly utilized in explanatory (efficacy) research with methods that are necessary to enhance external validity

Element of research design	Common explanatory (efficacy) clinical trial methods that emphasize internal validity	Recommendations for explanatory research studies to optimize external validity with minimal impact on internal validity	Recommendations for pragmatic studies to optimize external validity with limited impact on internal validity
Participating units (populations/settings)	Uses exclusion criteria to identify homogeneous participants and settings that have the least noise/complexity and are most likely to respond to treatment; often not representative of intended target audience	Uses exclusion criteria to identify relatively homogeneous participants from settings that reflect the target audience Compares the similarity of participants, settings, and delivery staff with the intended target audience Analyzes the differences between participants and those who decline	Most relevant for pragmatic trials: <ul style="list-style-type: none"> ■ Enroll participants/settings that represent intended target audience ■ Assess the reach (participation rate and representativeness) among potential participants/settings, and compare with intended target audience ■ Analyze differences between participants and those who decline ■ Investigate how/why differences occur
Development of intervention (or program/policy)	Developed from prior evidence published in the literature or basic science research	Developed from prior evidence published in the literature If programs are meant for downstream translation, they should also draw on input from stakeholders in terms of the feasibility, acceptability, and appropriateness of the program studied	Developed from prior evidence published in the literature or existing guidelines Multilevel input from stakeholders: <ul style="list-style-type: none"> ■ Feasibility to deliver program given typical organizational infrastructure and resources ■ Acceptability and feasibility of program to meet needs of all relevant stakeholders (e.g., participants, staff, health system leaders) ■ Potential sustainability of program
Implementation and adaptation of intervention (or program/policy)	Fidelity to intervention is optimized by strict adherence to a predefined rigid protocol and delivered by a limited number of highly trained and carefully supervised intervention staff	Fidelity to intervention is optimized by using strict protocol adherence and a limited number of highly trained and carefully supervised intervention staff The type of intervention staff utilized should be feasible for downstream target settings	Fidelity to core components of program is emphasized, but adaptation to local context is allowed (and reported) to ensure flexibility to both site needs and participant needs Site needs include priorities, change capability, culture, context, local site staff experience, training, workflow, and resources Participant needs include individual-level characteristics such as culture, literacy, motivation, and family and social factors

(Continued)

Table 1 (Continued)

Element of research design	Common explanatory (efficacy) clinical trial methods that emphasize internal validity	Recommendations for explanatory research studies to optimize external validity with minimal impact on internal validity	Recommendations for pragmatic studies to optimize external validity with limited impact on internal validity
Outcomes for decision making	Focus on a single primary outcome, usually a biological intermediate outcome to prove efficacy	Primary outcome may be a biological intermediate outcome to prove efficacy; if delivered in health settings, cost and resource demands, feasibility, and acceptability should be reported	Multiple outcomes (to assess generalizability and context): The primary outcome(s) should directly impact population health and be relevant to participants and the settings to which results are intended to apply
Maintenance/sustainability	Not assessed	Not typically necessary to assess actual maintenance during an explanatory trial, but intent to maintain (or adapt) should be reported and the potential downstream sustainability should be considered if the program is ultimately intended for clinical use	Utilize designing for dissemination principles to inform the dissemination strategy: <ul style="list-style-type: none"> ■ Identify systems needed for dissemination of program ■ Obtain multilevel stakeholder input throughout the project: across project design, implementation, and dissemination stages ■ Develop a value proposition for the product for future adopters ■ Identify communication channels for dissemination ■ See Table 2 for further details

including populations and settings/sites (80); (b) intervention development; (c) implementation and adaptation of the intervention; (d) implementation strategies used (and their adaptations); (e) outcomes of relevance to populations; and (f) factors that will promote or inhibit program sustainability (**Table 1**).

Table 1 summarizes the status quo and future recommendations for addressing these six key elements of health outcomes research design. We contrast how the tension between internal validity and external validity is typically handled by traditional explanatory or efficacy-based clinical trials (39, 56) with our recommendations for more optimally addressing this concern in explanatory trials and in pragmatic D&I study designs. We do consider that explanatory research seeks to maximize homogeneity across settings, participants, and staff, while also tightly controlling intervention condition(s) and minimizing other extraneous factors, in order to assess treatment efficacy and/or mechanisms optimally. In contrast, pragmatic studies seek to produce findings of relevance to real-world settings and, with some exceptions, to conduct research under real-world conditions using heterogeneous samples, settings, and delivery conditions (5, 39, 56, 62). We note that **Table 1** presents rather extreme cases for pedagogical purposes; in reality, no trial is completely explanatory or totally pragmatic on all dimensions (56). Both explanatory and pragmatic studies should consider designing for dissemination (D4D) principles (**Table 2**) to enhance the potential downstream sustainability of the program(s) studied (5, 62).

Table 2 Designing for dissemination: processes, outcomes, and products

Principles of designing for dissemination	Explanation
Processes	Engaging a range of relevant stakeholders at multiple levels (citizens or patients, delivery agents, supervisors, administrators, policy makers) from a project's beginning to end Identifying relevant theoretical models and communication channels for dissemination
Outcomes	Assessments of and methods to increase readiness for dissemination: <ul style="list-style-type: none"> ■ pragmatic outcome measures ■ data on impact, burden, and costs that convey return on investment information to future potential program adopters
Products	<ul style="list-style-type: none"> ■ Dissemination and step-by-step adaptation guides (also known as tool kits) ■ Business models ■ Dissemination strategies

Table adapted from References 7 and 64.

What Can We Do About It?

The Pragmatic Explanatory Continuum Indicator Summary (PRECIS) tool (79) and the more recently revised and updated PRECIS-2 tool (accessible at <http://www.precis-2.org>) (56) have been developed to guide researchers to explicitly consider nine factors (originally 10) related to external validity when designing a study. PRECIS-2 visually summarizes the extent to which a trial is more versus less pragmatic for each of the nine factors, recognizing that no study is entirely pragmatic or entirely explanatory. We highlight key differences between a more explanatory trial and a more pragmatic trial that have been posted on the PRECIS-2 website (**Figure 1**).

Figure 1 illustrates some potential differences that may be expected between a study that is generally explanatory (low PRECIS-2 scores) and one that is generally more pragmatic (high PRECIS-2 scores). As shown, explanatory studies often have low scores for the flexibility domain, as the staff in the trial are required to follow the intervention protocol quite strictly to ensure high fidelity and strong internal validity. However, as different settings may need to adapt programs to fit their local contexts, pragmatic D&I studies typically have much higher scores for flexibility. To ensure adequate internal validity, while allowing optimal local site flexibility, experts have suggested that pragmatic researchers should allow adaptation of intervention components such as mode of delivery (10, 77), while ensuring fidelity to the core components of an intervention/program. **Figure 1** also illustrates high versus low scores on other specific PRECIS-2 domains, such as recruitment, settings, and organization. We strongly recommend inclusion of a PRECIS-2 graphic when proposing or reporting on a study to allow both reviewers and potential adopters to interpret the relative strengths and weaknesses of the study from an external validity perspective.

Context Matters

One of the fundamental principles of external validity is that context matters (24). In this era of personalized medicine, great attention is paid to variations in our genetic code, but there is

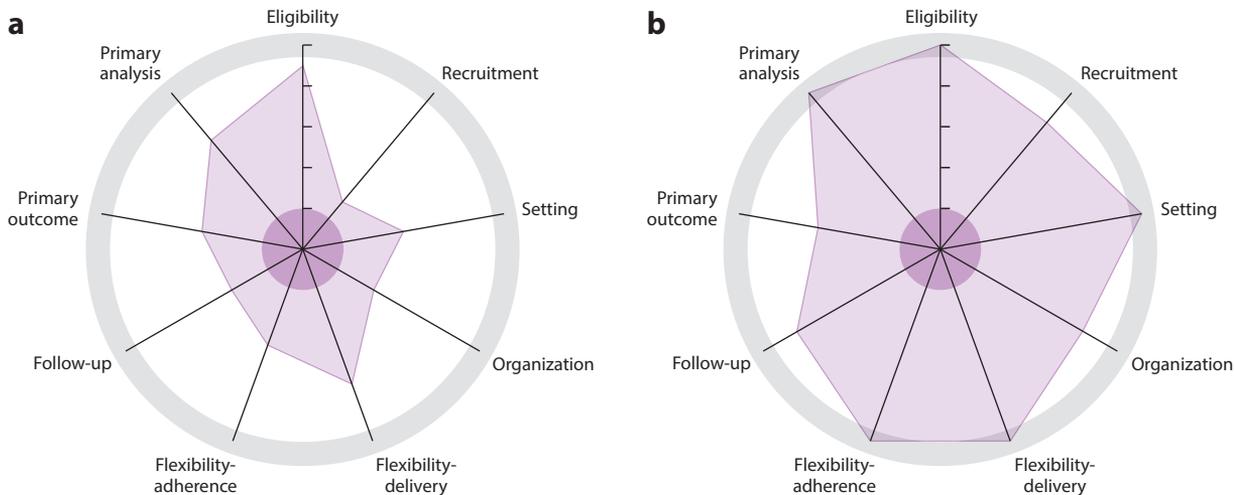


Figure 1

Contrasting examples of Pragmatic Explanatory Continuum Indicator Summary (PRECIS-2) diagrams from (a) an explanatory study with limited attention to external validity and (b) a pragmatic study with major attention to external validity. Note that the explanatory study in panel *a* has a low PRECIS-2 score for organization, which means the resources for this trial are not common in real-world settings. In addition, the explanatory study has a high score for eligibility but low scores for recruitment and setting: The results are thus likely to be relevant only to patients in the types of settings studied. In contrast, this study is likely to have excellent generalizability to other patients and settings on the basis of the high PRECIS-2 scores for eligibility, recruitment, and setting, and the high scores for organization suggests that most settings could deliver this program and that real-world implementation efforts are likely to find the same results as in the study. Abbreviation: PRECIS-2, Pragmatic Explanatory Continuum Indicator Summary-2. Panels *a* and *b* adapted from original diagrams available at <https://www.precis-2.org/>.

still insufficient attention to the context affiliated with one's zip code (or as Alice Ammerman has quipped, "Neighborhood-omics") (A. Ammerman, personal communication; 10, 11, 20, 28, 63). As housing areas in the United States are often segregated by social class, the behavioral health and social determinants of health that vary by zip code are related to many factors, including household income and levels of community health resources, education, social support, and social capital in the community and home environments (11, 20, 28, 63). When location and race are considered together, health disparities tend to be even more prominent (60). Given the strong linkages among zip code, wealth, behavioral health, and social determinants of health and health outcomes/inequities, it is important to consider context in the design and evaluation of research.

Context is also important in terms of the settings in which research is conducted (74, 80). For example, results can be quite different across academic medical settings or university mental health clinics, with residents/staff following highly detailed intervention protocols versus a Veterans Affairs center or community health center with fewer resources and busy staff balancing competing demands (80). To measure context adequately, mixed methods are often necessary, as further described in a review by Leviton (55). Additionally, for researchers who are specifically planning to disseminate research programs, it is important to consider the D4D literature to incorporate those principles into the research design of both explanatory trials and pragmatic trials (5, 7). This section summarizes methods to measure context and further addresses the importance of context for research design and for the evaluation and synthesis of research in systematic review articles.

As the adage often attributed to Albert Einstein goes, “Not everything that matters can be measured, and not everything that is measured matters” (22, p. 170). Several approaches have been developed to capture key elements of context, even when fully measuring context remains an elusive goal (24, 55). The frameworks that have been developed to capture context are part of the emerging D&I science field that seeks to embrace methods to recognize and utilize context for optimal program implementation and dissemination (6). A website compiling most of the D&I frameworks, their associated constructs, and relevant measures for each may be accessed at <http://www.dissemination-implementation.org/>. Researchers may use the search terms on this website to identify frameworks of particular relevance to one’s project. Several frameworks address external validity factors, including PRECIS-2; the Practical Robust Implementation and Sustainability Model (PRISM); the Research, Effectiveness, Adoption, Implementation and Maintenance (RE-AIM) model (34); the Consolidated Framework for Implementation Research (CFIR) (19); and others (36, 41, 49). Here we compare some of the most highly utilized frameworks and discuss how they address external validity (56).

As previously mentioned, the PRECIS-2 framework directs attention to certain contextual factors, such as the representativeness of participants relative to the target population, and the relevance of these factors to the organizational resources and infrastructure of future site/staff adopters. PRECIS-2 has been used primarily to inform research design (although see Gaglio et al. for its additional use in literature review and reporting) (26), but other frameworks have informed both the design and the evaluation of research programs, such as the RE-AIM model (27, 31, 34). More recently, the PRISM framework (**Figure 2a**) has been developed, both mapping onto RE-AIM outcomes and addressing contextual factors more explicitly (23). PRISM identifies specific contextual aspects of the organization delivering the program, the intervention itself (and how it fits within the multilevel context), the patient/participant receiving the program, and how the external environment and factors related to sustainability and infrastructure influence the RE-AIM outcomes. The widely used CFIR framework was developed by Damschroder and colleagues (19) to assess several contextual factors in the following domains: intervention characteristics, inner setting of the organization delivering the intervention, the outer setting of patient factors and external policies, staff characteristics, and the phase of the implementation process (**Figure 2b**). Summaries of applications of these models, recommendations for their use, and resources are available at both the RE-AIM (<http://www.re-aim.org>) and CFIR (<http://www.cfirguide.org/>) websites.

As outlined in **Table 1**, key elements of planning for future dissemination need to be included in the research design phase for both explanatory and pragmatic studies for their findings to be applicable to a variety of settings and for future use (7, 62). For pragmatic research, using a D&I framework such as PRISM, Knowledge to Action (36), Replicating Effective Programs (49), or CFIR (19) allows researchers to plan for and evaluate the role of contextual factors (34, 41, 68). Successfully implementing programs across a few diverse settings is a potential step toward broader dissemination. As highlighted by Steensma et al. (76), many evidence-based public health programs that are locally effective have failed to disseminate broadly. Thus, utilizing a D&I framework is helpful for implementation and replication but may not be sufficient for broad dissemination. This conundrum had led to the new D4D movement (7). The key role that external validity plays in D4D is highlighted by Brownson and colleagues (7): “We need to better understand how to better design interventions with the elements most critical for external validity in mind, addressing these issues during early developmental stages and not near the end of a project” (p. 1693).

D4D is still an emerging field but has recently been defined as a set of processes that are considered and activities that are undertaken throughout the planning, development, and evaluation of an intervention to increase its dissemination potential (66). Brownson and colleagues (7) have proposed several D4D processes, outcomes, and products that research designers can consider

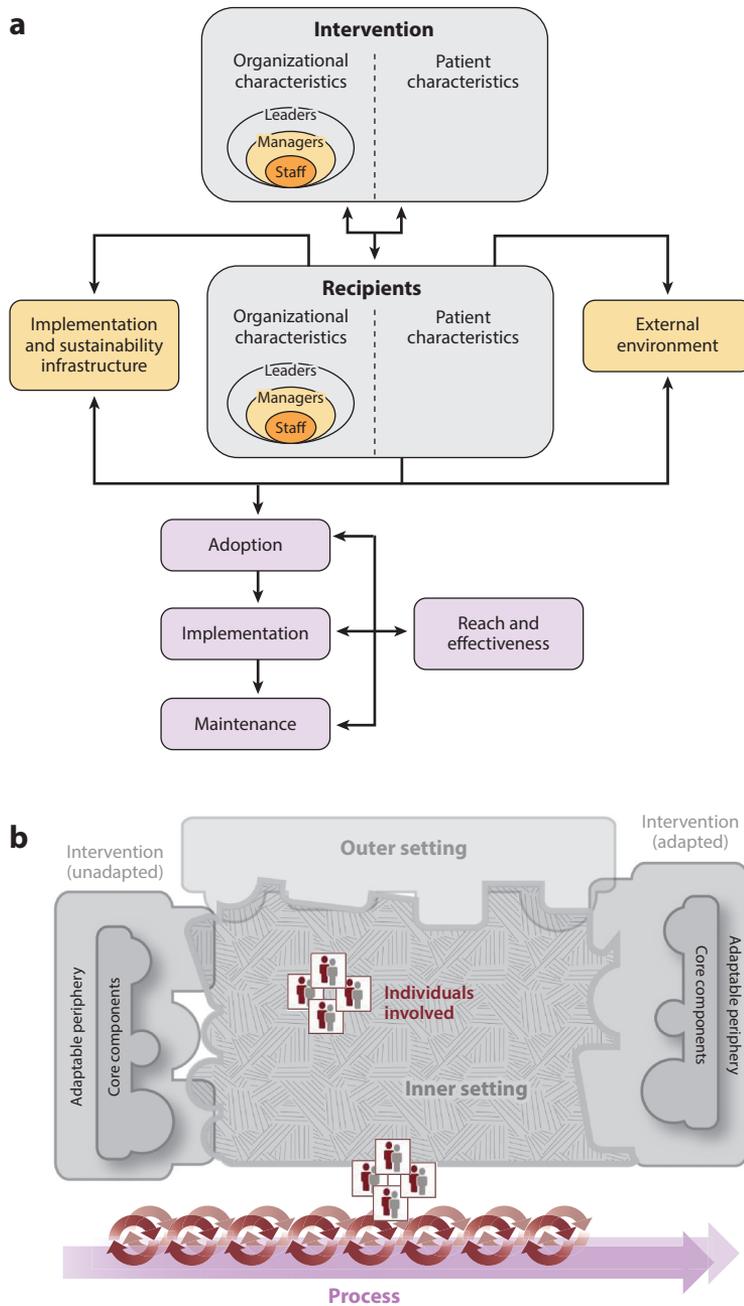


Figure 2

Sample frameworks that may be used to consider contextual factors related to external validity. (a) Practical Robust Implementation and Sustainability Model (PRISM); (b) Consolidated Framework for Implementation Research (CFIR).

(Table 2). In contrast with D&I frameworks that are developed explicitly for pragmatic trial researchers, the D4D movement intends to influence both explanatory and pragmatic trial designs, thus seeking to move explanatory trialists to consider more principles of pragmatic trials to ensure that the interventions developed would be relevant to key stakeholders and end users. Of course, it may be premature for explanatory studies to assess certain D4D outcomes and products, such as cost data and dissemination and adaptation guides; however, D4D processes such as multilevel stakeholder engagement are relevant for both explanatory and pragmatic research (Table 1). Recent reviews of stakeholder engagement (15, 16, 58) provide a more thorough explanation of this key D4D process.

In addition to considering external validity in the design and reporting of original research, researchers should consider factors related to external validity and generalizability in systematic reviews that evaluate the literature to identify evidence-based research interventions for broader implementation and dissemination (29, 32, 51, 54, 57, 72). To this end, innovations in the design of systematic reviews have permitted investigators to assess the relative external validity and pragmatism of research interventions, rather than assessing only the relative effectiveness of interventions. Since 2009, six review articles have assessed external validity or pragmatism (29, 32, 51, 54, 57, 72). Of these, five used the PRECIS-2 model (56) (or the first-generation PRECIS criteria), and half of the articles also assessed the reporting of additional contextual external validity factors related to the RE-AIM framework (Reach and Representativeness of populations studied, Effectiveness, Adoption by sites and site staff, Implementation factors such as cost of the program, Maintenance of the program after grant funding ends). The most recent of these six reviews, led by Luoma, Leavitt, and colleagues (57), also assessed the relative effectiveness of interventions among programs in the top tier of pragmatism by PRECIS-2 scoring in order to identify a subset of programs that are ripe for dissemination owing to their comparatively high ratings of pragmatism and proven effectiveness. We propose that review articles that intend to identify interventions for future translation should uniformly incorporate comparative PRECIS-2 rankings of pragmatism and external validity factors in tandem with reporting relative effectiveness. Furthermore, such pragmatic review articles should also report on external validity factors related to the staff, site, and patient population to help future potential adopters identify if an intervention is feasible and if it fits the priorities, needs, and resources of their site, staff, and potential participants. Examples of this proposed approach may be found in the review articles published by Luoma, Leavitt, and colleagues (57) and by Sanchez et al. (72).

Additional Future Opportunities

Another critical step to enhance the external validity of health research is to utilize sites/settings that are relatively similar to the intended sites/settings for which a given intervention is likely to be utilized. As Green (37) has opined, “If we want more evidence-based practice, we need more practice-based evidence” (p. 406). As public health centers and medical practices develop a digital footprint for their care processes, the number of electronic health solutions that can integrate research into practice in a way that is feasible and acceptable to patients, clinics, and public health systems increases (35; <https://apporchard.epic.com/>). An underresearched area of translating research into practice is the integration of intervention protocols into electronic health records. For example, behavioral counseling protocols and community referral actions to address social determinants of health can be embedded into “smart phrase” templates in an electronic health record, and this protocol serves at least five purposes:

- It creates a D4D product of a health record—an embedded protocol that could be broadly disseminated;
- It enhances staff fidelity to recommended evidence-based procedures;
- It promotes efficient documentation of activities performed by staff, thus minimizing the time/cost of recording and tracking the program and improving its potential for dissemination;
- It provides data for audit, feedback, and recognition for quality improvement and incentive programs; and
- It communicates with patients’ primary health teams, supervisors, and accountable care or public health agencies about care and services provided.

In addition to integrating evidence-based procedures into real-world practice, the potential is emerging for researchers to upload objective patient data [e.g., physical activity, glucose, and/or blood pressure from wearable devices (61)] and patient-reported data [e.g., on health behaviors, mental health, and social determinants of health (52)] to enhance population health monitoring and to prompt outreach to at-risk patients prior to already scheduled visits.

A Call to Action

Throughout this article, we have (a) summarized key issues in external validity; (b) reviewed the low level of reporting on external validity and generalizability in the published literature; (c) provided a rationale for increasing attention to external validity concerns; and (d) provided some specific resources and tools for scientists, journal editors, and funders to apply. In this section, we aim to help move the field from knowledge to action by enumerating steps to enhance the external validity of health outcomes research and to provide tools to address the crisis of failure to replicate—both in terms of the scientific failure to replicate studies and in terms of the failure to replicate successful programs in public health practice and policy. We specify a call to action for researchers, journal editors, funders, and policy makers.

1. For researchers conducting explanatory studies, we call for additional consideration of the elements of study design related to external validity, including the use of study designs that consider the potential for future dissemination, even prior to proving efficacy (**Tables 1 and 2**).
2. For researchers conducting pragmatic studies, we call for the use of PRECIS-2 for both study reporting and study design to ensure the consideration of key external validity factors and to design studies with frameworks, such as PRISM or CFIR, that explicitly address and facilitate reporting on the role of contextual factors (**Figures 1 and 2**), as recommended by the Standards for Reporting Implementation studies (65). In addition, we emphasize the need to design trials with the goals of broad dissemination, population health, and health equity in mind. To that end, it is key to utilize processes, assess outcomes, and develop products that will facilitate dissemination and broad application across diverse settings and populations (**Table 2**).
3. For journal editors and reviewers, we emphasize the importance of holding researchers accountable to report and editors to provide space to publish on the internal and external validity factors that influence researchers’ findings and to require an expanded CONSORT diagram (or alternative presentation of these types of data) to adequately inform future adopters about the relevance of this work to their practice and to inform researchers about factors potentially related to the replication of findings.

4. For funders, we call for requirements for grant applications to provide data or descriptions on the applicability of the proposed research to potential future adopters and to diverse settings, including low-resource systems and populations, by way of a PRECIS-2 diagram (**Figure 1**) or similar method.
5. For policy makers, we call for further promotion of the need for researchers and publishers to collect and report data on external validity in a way that public health organizations may readily interpret, as these data are necessary to inform decisions for real-world application.

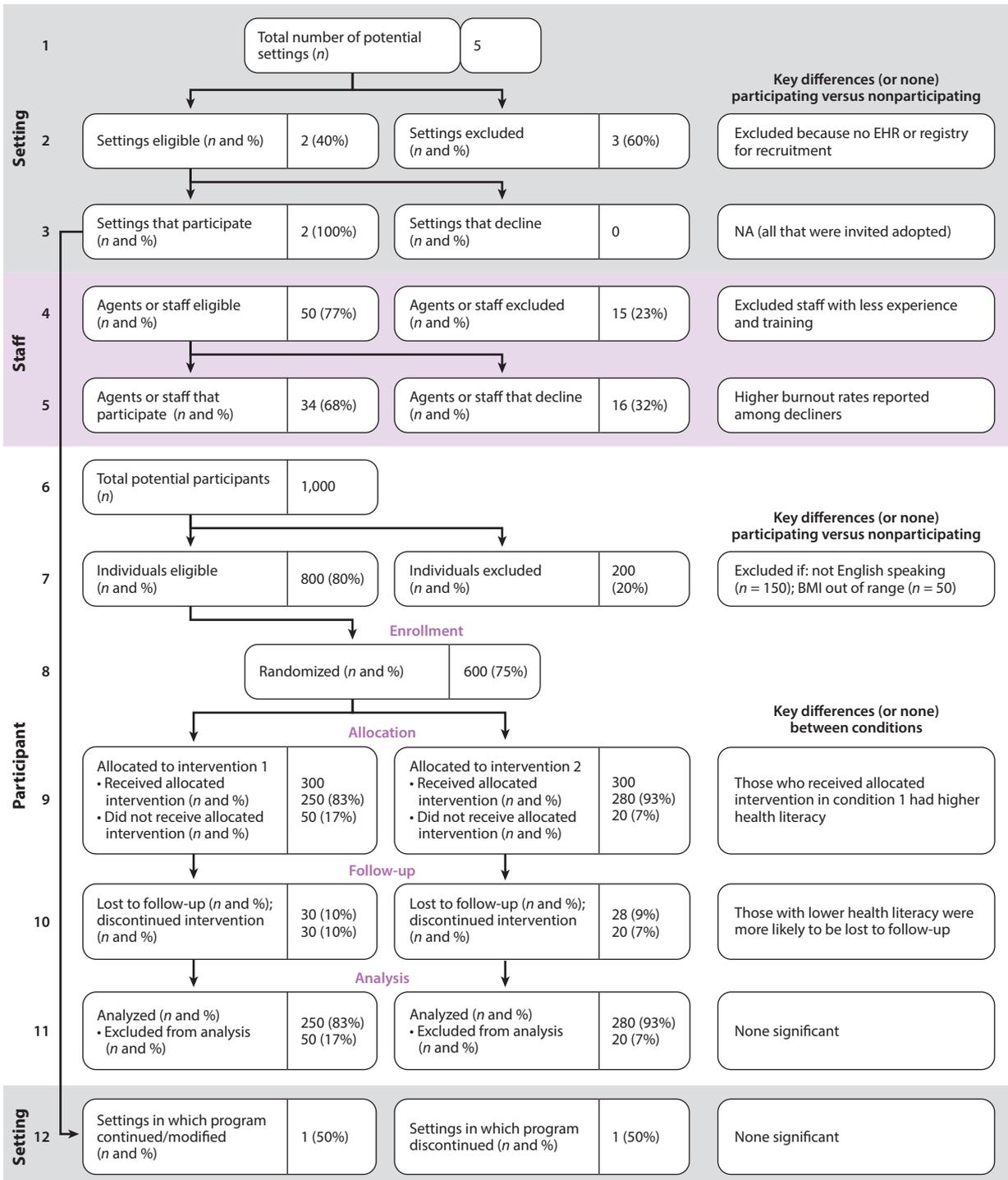
CONCLUSIONS

This article reviews and highlights what others have clearly documented over the last two decades: the absence of reporting on key external validity factors that are necessary both to inform adoption by health system and public health decision makers as well as to allow scientists to replicate each other's work (14, 27, 38, 39, 42, 43, 45, 46, 50, 55, 57, 65, 72). Researchers must pay greater attention to the context of their findings across multiple levels of participants, settings, and systems (**Table 1**) in order to compile answers to what is seemingly the ultimate use question: Which programs/policies work for which participants in which settings in order to produce which outcomes, under which circumstances, at what cost, and with what means?

Because the needle has not moved substantially on these issues despite decades of others citing these concerns, we have specifically called researchers and other research stakeholders to action. This review also offers research stakeholders (e.g., researchers, editors, funders, and policy makers) several specific tools for assistance with this endeavor. To improve the design and reporting of original research, researchers and funders could consider frameworks such as PRECIS-2 (**Figure 1**), PRISM (**Figure 2a**), CFIR (**Figure 2b**), and several recent review articles that address this topic in further depth (44, 48, 67). We have also summarized the value of using D4D processes of ongoing multilevel stakeholder engagement for all research and of developing user-friendly implementation and adaptation guides for programs that aspire to downstream translation, be it explanatory or pragmatic research (**Table 2**). For pragmatic research (30), investigators need to utilize additional processes, outcomes, and products that promote future dissemination (**Table 2**). In addition, we have discussed the use of tools such as the expanded CONSORT diagram (**Figure 3**) to encourage more transparent reporting of original research and have developed a user's guide on how to do this, including a fillable expanded CONSORT diagram, which may be accessed online (<https://goo.gl/jw3HXa>). We have also shown how systematic reviews can assess elements of external validity, in addition to evaluating relative effectiveness and identifying programs that are both pragmatic and effective. To guide interested researchers in this methodology, manuscripts using this approach have been published; an online user's guide (<https://goo.gl/dYjc5Y>, located in the "Guides and Tools" tab) provides the review criteria used in prior studies to rate manuscripts according to PRECIS-2, RE-AIM, and other external validity-related factors. Finally, researchers should seek to study pragmatic interventions within sites and settings that are representative of the future program adopters: communities of public health practice.

It is not necessary to sacrifice internal validity to anticipate, design for, and report on external validity issues; however, we do need to rectify the well-documented current imbalance (**Table 1**). As PRECIS-2 demonstrates visually, there is a continuum of external validity and public health relevance for both explanatory trials and pragmatic trials (**Figure 1**). Ultimately, it is a matter of carefully considering our priorities. If we are serious about translating more than 14% of evidence-based research into practice, and if we would like it to take fewer than 17 years for these findings

**Contextual data
and participation rates**



(Caption appears on following page)

Figure 3 (Figure appears on preceding page)

An expanded version of the Consolidated Standards of Reporting Trials (CONSORT) diagram that increases reporting on factors related to external validity. Abbreviations: BMI, body mass index; EHR, electronic health record; *n*, number of participants; NA, not applicable. Figure adapted with permission from the American Journal of Preventive Medicine (Glasgow RE, Huebschmann AG, Brownson RC. 2018. Expanding the CONSORT figure: increasing transparency in reporting on external validity. *Am. J. Prev. Med.* 55:422–30) (33).

to be translated (3), then we need to respond collectively to this call to action. Failure to act will constitute a waste of time, money, effort, and opportunities to improve public health.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

LITERATURE CITED

1. Ahmad N, Boutron I, Dechartres A, Durieux P, Ravaud P. 2010. Applicability and generalisability of the results of systematic reviews to public health practice and policy: a systematic review. *Trials* 11:20
2. Ammerman A, Smith TW, Calancie L. 2014. Practice-based evidence in public health: improving reach, relevance, and results. *Annu. Rev. Public Health* 35:47–63
3. Balas EA, Boren SA. 2000. Managing clinical knowledge for health care improvement. *Yearb. Med. Inform.* 1:65–70
4. Boutron I, Altman DG, Moher D, Schulz KF, Ravaud P, CONSORT NPT Group. 2017. CONSORT statement for randomized trials of nonpharmacologic treatments: a 2017 update and a CONSORT extension for nonpharmacologic trial abstracts. *Ann. Intern. Med.* 167:40–47
5. Brown CH, Curran G, Palinkas LA, Aarons GA, Wells KB, et al. 2017. An overview of research and evaluation designs for dissemination and implementation. *Annu. Rev. Public Health* 38:1–22
6. Brownson RC, Colditz GA, Proctor EK, eds. 2018. *Dissemination and Implementation Research in Health: Translating Science to Practice*. New York: Oxford Univ. Press. 2nd ed.
7. Brownson RC, Jacobs JA, Tabak RG, Hoehner CM, Stamatakis KA. 2013. Designing for dissemination among public health researchers: findings from a national survey in the United States. *Am. J. Public Health* 103:1693–99
8. Campbell DT, Stanley JC. 1963. *Experimental and Quasi-Experimental Designs for Research*. Chicago: R. McNally
9. Chambers DA, Glasgow RE, Stange KC. 2013. The dynamic sustainability framework: addressing the paradox of sustainment amid ongoing change. *Implement. Sci.* 8:117
10. Chambers DA, Norton WE. 2016. The adaptome: advancing the science of intervention adaptation. *Am. J. Prev. Med.* 51:S124–31
11. Chetty R, Stepner M, Abraham S, Lin S, Scuderi B, et al. 2016. The association between income and life expectancy in the United States, 2001–2014. *JAMA* 315:1750–66
12. Collins FS. 2010. Research agenda. Opportunities for research and NIH. *Science* 327:36–37
13. Community Prev. Serv. Task Force. 2016. *Using evidence to improve health outcomes*. Annu. Rep. Congr., Community Guide, Atlanta. <https://www.thecommunityguide.org/sites/default/files/assets/2016-congress-report-full.pdf>
14. Compennolle S, De Cocker K, Lakerveld J, Mackenbach JD, Nijpels G, et al. 2014. A RE-AIM evaluation of evidence-based multi-level interventions to improve obesity-related behaviours in adults: a systematic review (the SPOTLIGHT project). *Int. J. Behav. Nutr. Phys. Act.* 11:147
15. Concannon TW, Fuster M, Saunders T, Patel K, Wong JB, et al. 2014. A systematic review of stakeholder engagement in comparative effectiveness and patient-centered outcomes research. *J. Gen. Intern. Med.* 29:1692–701

16. Concannon TW, Meissner P, Grunbaum JA, McElwee N, Guise JM, et al. 2012. A new taxonomy for stakeholder engagement in patient-centered outcomes research. *J. Gen. Intern. Med.* 27:985–91
17. Cook TD, Campbell DT. 1979. *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Chicago: Rand McNally
18. Cronbach LJ, Shapiro K. 1982. *Designing Evaluations of Educational and Social Programs*. San Francisco: Jossey-Bass
19. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. 2009. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. *Implement. Sci.* 4:50
20. Daniel H, Bornstein SS, Kane GC, Health Public Policy Comm. Am. Coll. Phys. 2018. Addressing social determinants to improve patient care and promote health equity: an American College of Physicians position paper. *Ann. Intern. Med.* 168:577–78
21. De Maeseneer JM, van Driel ML, Green LA, van Weel C. 2003. The need for research in primary care. *Lancet* 362:1314–19
22. Eisner EW. 2002. *The Arts and the Creation of Mind*. New Haven, CT: Yale Univ. Press
23. Feldstein AC, Glasgow RE. 2008. A practical, robust implementation and sustainability model (PRISM) for integrating research findings into practice. *Jt. Comm. J. Qual. Patient Saf.* 34:228–43
24. Fisher EB. 2008. The importance of context in understanding behavior and promoting health. *Ann. Behav. Med.* 35:3–18
25. Flay BR. 1986. Efficacy and effectiveness trials (and other phases of research) in the development of health promotion programs. *Prev. Med.* 15:451–74
26. Gaglio B, Phillips SM, Heurtin-Roberts S, Sanchez MA, Glasgow RE. 2014. How pragmatic is it? Lessons learned using PRECIS and RE-AIM for determining pragmatic characteristics of research. *Implement. Sci.* 9:96
27. Gaglio B, Shoup JA, Glasgow RE. 2013. The RE-AIM framework: a systematic review of use over time. *Am. J. Public Health* 103:e38–46
28. Galea S, Tracy M, Hoggatt KJ, Dimaggio C, Karpati A. 2011. Estimated deaths attributable to social factors in the United States. *Am. J. Public Health* 101:1456–65
29. Gartlehner G, Thieda P, Hansen RA, Morgan LC, Shumate JA, Nissman DB. 2009. Inadequate reporting of trials compromises the applicability of systematic reviews. *Int. J. Technol. Assess. Health Care* 25:323–30
30. Glasgow RE. 2013. What does it mean to be pragmatic? Pragmatic methods, measures, and models to facilitate research translation. *Health Educ. Behav.* 40:257–65
31. Glasgow RE, Estabrooks PE. 2018. Pragmatic applications of RE-AIM for health care initiatives in community and clinical settings. *Prev. Chronic Dis.* 15:170271
32. Glasgow RE, Gaglio B, Bennett G, Jerome GJ, Yeh H-C, et al. 2012. Applying the PRECIS criteria to describe three effectiveness trials of weight loss in obese patients with comorbid conditions. *Health Serv. Res.* 47:1051–67
33. Glasgow RE, Huebschmann AG, Brownson RC. 2018. Expanding the CONSORT figure: increasing transparency in reporting on external validity. *Am. J. Prev. Med.* 55:422–30
34. Glasgow RE, Vogt TM, Boles SM. 1999. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. *Am. J. Public Health* 89:1322–27
35. Glooko. 2018. *Glooko API for EHRs and clinical platforms*. Fact Sheet, Glooko, Mountain View, CA. http://get.glooko.com/rs/352-OEL-682/images/PL_API%20for%20Platforms_PRM%200239_16-06_rA.pdf
36. Graham ID, Logan J, Harrison MB, Straus SE, Tetroe J, et al. 2006. Lost in knowledge translation: time for a map? *J. Contin. Educ. Health Prof.* 26:13–24
37. Green LW. 2006. Public health asks of systems science: to advance our evidence-based practice, can you help us get more practice-based evidence? *Am. J. Public Health* 96:406–9
38. Green LW, Glasgow RE. 2006. Evaluating the relevance, generalization, and applicability of research: issues in external validation and translation methodology. *Eval. Health Prof.* 29:126–53
39. Green LW, Nasser MN. 2018. **Furthering dissemination and implementation research: the need for more attention to external validity. See Ref. 6, pp. 301–16**

39. Gives an expanded background on the origins of researchers' bias toward prioritizing internal validity over external validity and a thorough grounding in the domains of external validity that need to be assessed in order to ensure future translation to practice and policy.

40. Green LW, Ottoson JM, García C, Hiatt RA. 2009. Diffusion theory and knowledge dissemination, utilization, and integration in public health. *Annu. Rev. Public Health* 30:151–74
41. Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O. 2004. Diffusion of innovations in service organizations: systematic review and recommendations. *Milbank Q.* 82:581–629
42. Harden SM, Gaglio B, Shoup JA, Kinney KA, Johnson SB, et al. 2015. Fidelity to and comparative results across behavioral interventions evaluated through the RE-AIM framework: a systematic review. *Syst. Rev.* 4:155
43. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, et al. 2014. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. *BMJ* 348:g1687
44. Holtrop JS, Rabin BA, Glasgow RE. 2018. Dissemination and implementation science in primary care research and practice: contributions and opportunities. *J. Am. Board Fam. Med.* 31:466–78
45. Ioannidis JPA. 2005. Why most published research findings are false. *PLOS Med.* 2:e124
46. Ioannidis JPA. 2014. How to make more published research true. *PLOS Med.* 11:e1001747
47. Kennedy-Martin T, Curtis S, Faries D, Robinson S, Johnston J. 2015. A literature review on the representativeness of randomized controlled trial samples and implications for the external validity of trial results. *Trials* 16:495
48. Kessler RS, Purcell EP, Glasgow RE, Klesges LM, Benkeser RM, Peek CJ. 2013. What does it mean to “employ” the RE-AIM model? *Eval. Health Prof.* 36:44–66
49. Kilbourne AM, Neumann MS, Pincus HA, Bauer MS, Stall R. 2007. Implementing evidence-based interventions in health care: application of the replicating effective programs framework. *Implement. Sci.* 2:42
50. Klesges LM, Dziewaltowski DA, Glasgow RE. 2008. Review of external validity reporting in childhood obesity prevention research. *Am. J. Prev. Med.* 34:216–23
51. Koppelaar T, Linmans J, Knottnerus JA, Spigt M. 2011. Pragmatic vs. explanatory: an adaptation of the PRECIS tool helps to judge the applicability of systematic reviews for daily practice. *J. Clin. Epidemiol.* 64:1095–101
52. Krist AH, Phillips SM, Sabo RT, Balasubramanian BA, Heurtin-Roberts S, et al. 2014. Adoption, reach, implementation, and maintenance of a behavioral and mental health assessment in primary care. *Ann. Fam. Med.* 12:525–33
53. Laws RA, St. George AB, Rychetnik L, Bauman AE. 2012. Diabetes prevention research: a systematic review of external validity in lifestyle interventions. *Am. J. Prev. Med.* 43:205–14
54. Levine DM, Savarimuthu S, Squires A, Nicholson J, Jay M. 2015. Technology-assisted weight loss interventions in primary care: a systematic review. *J. Gen. Intern. Med.* 30:107–17
55. Leviton LC. 2017. Generalizing about public health interventions: a mixed-methods approach to external validity. *Annu. Rev. Public Health* 38:371–91
56. Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe KE, Zwarenstein M. 2015. The PRECIS-2 tool: designing trials that are fit for purpose. *BMJ* 350:h2147
57. Luoma KA, Leavitt IM, Marrs JC, Nederveld AL, Regensteiner JG, et al. 2017. How can clinical practices pragmatically increase physical activity for patients with type 2 diabetes? A systematic review. *Transl. Behav. Med.* 7(4):751–72
58. Manafó E, Petermann L, Vandall-Walker V, Mason-Lai P. 2018. Patient and public engagement in priority setting: a systematic rapid review of the literature. *PLOS ONE* 13:e0193579
59. McLeroy KR, Bibeau D, Steckler A, Glanz K. 1988. An ecological perspective on health promotion programs. *Health Educ. Q.* 15:351–77
60. Murray CJ, Kulkarni SC, Michaud C, Tomijima N, Bulzacchelli MT, et al. 2006. Eight Americas: investigating mortality disparities across races, counties, and race-counties in the United States. *PLOS Med.* 3:e260
61. Offringa R, Sheng T, Parks L, Clements M, Kerr D, Greenfield MS. 2018. Digital diabetes management application improves glycemic outcomes in people with type 1 and type 2 diabetes. *J. Diabetes Sci. Technol.* 12(3):701–8

62. Owen N, Goode A, Sugiyama T, Koohsari MJ, Healy G, et al. 2018. Designing for dissemination in chronic disease prevention and management. See Ref. 6, pp. 107–20
63. Pedrana L, Pamponet M, Walker R, Costa F, Rasella D. 2016. Scoping review: national monitoring frameworks for social determinants of health and health equity. *Glob. Health Action* 9:28831
64. Phippasone-Brady P, Kwan BM, Holtrop JS, Loskutova N, Rabin BA. 2018. *Processes, outcomes, and products to make designing for dissemination POP!* Presented at the Society of Behavioral Medicine Annual Meeting, New Orleans, LA, April 12
65. Pinnock H, Barwick M, Carpenter CR, Eldridge S, Grandes G, et al. 2017. Standards for Reporting Implementation Studies (StaRI) statement. *BMJ* 356:i6795
66. Rabin BA, Brownson RC, Haire-Joshu D, Kreuter MW, Weaver NL. 2008. A glossary for dissemination and implementation research in health. *J. Public Health Manag. Pract.* 14:117–23
67. Rabin BA, Lewis CC, Norton WE, Neta G, Chambers D, et al. 2016. Measurement resources for dissemination and implementation research in health. *Implement. Sci.* 11:42
68. Rogers EM. 1962. *Diffusion of Innovations*. New York: Free Press of Glencoe
69. Rothwell PM. 2005. External validity of randomised controlled trials: “To whom do the results of this trial apply?” *Lancet* 365:82–93
70. Rowling L, Jeffreys V. 2006. Capturing complexity: integrating health and education research to inform health-promoting schools policy and practice. *Health Educ. Res.* 21:705–18
71. SAMHSA (Subst. Abuse Ment. Health Serv. Adm.). 2018. *Statement of Elinore F. McCance-Katz, MD, PhD, Assistant Secretary for Mental Health and Substance Use regarding the National Registry of Evidence-based Programs and Practices and SAMHSA’s new approach to implementation of evidence-based practices (EBPs)*. Press Release, Jan. 11. <https://www.samhsa.gov/newsroom/press-announcements/201801110330>
72. Sanchez MA, Rabin BA, Gaglio B, Henton M, Elzarrad MK, et al. 2013. A systematic review of eHealth cancer prevention and control interventions: new technology, same methods and designs? *Transl. Behav. Med.* 3:392–401
73. Saunders C, Byrne CD, Guthrie B, Lindsay RS, McKnight JA, et al. 2013. External validity of randomized controlled trials of glycaemic control and vascular disease: How representative are participants? *Diabet. Med.* 30:300–8
74. Shippee ND, Allen SV, Leppin AL, May CR, Montori VM. 2015. Attaining minimally disruptive medicine: context, challenges and a roadmap for implementation. *J. R. Coll. Phys. Edinb.* 45:118–22
75. Steckler A, McLeroy KR. 2008. The importance of external validity. *Am. J. Public Health* 98:9–10
76. Steensma JT, Kreuter MW, Casey CM, Bernhardt JM. 2018. Enhancing dissemination through marketing and distribution systems: a vision for public health. See Ref. 6, pp. 191–200
77. Stirman SW, Miller CJ, Toder K, Calloway A. 2013. Development of a framework and coding system for modifications and adaptations of evidence-based interventions. *Implement. Sci.* 8:65
78. Thomson HJ, Thomas S. 2012. External validity in healthy public policy: application of the RE-AIM tool to the field of housing improvement. *BMC Public Health* 12:633
79. Thorpe KE, Zwarenstein M, Oxman AD, Treweek S, Furlberg CD, et al. 2009. A pragmatic-explanatory continuum indicator summary (PRECIS): a tool to help trial designers. *CMAJ* 180:E47–57
80. Tomoia-Cotisel A, Scammon DL, Waitzman NJ, Cronholm PF, Halladay JR, et al. 2013. Context matters: the experience of 14 research teams in systematically reporting contextual factors important for practice change. *Ann. Fam. Med.* 11(Suppl. 1):S115–23

RELATED RESOURCES

1. CFIR (Consol. Framew. Implement. Res.). <https://cfirguide.org>

This website includes an overview of the CFIR framework, its domains and constructs, and libraries of citations using CFIR. The site also includes tools and templates for data collection and analysis, including a tool to assist researchers to build interview guides based on CFIR.

2. Lavis JN, Oxman AD, Souza NM, Lewin S, Gruen RL, Fretheim A. 2009. SUPPORT tools for evidence-informed health policymaking (STP) 9: assessing the applicability of the findings of a systematic review. *Health Res. Policy Syst.* 7(Suppl. 1):S9

This article guides health policy decision makers to assess how the findings of a systematic review apply to one's specific setting.

3. PRECIS-2 (Pragmatic-Explanatory Continuum Indicator Summary 2). <https://www.precis-2.org>

This website provides an interactive guide for researchers and policy makers on how to use PRECIS-2. It is specifically set up to allow researchers to proactively consider and rate the level of emphasis their proposed trial places on 9 domains of external validity factors, such as the eligibility criteria, the settings involved, and the level of organizational infrastructure required.

4. RE-AIM (Reach, Effectiveness, Adoption, Implementation, Maintenance). <http://www.re-aim.org>

This website provides researchers with resources and recent presentations and publications related to the RE-AIM framework and external validity. Information on how to apply the framework is included, as well as information on its components: reach, effectiveness, adoption, implementation, and maintenance. Examples of optimal use are also included.

5. Tabak RG, Chambers DA, Hook R, Brownson RC. 2018. The conceptual basis for dissemination and implementation research: lessons from existing models and frameworks. In *Dissemination and Implementation Research in Health: Translating Science to Practice*, ed. RC Brownson, GA Colditz, EK Proctor, pp. 73–88. New York: Oxford Univ. Press. 2nd ed.

This chapter provides an overview of D&I frameworks and provides guidance on applying three frameworks to case scenarios.