Balancing rigor against the inherent limitations of investigating hard-to-reach populations

Richard A. Crosby¹*, Laura F. Salazar²,³, Ralph J. DiClemente²–⁵ and Delia L. Lang²,³

Abstract

Maintaining rigor in research is critical; however, this need must be balanced by the necessity of conducting studies in populations where inherent barriers exist relative to key issues such as recruitment, attrition, sampling, sample size, assessment techniques, psychometric rigor, the identification of mediators and moderators and the practical relevance of the research question itself. Ultimately, the value of a study in health promotion should be judged on the practicality of the research question within the context of the target population. Striking the perfect balance between rigor and practicality to the field is a question that health promotion researchers and professionals need to determine through ongoing dialogue and debate.

Introduction

The field of health promotion is rapidly becoming a cornerstone of public health practice. One inescapable observation is that the elimination of health disparities is indeed a large part of the overall mission and scope of health promotion efforts [1]. In thinking about populations most affected by health disparities, it is quickly apparent that issues such as literacy and lack of stability could be quite problematic for a researcher trying to construct a study designed to understand and change health behavior. In contrast, for example, to studies of middle class, well-educated people residing in stable home environments, a study of homeless women who trade sex for drugs clearly poses an entirely different set of challenges and magnifies the obstacles to achieving rigor in research. This stark difference demands that the scholars involved in designing and conducting research, as well as researchers and practitioners involved in utilizing the published research findings, be aware of the need to judge research based on a balance between methodological rigor and the inherent limitations imposed by working with any given hard-to-reach population.

As a translational multi-disciplinary field, emphasizing a ‘research-to-practice’ paradigm (and a practice-to-research paradigm), health promotion should be reliant on rigorous research methodology to identify ‘best practice’ approaches to addressing health problems. The external validity of health promotion research is clearly a paramount concern. Although all health promotion research is inherently valuable, studies with high external validity are more readily translatable, can be more easily brought to scale and thus can have a more widespread impact. The external validity of health promotion research is clearly a paramount concern given that findings from research...
only have value when they are successfully translated into widespread practice. Tightly controlled randomized trials, for example, may produce significant findings that have little practical value simply because the conditions created in the trial cannot feasibly be replicated in typical settings. This issue of external validity is important and it has been the subject of previous commentaries [2–5]. The purpose of this commentary is to depart from past discourse on external validity to identify several issues pertaining to methodological rigor in the context of practice-based research. Although some researchers may feel compelled to find solutions to these issues, we suggest that many are intractable and thus their identification alone may be extremely valuable. Populations most in need of empirical study typically pose these types of intractable challenges to researchers. For example, the luxury of conducting a prospective study among highly literate people leading stable lives is not one that can be typically expected by a researcher dedicated to HIV prevention, substance abuse prevention or the reduction of gender-based violence. Instead, health promotion studies must strike a balance between the need for rigor and the realities of working with populations that may be less than amenable to engaging in research.

Achieving methodological rigor is, without question, challenging. Although we acknowledge that no study can be ‘perfect’ in terms of methodological rigor, nonetheless, numerous prescriptions for enhancing methodological rigor exist and a number of key issues need to be carefully weighed in designing health promotion studies [6]. In this commentary, we address seven issues typically associated with methodological rigor (attrition, probability sampling, recruitment, effect size, assessment, psychometric validity and mediation) and we note the need for practical relevance of the research questions being studied. For each issue, we urge the reader to consider a balance between rigor and practicality.

### Issue 1: attrition

The issue of high attrition is an ongoing problem in studies of hard-to-reach populations. For example, in studies that evaluate multi-session intervention programs, attrition issues may be resolved by interspersing periodic assessments that provide a hearty financial incentive for return. While this strategy may increase return rates, its utility in practice is highly questionable as people are not ‘paid’ to return. Although minimizing attrition is important, in hard-to-reach populations (e.g. injection drug users, homeless, drug-dependent populations, commercial sex workers and persons released from prison), application of even the most comprehensive retention strategies may not be sufficient. Published standards for determining ‘acceptable attrition’ do not exist. Thus, a key question for health promotion is how much attrition is tolerable when studying high-risk, less-stable populations?

Applying the same scientific standard for what constitutes acceptable levels of attrition to all populations, whether they are readily accessible or not, may discourage the research community from fully pursuing the mandate of health promotion to address health disparities among hard-to-reach populations. We acknowledge that high attrition may result in imprecise estimates and, if attrition is differential between study conditions, it may also threaten the internal validity of the study. If the appropriate measurement of key variables has been made at baseline, differential attrition can be assessed and, if detected, statistical analyses can be adjusted accordingly. Unfortunately, it may be quite problematic to determine what the key variables are thereby suggesting that this approach may not be an optimal solution. Indeed, we assert that there may not be an optimal solution. What we must confront is the need to balance our interest in hard-to-reach populations against the use of scientific standards that may be difficult to meet given the challenges posed by these populations.

### Issue 2: probability sampling

The value of probability sampling is the added ability to make inferences from the sample to the
population. The capacity for making inferences is based on the degree of representativeness achieved by the probability sampling technique. Of note, the ability to make inferences is different than generalizability in that generalizability refers to how well the results might apply to populations other than the sample population. Research methods textbooks for the social and behavioral sciences promote probability sampling over non-probability sampling because of the advantage of the former to infer results to the larger population. The downside is that probability sampling relies on enumerating a sampling frame that is rarely available for many of the populations prioritized for health promotion intervention. Under these circumstances, we suggest that other sampling methods such as convenience sampling, purposive sampling, time–venue-based sampling or respondent-driven sampling may be useful and should be viewed more as a cautionary note rather than a flaw in the study design.

**Issue 3: recruitment, sample size and effect size**

Naturally, recruitment challenges differ greatly across various populations, with some populations being far more difficult to access and recruit than others. For example, studies of drug addicts may not be able to amass the large sample sizes that characterize studies of the general population. Thus, it is important to consider whether there is a tendency within our discipline to overvalue findings derived from large samples and undervalue findings stemming from small samples. Such bias could inadvertently steer research interest and focus away from studies involving under-served and hard-to-reach populations. Although smaller samples may be limited in their statistical power to effectively test hypotheses and the precision of effect estimates may be less stable, a key consideration is the value of the findings in addressing gaps in the empirical literature as these gaps may be valuable for informing public health policy and practice.

Along this vein, in any study, whether the sample is large or small, effect sizes and the corresponding confidence limits should be reported. In studies of hard-to-reach populations, medium or large effects [7] are substantive and may have tremendous importance regardless of the limited sample size and under-powered statistical testing. Indeed, the real fly in the ointment involves studies that have very small effects but because of the inflated power afforded by very large sample sizes the results are deemed ‘significant’. Using standard hypothesis testing, with an alpha of 0.05 as the guard against type 1 error, these large and expensive studies may easily lead readers to overlook the fact that the effect size may have been small.

Although finding large effect sizes is often the ‘holy grail’ of research, this too should be viewed in the context of practice. The hard reality of current funding for public health is that cost-constrained environments are indeed the norm. Thus, studies that report effect sizes that were generated by labor- and resource-intensive intervention activities may be less than ideal from the standpoint of external validity. An important consideration is also to demonstrate that medium to large effect sizes can be replicated with fidelity and, if adapted, can be used effectively in the context of practice settings [7].

**Issue 4: assessment**

Without question, the collection of valid behavioral data and the corresponding psychosocial variables are essential to health promotion research. The challenges posed by this task may be magnified for populations with low literacy thereby necessitating the use of face-to-face interviews or audio-assisted methods of data collection. Thus, unlike research with highly literate populations, much of the health promotion research literature is dependent upon the use of data collection methodology that does not require reading. Here, we suggest that literacy levels of the sample be thoroughly assessed and described by the researchers and that carefully selected
strategies should be devised to accommodate literacy problems.

**Issue 5: reliability and validity**

First, it is axiomatic that the validity of self-reported behaviors most often cannot be tested empirically. This reality may lead to a bias against or devaluation of research findings derived solely from self-report data. However, given the valuable role of behavioral intervention in the prevention of disease, research efforts will clearly continue to rely on observational and experimental studies that seek to understand and change health behavior thereby necessitating the use of self-reported measures.

A second issue related to validity applies to the use of scales designed to assess psychosocial constructs pertaining to health behaviors. While scale validity is important and we recommend using well-established scales characterized by high reliability and validity, often scales are not available to assess particular constructs relevant to our study populations. In such cases, a scale may have to be created to capture the relevant information. Thus, studies with newly developed scales, designed to capture psychosocial constructs of particular importance to a population, can provide useful information and further the research enterprise. Using scales that have been through the proverbial ‘psychometric ringer’ to establish all forms of validity and reliability may partly define rigor, but we suggest that this feature may be a relatively impractical luxury in many emerging areas of research among diverse populations.

**Issue 6: mediators and moderators**

In the context of efficacy trials, one key question addressed is ‘does the intervention operate through the theoretically derived psychosocial mediators?’ Thus, in efficacy trials, a rigorous study would provide a mediation analysis to empirically establish that the intervention changed targeted hypothesized mediators of behaviors and that favorable changes in these constructs led to the adoption of health-protective behaviors. In the absence of this empirical evidence, the success or failure of an intervention program cannot be traced back to the actual protocol or curriculum that guided implementation. Clearly, this feedback may also become informative relative to the use of theory to guide behavioral intervention efforts. Indeed, using evaluation findings to gauge the value of applied theories is an important obligation of both researchers and practitioners in health promotion. Nonetheless, from a practice-based perspective, it may be less important to determine why a program worked than to identify optimal conditions of effectiveness. Thus, increased attention to moderating factors in addition to mediating factors is warranted.

**Issue 7: practical relevance**

Finally, and perhaps of most importance, we suggest that the most valuable studies are those which are firmly grounded in a research question with direct relevance to practice (i.e. practicality). This relevance should be made explicit. Research that fails to produce conclusions about practice falls short of being valuable. Indeed, the ‘value’ of any given study should be judged, in large part, based upon the potential utility of the findings to reasonably inform health promotion practice. In essence, the utility of the research question to practice is paramount [3]. This point does not name ‘empirical rigor’ as the paramount value as doing so would imply that studies meeting the highest standards are inherently valuable regardless of the study population and (the converse) that studies failing to meet the highest standards are inherently flawed regardless of their potential value to practice.

Striking the perfect balance between rigor and practicality to the field is a question that health promotion researchers and professionals need to determine through ongoing dialogue and debate. Without question, populations such as African American men having sex with men, methamphetamine-using teens, homeless women who exchange sex for money and exploited children warrant ongoing research attention as do populations of marginalized and under-served people in developing countries.
Although these studies should strive to meet all established expectations of rigor (or seek to compensate for lack of rigor through analytic and design variations), we suggest that practicality may sometimes preclude rigor. The ‘balance’ then is defined by the degree of departure from these expectations that represent an acceptable trade-off in exchange for valuable data from under-served and difficult to study populations. Although these ‘trade-offs’ have previously been described with respect to randomized trials [8–10], we suggest that they also apply to the study of hard-to-reach populations. Ultimately, determining the acceptable degree of trade-offs is a responsibility of the professionals reading published articles as well as being the responsibility of investigators conducting the research, sponsors of the research, peer reviewers of manuscript, etc. It is a shared responsibility among many. We suggest that these professionals all should carefully balance the burden of evidence against the inherent limitations imposed by the choice of the study population before making conclusions about the value of the empirical findings.

Conflict of interest statement

None declared.