Published in final edited form as:

Prev Sci. 2015 October ; 16(7): 1026–1032. doi:10.1007/s11121-015-0577-4.

Why Small is Too Small a Term: Prevention Science for Health Disparities, Culturally Distinct Groups, and Community-level Intervention

David Henry,

University of Illinois-Chicago

Carlotta Ching Ting Fok, and

University of Alaska Fairbanks

James Allen

University of Minnesota Medical School, Duluth Campus

Abstract

Implications of the *Advancing Small Sample Prevention Science* Special Section are discussed. Efficiency and precision are inadequately considered in many current prevention science methodological approaches. As a result, design and analytic practices pose difficulties for the study of contextual factors in prevention, which often involve small samples. Four primary conclusions can be drawn from the Special Section. First, contemporary statistical and measurement approaches provide a number of underutilized opportunities to maximize power. These strategies maximize efficiencies by optimizing design and resource allocation parameters, allowing for the detection of effects with small samples. Second, several alternative research designs provide both rigor and further optimize efficiencies through more complete use of available information, allowing study of important questions in prevention science for which only small samples may be accessible. Third, mixed methods hold promise for enhancing the utility of qualitative data in studies with small samples. Finally, Bayesian analytic approaches, through their use of prior information, allow for even greater efficiencies in work with small samples, and through their introduction in the routines of mainstream software packages, hold particular promise as an emergent methodology in prevention research.

Keywords

Small samples; research methods; statistical methods; health disparities; ethnic minority research

“Great things are done by a series of small things brought together”

Correspondence to this article should be addressed to James Allen, Department of Biobehavioral Health and Population Sciences, University of Minnesota Medical School, Duluth Campus, 231 SMed, 1035 University Drive, Duluth, MN 55812-3031, jallen@d.umn.edu.

David Henry, Institute for Health Research and Policy, University of Illinois at Chicago; Carlotta Fok, Center for Alaska Native Health Research, University of Alaska Fairbanks; James Allen, Department of Biobehavioral Health and Population Sciences, University of Minnesota Medical School, Duluth Campus.

The authors declare that they have no conflict of interest.

(Vincent Van Gogh, 1882)

Although the articles of the *Advancing Small Sample Prevention Science* Special Section are particularly relevant to the analysis of small samples, their implications extend beyond small data sets. Together they provide important reflections on current design and analytic practices in prevention and intervention science, and question some of its assumptions. A primary observation through the lens of the small samples articles of this Special Section is that current practices in prevention science rarely weight considerations of efficiency in design and analytic choices. This has resulted in a current set of assumptions aligned to the necessity of amassing ever larger, 'big' data sets in order to (a) detect measureable effects of preventative interventions, (b) establish causality in intervention impact, and (c) facilitate rigorous prevention theory elaboration. However, with these data set size requirements comes a cost; in this case, a diminished capacity to study local contextual factors often crucial to effective prevention programs, and to research preventive solutions for groups that experience health disparities, that are also often culturally distinct, small population groups. In other words, the capacity to effectively study those factors that are of potentially the greatest societal importance is diminished through many of the current conventions of prevention research design and analysis.

The introduction to this Special Section (Fok, Henry, & Allen, 2015) noted an irony in how the original premise and justification for introduction of statistics in research was its ability to draw inferences using a relatively small sample drawn from the population of interest. Current advances in statistical methods allow for the kinds of finely grained analysis and detailed description of subtle intervention effects undreamed of by its earliest developers. However, as the papers in this Special Section highlight, the last several decades have witnessed an ever-increasing escalation in minimum requirements for sample size. A recent exemplar of prevention research, the Diabetes Prevention Program intervention, screened over 158,000 individuals to achieve a final sample size of 3234 (The Diabetes Prevention Program Research Group, 2000). The high impact arm of the Diabetes Prevention Program included an elaborate recruitment process, dietary and exercise components culturally tailored to setting, curriculum delivered by case managers carefully trained to the context, and local staff who monitored adherence in a culturally tailored way. The question of generalizability of such findings calls forth scientific humility; moreover, the size of the screened sample required for the final sample precludes systematic study of local variation in contextual factors and the adaptations to them crucial to the intervention effectiveness (Trickett, 2011). Consideration of these issues highlights ways the analysis of small samples is relevant to all prevention science.

The articles of the Special Section describe ways in which much of the current need for large sample sizes is driven through inefficiencies in the particulars of application of a number of prevention research methodologies, as well as by built-in factors of research designs that are currently promoted as best practices, but are often inefficient in their statistical power. The result of these inefficiencies has erected more than just barriers for small samples researchers, who find current state of the art analytic techniques out of reach given their large sample size requirements. It has also imposed limitations in what is possible for health disparities research, creating restrictions that can prevent exploration of

solutions for those small population groups most adversely affected in health outcomes. This has led to a lack of empirical data to guide the development of intervention for these populations, and limited research to establish what is effective in interventions to address their health disparities. Time is ripe for re-examination of ways current approaches tie prevention science to inefficient methodologies that impose obstacles to scientific progress, leaving unaddressed critical research questions in health disparities and in effective intervention for small population, culturally distinct groups.

Issues called forth in the first subsection of the Special Section emphasize efforts to more accurately model context, through enhancement of measurement precision to minimize measurement error, or through use of alternative research designs that more adequately consider elements of context. Current experimental design approaches in prevention science primarily attempt to control extraneous variables through randomization in order to identify causal effects from intervention; this can also have the effect of decontextualizing local variation in setting that are critical considerations in making community-level intervention effective (Trickett et al., 2011). The Affordable Care Act emphasis on preventative medicine with a population health orientation drives increased need for methods that facilitate study of group interventions whose target is the population or community level. In these interventions the unit of randomization is group; when degrees of freedom are determined by number of communities sampled and not the number of individual participants, all community level intervention becomes small sample research.

The second subsection of the Special Section provides recommendations regarding small sample analytic procedures. The opening article (Hoyle, 2015) focuses on two widely used procedures in current prevention research, multilevel modeling (MLM) and structural equations modeling (SEM), providing guidance on optimizing their application with small samples. The next three articles offer three innovations in analytic methods for potential adoption by small sample researchers: (a) a statistical modeling approach for multilevel ordinal data (Hedeker, 2015), (b) three mixed-methods approaches that embed quantitative analyses in support of more complex interpretation of coded qualitative data (Henry, Dymnicki, Mohatt, Allen, & Kelly, 2015), and (c) a Bayesian approach as an alternative to current frequentist statistical approaches (Kadane, 2015).

Four primary conclusions can be drawn from these articles in the Special Section on the analysis of small samples. First, the articles highlight a number of optimization strategies applicable to contemporary statistical and measurement approaches that can facilitate small sample research. Second, the articles introduce an array of alternative research designs that offer both rigor and efficiency optimization through more complete use of the available information in a small sample study. Third, mixed methods provide new possibilities in qualitative research, allowing for more complex inference from samples often too small to allow for quantitative testing. Fourth, Bayesian approaches allow for even greater efficiencies in quantitative work with small samples through their use of prior information. The introduction of Bayesian analysis to routines in mainstream software packages holds particular promise as an emergent methodology for prevention research, and Bayesian thinking about probability has important benefits for prevention science as well. Implications of these four conclusions are next considered in greater detail. Together, these

four broad strategies open possibility for studies on a number of critical questions in prevention science that are currently excluded from research because only small sample sizes are accessible and current approaches can result in unpowered studies.

Refinements in Statistical Applications and Measurement Provide Strategies that Permit Analyses with Small Samples

Careful attention to features that optimize the implementation of current statistical procedures and enhance measurement rigor are immediate steps that prevention researchers can take to advance small sample research. In particular, advances in multilevel statistical approaches and item response theory alternatives to classical psychometric theory afford researchers fresh opportunities to maximize statistical power and detect change when sample size cannot be increased.

Hopkin, Hoyle, and Gottfredson (2015) provide a number of general recommendations for optimization of small sample research. As each observation in a small sample carries significant weight, missing data can both deleteriously impact power and introduce significant potential for bias. While a number of the complexities in theory and the mechanisms of missingness (Baraldi, 2013) extend beyond the scope of this Special Section, these authors open their discussion by emphasizing the critical importance of modern imputation procedures to small sample research. Next, they focus on context. They emphasize processes of variance reduction as a crucial optimization procedure with small samples, highlighting an array of options. Some of the most easily implemented include thoughtful model specification, based in theory or on previous research, to account for variance unrelated to the intervention effect through covariates or levels, and the use of within subjects designs to account for unmeasured personality and context factors. In conjunction with these variance reduction strategies, approaches that simultaneously maximize construct relevant variance constitute a complementary strategy to uncover effects. The authors discuss the potential role of purposive sampling in small sample research to increase construct relevant variance within the sample. They also make a crucial distinction, distinguishing between the differential power in analyses to clearly establish main effects, in contrast to tests of interaction effects, that make interactions more difficult to establish with small samples. This amplifies the importance of recommendations for use of fractional factorial designs (Dziak, Nahum-Shani, & Collins, 2012) in small sample research, in which design specification is based on a theory of change and previous research experience. Together these approaches suggest strategies by which all prevention study analyses can be more efficiently designed and conducted; all prevention research can benefit from these optimization strategies of study features and resource allocation to maximize the ability to detect effects. While with small samples, these considerations may attain a critical pragmatic importance; they also implicate an ethical responsibility for all researchers in their use of the services of human research participants wisely.

IRT procedures provide researchers with a set of more fine-tuned alternatives to the tools associated with classical measurement theory. IRT techniques afford enhanced precision to minimize measurement error. Graded response level approaches to scale construction additionally create measures tapping a continuum of item difficulty of the latent trait being

assessed, composed of non-overlapping items spanning different levels of the trait. This enhanced granularity in understanding item level functioning allowed in IRT affords more systematic exploration of an under recognized but highly salient consideration in prevention research, related to responsiveness or sensitivity to change, which describes a measure's accuracy in assessing change in response to intervention. Fok and Henry (2015) persuasively advocate for more routine assessment of responsiveness as a third measurement characteristic that prevention researchers should consider in addition to reliability and validity. The authors present several practical strategies that allow researchers to assess and optimize responsiveness of measurement scales. Scales are most effective for small sample prevention research when they tap a continuum of item difficulty and tap behaviors maximally sensitive to the processes of change instituted by intervention; reliability, validity, and responsiveness are together particularly salient in small sample prevention research. These first two articles highlight how research with small samples requires a level of attention to rigor in design and measurement in order to successfully measure effects and establish impact that is not always as essential to work with larger samples.

When conducting small population or community level research, what represents an adequate sample size for MLM and SEM statistical applications? Hoyle and Gottfredson (2015) address this question directly through careful examination of a growing literature of simulation studies, concluding modest sample sizes, even as low as 50, are justifiable in use with both these techniques when other conditions are satisfied. In many applications, MLM provides the capacity to model complex multi-level effects across two or even three levels in samples at these limited degrees of freedom, while SEQ exhibits greater instability at samples smaller than 200, requiring more careful specification and description of limitations in its reporting. Importantly, considerations of power are impacted as a function of accurate model specification and characteristics of the data, and this varies somewhat by statistical technique. For example, among several issues to attend to, in MLM particular concern arises if the outcome modeled is non-normally distributed, while in SEM, a major concern is in the degree to which the observed covariance matrix reflects the population covariance matrix; this again highlights the importance of measurement and sampling in small sample research. The authors provide examples of how optimization can be achieved in both MLM and SEM through selection of alternative estimation and testing procedures that perform best with small samples.

Given MLM's sensitivity to outcomes that exhibit non-normality, ordinal data may be more appropriate for small sample prevention research than the assumptions associated with normal, continuous data. To address this, Hedeker (2015) proposes use of mixed proportional odds models to provide an approach for clustered or longitudinal multilevel ordinal outcomes that may be particularly well suited for the analysis of small sample data. The approach addresses key weaknesses in current MLM implementations given the instability of findings with non-normal outcomes, the biased estimates produced by mixed normal and ordinal models to ordinal outcomes, and the severe loss of power when dichotomizing ordinal outcomes. By addressing this issue in current MLM application with non-normal outcomes, in which the procedure is especially sensitive to risk for providing biased estimates (Hoyle and Gottfredson (2015), the approach potentially represents a significant enhancement to existing MLM procedures. Here again, attending to an issue with

particular relevance to the analysis of small samples has broader implications for all of prevention research.

In summary, the articles of the Special Section conclude that research with modest samples sizes is clearly justifiable using contemporary analytic techniques now commonly used in prevention science. Refined scaling approaches can identify effects and minimize error, and optimization of design, model specification, and testing can maximize statistical power and minimize bias with small samples. These articles suggest some basic standards for statistical analysis and measurement with regard to small sample research, with increased attention to tests and model specification elements most appropriate to small samples. Small samples analysis requires thoughtful, theory driven application of current approaches to carefully model the distinctive attributes of context and individuals that impact upon the outcomes of interest, carefully refined scale development to provide precise measures of constructs through item pools that are maximally sensitive to the change processes an intervention initiates, and techniques that fit characteristics of the data often produced by small samples.

Use of Alternative Research Designs that Maintain Rigor While Maximizing Power

Small sample research creates several significant challenges to effective implementation of the randomized controlled trial (RCT) design. Wyman, Henry, Knoblauch, and Brown (2015) outline potential logistical and ethical issues that can arise in small sample applications of the RCT. The authors also note inefficiencies in how the RCT makes use of information that can result in low power and low external validity in small sample research, and discuss two alternative designs, the Dynamic Wail-listed Design (DWLD) and the Regression Point Displacement Design (RPDD). Both designs make us of optimization strategies to more efficiently use available information to maximize power with modest sample size. DWLD employs randomization in the ordering of the roll out of the intervention across multiple time periods among individual participants or the clusters formed by communities or groups (collectively termed “units” of analysis), while also treating units equitably regarding resources and intervention. One of the most significant optimization pathways DWLD uses to enhance efficiency, in contrast to RCT, involves maximization of power through the longer time periods that DWLD allows for comparing units that have received intervention, both through comparison to the unit’s own time series before intervention and to other units that have not yet received intervention. The approach also provides additional information to intervention designers unavailable in an RCT. For example, it can identify if the immediate response to intervention diminishes over time, or if growth in effects continues.

The RPDD is a still underutilized design that first appeared in the literature over 50 years ago (Cook, 2008). It offers a quasi-experimental design alternative allowing prevention research with samples as small as a single intervention unit if archival data exists from multiple units from the same population prior to and following the intervention. The more the selection process for comparison of units approximates random selection, the more valid a causal inference can be made. In accord with a second theme of this Special Section, attention to the modeling of context through the selection and measurement of covariates

(Steiner, Cook, Shadish, & Clark, 2010) has potential to greatly improve the accuracy of estimates in the RPDD.

Continuing this discussion of alternative designs, Fok, Henry, and Allen (2015) describe the stepped wedge design (SWD) and the interrupted time-series design (ITSD). The SWD is a related design to the DWLD, wherein the timing of the crossover to intervention for each unit is randomized, and power is maximized when each unit has its own randomization step. The SWD and DWLD originated across different disciplinary and research backgrounds, with some differences in implementation, such as different emphases on the use of blocking prior to randomization, and slight differences in conventions for statistical modeling. However both designs share in common the unidirectional roll-out of intervention across multiple random time periods.

ITSD is a quasi-experimental design that makes use of multiple time points to separate intervention from extraneous effects; multiple time points both increase internal validity and power. Much like the RPDD, the ITSD can be applied to evaluate prevention with a single unit, and its power and utility are increased through the number of time points available to segmented regression analysis, as well as careful selection and measurement of covariates.

The RCT is an inefficient design with regards to issues of statistical power. The Special Section describes four research designs that make more efficient use of available information, (i.e., DWLD, RPDD, SWD, and ITSD), that provide rigor and, in the cases of the DWLD and SWD, an element of randomization. It should be noted that the ITSD can also include randomization of the time of cross-over to intervention across units. Scientific skepticism incorporates a continual questioning of its methods, guiding researchers to question rigid adherence to methods that may exclude inferences regarding important research questions among those populations at greatest need, such as health disparities and small population, culturally distinct groups. An emerging body of evidence suggests conditions wherein methods departing from 'true' experiments can produce comparable results (Cook, Shadish, & Wong, 2008; Shadish, Clark, & Steiner, 2008). In addition, despite being held up as a gold standard for research, RCTs possess their own set of methodological deficiencies (Kraemer, 2000), and a particular set of challenges when applied to clusters such as communities and groups, including bias introduced by differences in sample sizes across clusters and issues associated with fixed numbers of clusters (Eldridge, Ashby, & Kerry, 2006; Hemming, Girdling, Stich, Marsh, & Lilford, 2011). RCTs also possess limits regarding the description and modeling of context. Rather than describe, RCTs attempt to control for variation across units. In group level interventions such as community intervention, this results in an attempt to control for the local distinctiveness central to the community intervention paradigm (Trickett et al., 2011). An intervention science for community intervention might instead model this distinctiveness and intervention adaptations to it as part of analysis. These alternative designs optimize understanding of context.

Current emphases prioritizing implementations of the RCT as uniformly applicable and suitable to all settings and all populations risks the unintended effect of privileging interventions focused on individuals or accessible, large samples, rather than on

communities or health disparities groups that often involve small samples, but that might have a larger societal impact (West, 2009). The alternative designs presented here provide an important advantage over RCTs by permitting prevention scientists to address significant research questions, instead of abandoning them because they cannot be investigated through an RCT. A primary reason for this advantage is that these alternative designs are not as dependent on having a sufficient number of units. An additional benefit of particular relevance to prevention research is that well designed studies using these alternative designs also can often permit more detailed generalizations of causal effects specific to settings, populations, outcomes, and intervention adaptations.

Qualitative and Mixed Methods Research

Hopkin, Hoyle, and Gottfredson (2015) note how important statistical inferences can still be extracted from samples too small to allow for detection of impact, and how data visualization methods can be useful for sample sizes too small for hypothesis testing or even parameter estimation. Further opportunities in small sample research are provided by the rich possibilities presented through qualitative methods. Qualitative research is typically conducted with small samples, and can produce a level of nuance and complexity in description unrealizable using quantitative methods with small samples.

One exciting recent development within qualitative methodology has been progress in mixed methods, which involve an intentional collection of both quantitative and qualitative data with the goal of some type of integration of the methods that maximizes the strengths of each (Creswell, Plano Clark, Petska, & Creswell, 2011). Qualitative research is inductive and discovery based, and include ethnography and case studies, life history interviews, and structured interviewing. Quantitative research is deductive, and tests theories or hypotheses about the relationship among variables through measurement, statistical analysis to establish causality, and generalization to populations or group comparisons. Quantitative approaches include randomized controlled trials, time-series and other quasi-experimental designs, observational studies, case-control studies, and descriptive surveys. Creswell and Plano Clark (2011) identify three general mixed methods approaches at increasing levels of integration.

Merging data can involve reporting statistical results with qualitative quotes or themes that support or refute these results, or transforming qualitative to quantitative data (e.g., occurrence of themes from qualitative coding). Connecting data uses analysis from one data set (e.g., a quantitative survey) to inform a subsequent project (e.g., a qualitative interview study). Embedding data implants data collection and analysis from one method that is of secondary importance within a primary research design using a different method. Henry, Dymnicki, Mohatt, Kelly, and Allen (2015) propose new directions in embedded analysis.

Most typically, qualitative analysis has been used secondarily, to support quantitative research as the primary aim of a study. In contrast, Henry et al. propose an approach in which the quantitative analysis is secondary, and used to support and advance qualitative study aims. Their approach uses cluster analysis to clarify the structure of researcher coding of qualitative data. The approach is notable in that recently, qualitative data analysis (QDA)

packages have begun providing basic quantitative tools, such as coding co-occurrence indices (e.g., the c-index; Garcia, 2005) that allow study of the interrelation of codes and their networks. Henry et al.'s demonstrations of clustering approaches provide additional tools allowing for more complex description of the configural networks of codes, and for the grouping of research participants within different configural structures. The authors test the functioning and relative strengths and weaknesses of three newer clustering approaches—latent class analysis, hierarchical cluster analysis, and K-mean analysis—to assist qualitative researchers seeking to organize their qualitative findings.

An intriguing application of these methods involves their potential to facilitate more systematic exploration of the meaning of relational configurations of code structures as part of the interpretative work in qualitative analysis; this can provide qualitative researchers with new tools to systematically explore their data beyond the interpretation of individual codes in isolation. The application of cluster methods in qualitative data interpretation can only be used effectively by a researcher with grounding in theory and context, as well as rich immersion in the narrative produced through qualitative work. The approach provides a realization of Jick's (1979) concept of triangulation from multiple methods, itself an extension of ideas originally proposed by Campbell & Fiske (1959). Future, implementations of QDA packages are likely to include these types of new analytic possibilities, much in the way current statistical packages are now beginning to introduce Bayesian analytic capabilities.

Bayesian Analysis

There has been a resurgence in interest in Bayesian perspectives in recent years as a viable alternative to the frequentist paradigm of statistics currently practiced in the establishment of a research evidence base, largely due to recent advances in statistical software packages that are both widely available yet capable of its computational complexity, and new, more accessibility introductions to its implementation (Kaplan, 2014). A strength of the Bayesian analytic approach is that it includes prior information in the analysis of data. This makes possible analysis with small sample sizes not conceivable using frequentist approaches. While a sample size of 20 may prove unfeasible using frequentist statistics, as a series of studies in small batches that integrate prior knowledge from earlier studies using a Bayesian approach, the information collected in these earlier studies can be used to shape the analysis, making such an analysis feasible. The Bayesian paradigm uses optimization strategies that build on the efficiency allowed through the use of existing information from the present knowledge base that is not accessed in current frequentist analytic approaches.

This use of prior information is more in keeping with how people use information generally, and the research enterprise is not all that different. For example, in the task of clinical decision-making, clinicians also routinely take prior information, combining it to produce posterior information. Further, in several areas of analysis such as mediation effects, researchers cannot assume a normal distribution, and Bayesian approaches do not impose these assumptions nor rely on large samples for their approximations (Yuan & MacKinnon, 2009).

Kadane (2015) provides an introduction to the Bayesian paradigm and reviews how it addresses a number of the issues that arise in small samples. He then provides a case example of application of a Bayesian approach to the analysis of a small sample prevention data set. In the example, he describes the specification of a likelihood function and a prior distribution, then the computation of the posterior distribution and its interpretation, providing an overview of how these fundamentals of Bayesian analysis apply to prevention research. This relatively simple illustration of a Bayesian approach to a small sample is more broadly suggestive of a potential paradigm shift in approach to the analyses of prevention research. This is driven by need for methods to better study group and population level interventions, health disparities, and other areas where societally important and impactful research requires researchers to work with small samples.

Conclusion

A primary conclusion from this Special Section incorporates broader considerations of the consequences of current practices in prevention research that rarely consider efficiency in design and analytic choices. This imposes limitations on what is possible in health disparities research, creating restrictions that can inhibit exploration of solutions for those small population, culturally diverse groups most adversely affected by health disparities. It additionally privileges individual level interventions and those interventions that target problems that can produce large samples most easily accessible to researchers, over community and population level or other health disparities interventions that typically involve research with small samples. Despite the potential for these types of small sample research studies with health disparities, culturally distinct groups, and community and population level interventions to have broad societal impacts, current methods in prevention science have often led to a lack of empirical data to guide development and establish what is effective in intervention to address health disparities. Optimization strategies for contemporary statistical and measurement approaches, alternative research designs that make more complete use of available information, mixed methods that provide new possibilities in qualitative research, and Bayesian approaches that optimize through use of prior information together provide rigorous approaches for the study of important prevention science research questions that can only be addressed through small sample research.

Acknowledgments

This Special Section of the *Prevention Science* was supported through a grant from the National Institutes of Drug Abuse [R13DA030834, C.C.T. Fok, PI], which funded the conference “Advancing Science with Culturally Distinct Communities: Improving Small Sample Methods for Establishing an Evidence Base in Health Disparities Research” held on August 17–18, 2011 at the University of Alaska Fairbanks. We thank University of Alaska President’s Professors John Himes, William Knowler, Alan Kristal, Mary Sexton, Nancy Schoenberg, Beti Thompson, and Edison Trickett for their support and input to the application and this conference. Preparation and background to this article was funded by the National Institute of Drug Abuse, the National Institute of Alcohol Abuse and Alcoholism, the National Institute for Minority Health and Health Disparities, and the National Institute for General Medical Services [T32 DA037183, R21AA016098, RO1AA11446; R21AA01 6098; R24MD001626; P20RR061430].

References

- Baraldi, AN. Missing data methods. In: Little, TD., editor. *The Oxford handbook of quantitative methods* (Vol. 2): Statistical analysis. New York, NY: Oxford University Press; 2013. p. 635–664.

- Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*. 1959; 56:81–105. [PubMed: 13634291]
- Creswell, JW.; Klassen, AC.; Plano Clark, VL.; Smith, KC. for the Office of Behavioral and Social Sciences Research. Best practices for mixed methods research in the health sciences. National Institutes of Health; 2011. Retrieved February 16, 2015 from http://obssr.od.nih.gov/mixed_methods_research
- Cook TD. “Waiting for life to arrive”: A history of the regression-discontinuity design in psychology, statistics and economics. *Journal of Econometrics*. 2008; 142:636–654.
- Cook TD, Shadish WR, Wong VC. Three conditions under which experiments and observational studies produce comparable causal estimates: New findings from within-study comparisons. *Journal of Policy Analysis and Management*. 2008; 27:724–750.
- Dziak JJ, Nahum-Shani I, Collins LM. Multilevel factorial experiments for developing behavioral interventions: Power, sample size, and resource considerations. *Psychological Methods*. 2012; 17:153–175. [PubMed: 22309956]
- Fok CCT, Henry D. Increasing the sensitivity of measures to change. *Prevention Science*. 201510.1007/s11121-015-0545-z
- Eldridge SM, Ashby D, Kerry S. Sample size for cluster randomized trials: Effect of coefficient of variation of cluster size and analysis method. *International Journal of Epidemiology*. 2006; 35:1292–1300.10.1093/ije/dyl129 [PubMed: 16943232]
- Garcia, E. Keywords co-occurrence and semantic connectivity. 2005. Retrieved from <http://www.miislita.com/semantics/c-index-1.html>
- Hedeker D. Methods for multilevel ordinal data in prevention research. *Prevention Science*. 201510.1007/s11121-014-0495-x
- Hemming K, Girling AJ, Stich AJ, Marsh J, Liliford RJ. Sample size calculations for cluster randomised controlled trials with a fixed number of clusters. *BMC Medical Research Methodology*. 2011; 11:102. [PubMed: 21718530]
- Henry D, Dymnicki AB, Mohatt N, Kelly JG, Allen J. Clustering methods with qualitative data: A mixed methods approach for prevention research with small samples. *Prevention Science*. 2015
- Hopkin CR, Hoyle RH, Gottfredson NC. Maximizing the yield of small samples in prevention research: A review of general strategies and best practices. *Prevention Science*. 201510.1007/s11121-014-0542-7
- Hoyle RH, Gottfredson NC. Sample size considerations in prevention research applications of multilevel modeling and structural equation modeling. *Prevention Science*. 201510.1007/s11121-014-0489-8
- Kadane JB. Bayesian methods for prevention research. *Prevention Science*. 201510.1007/s11121-014-0531-x
- Kaplan, D. Bayesian statistics for the social sciences. New York, NY: Guilford; 2014.
- Kraemer HC. Pitfalls of multisite randomized clinical trials of efficacy and effectiveness. *Schizophrenia Bulletin*. 2000; 26:533–541. [PubMed: 10993394]
- Shadish WR, Clark MH, Steiner PM. Can nonrandomized experiments yield accurate answers? A randomized experiment comparing random to nonrandom assignment. *Journal of the American Statistical Association*. 2008; 103:1334–1343.
- Shadish WR, Cook TD. The renaissance of field experimentation in evaluating interventions. *Annual Review of Psychology*. 2009; 60:607–629.
- Steiner PM, Cook TD, Shadish WR, Clark MH. The importance of covariate selection in controlling for selection bias in observational studies. *Psychological Methods*. 2010; 15:250–267. [PubMed: 20822251]
- The Diabetes Prevention Program Research Group. The Diabetes Prevention Program: Baseline characteristics of the randomized cohort. *Diabetes Care*. 2000; 23:1619–1629. [PubMed: 11092283]
- Trickett EJ. Community-based participatory research as worldview or instrumental strategy: Is it lost in translation(al) research? *American Journal of Public Health*. 2011; 101:1353–1355. [PubMed: 21680920]

- Trickett EJ, Beehler S, Deutsch C, Green LW, Hawe P, McLeroy K, et al. Advancing the science of community-level interventions. *American Journal of Public Health*. 2011; 101:1410–1419. [PubMed: 21680923]
- Van Gough V. Letter to Theo van Gogh from The Hague. 1882 Oct 22.
- West SG. Alternatives to randomized experiments. *Current Directions in Psychological Science*. 2009; 18:299–304.
- Wyman PA, Henry D, Knoblauch S. Designs for testing group-based interventions with limited number of social units: The dynamic wait-listed and regression point displacement designs. *Prevention Science*. 2015 10.1007/s11121-014-0535-6
- Yuan Y, MacKinnon DP. Bayesian mediation analysis. *Psychological Methods*. 2009; 14:301–322. 10.1037/a0016972 [PubMed: 19968395]