



HHS Public Access

Author manuscript

Prev Sci. Author manuscript; available in PMC 2016 October 01.

Published in final edited form as:

Prev Sci. 2015 October ; 16(7): 943–949. doi:10.1007/s11121-015-0584-5.

Maybe Small is Too Small a Term: Introduction to Advancing Small Sample Prevention Science

Carlotta Ching Ting Fok,

Center for Alaska Native Health Research, Institute for Arctic Biology, University of Alaska Fairbanks

David Henry, and

Institute for Health Research and Policy, University of Illinois at Chicago

James Allen

Department of Biobehavioral Health and Population Sciences, University of Minnesota Medical School, Duluth Campus

Abstract

Prevention research addressing health disparities often involves work with small population groups experiencing such disparities. The goals of this Special Section are to (1) address the question of what constitutes a small sample, (2) identify some of the key research design and analytic issues that arise in prevention research with small samples, (3) develop applied, problem-oriented, and methodologically innovative solutions to these design and analytic issues, and (4) evaluate the potential role of these innovative solutions in describing phenomena, testing theory, and evaluating interventions in prevention research. Through these efforts, we hope to promote broader application of these methodological innovations. We also seek whenever possible, to explore their implications in more general problems that appear in research with small samples but concern all areas of prevention research. This Special Section includes two sections. The first section aims to provide input for researchers at the design phase, while the second focuses on analysis. Each article describes an innovative solution to one or more challenges posed by the analysis of small samples, with special emphasis on testing for intervention effects in prevention research. A concluding article summarizes some of their broader implications, along with conclusions regarding future directions in research with small samples in prevention science. Finally, a commentary provides the perspective of the federal agencies that sponsored the conference that gave rise to this Special Section.

Keywords

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

Correspondence to this article should be addressed to Carlotta Ching Ting Fok, Center for Alaska Native Health Research, Institute for Arctic Biology, University of Alaska Fairbanks 99775-7000. ctfok@alaska.edu.

The authors declare that they have no conflict of interest.

This Special Section of *Prevention Science* is the result of a conference that addressed the analysis of small samples as an important challenge impeding the progress of health disparities research. “Advancing Science with Culturally Distinct Communities: Improving Small Sample Methods for Establishing an Evidence Base in Health Disparities Research” brought together prominent methodologists to focus on this key issue emergent in contemporary health disparities research. The reasons why small samples tend to appear in research on health disparities are manifold. Some of the most common reasons are the research may involve a health disparity group that is a small population ethnic minority or culturally distinct group. In addition, the work may be conducted in logistically challenging circumstances where fiscal and other feasibility restrictions require small samples. The intervention may be a community-level intervention, where the unit of randomization is at the level of community, not the individual. However, in all such cases, the potential may be large for the research to produce important findings that can address a significant health equity issue. This Special Section aims to address issues associated with small sample sizes, meeting its challenges by proposing innovative research designs and statistical methodologies. Its goal is to provide practical guidance to applied, prevention researchers who work in such areas as health disparities, culturally distinct groups, and school, community, and population level intervention, where they are often engaged in small sample research.

The burden of health disparities in the United States falls disproportionately upon ethnic minority groups. The health disparities of these culturally and ethnically distinct populations include elevated disease rates (CDC, 2011), accompanied by a myriad of social determinants that include among many factors substandard sanitation and medical care (IOM, 2002; Murray et al., 2006), as well as media portrayals that dampen efficacy for improving health-related behaviors (Anderson, 2005). These minority groups may themselves be overlooked in research efforts because they often constitute small, geographically dispersed populations (Giger & Davidhizar, 2007). Achieving large sample sizes in prevention trials and other research to address health disparities with these distinct populations can prove difficult, often impractical, and cost-prohibitive. These same challenges associated with sample size are often shared by prevention researchers interested in working with numerous other groups, including rural residents, people with disabilities, people with substance abuse problems, and people with HIV/AIDS.

Challenges in Research with Small Sample Sizes

The seriousness of the problems associated with small sample sizes and the paucity of recent literature to address these issues have now reached a crisis point for many ethnic minority and health disparities researchers. Many recent advances in multivariate statistics require large samples, and state-of-the-art research designs often call for rigid design parameters rarely attainable given logistical demands of health disparities research, which is often conducted in difficult to access, or remote, or otherwise challenging locations. Small sample sizes frequently preclude use of an array of statistical methods that allow for the analysis of multilevel change processes in response to preventive interventions, and multivariate modeling of complex health related biomedical and psychosocial phenomena. Many of these current statistical techniques, including multilevel modeling (MLM), structural equation

modeling (SEM), generalized linear modeling (GLM), and item response theory (IRT), appropriate for testing intervention effects, hypothetical models, the adequacy of measures, and other research questions, often provide correct estimates only with large samples, as they have unknown or undesirable properties with small samples (Forero & Maydeu, 2009; Hoyle, 1999; Kenny, Mannetti, Perro, Livi, & Kashy, 2002). As a result, small samples prevention and health disparities researchers are often required to apply less sophisticated statistical methods that do not fit their research questions as well, or to alter their research questions in order to fit simpler statistical techniques (Hoyle, 1999). Such alternatives are problematic on numerous levels, most prominently; they erect large disincentives to doing this type of research.

Over the past three decades, some researchers have addressed the statistical issues associated with small sample sizes, suggesting potential solutions (Boomsma, 1983; Hoyle, 1999; Tanaka, 1987). For example, Kenny et al. (2002) provided extensions of MLM that address some of the statistical problems of small samples through (1) development of procedures to accommodate negative nonindependence in the model, and (2) use of an actor-partner interaction model that combines results from a between-within analysis to model group effects.

Ironically, with culturally distinct populations, a small sample can often represent a larger proportion of the population of interest than is the case in large research samples in studies drawn from majority populations. Experience highlights the need to develop methodologies to fit these contexts, rather than continuing to try to fit the context to the methods available. A final irony can be found in the original justifications made for the early uses of statistics in research, which were based in the ability of those early statistics to draw inferences about populations using a relatively small sample drawn from the population of interest.

What is Small?

The definition of ‘small’ can be surprisingly elusive. With assignment to intervention conditions at the community or school level, prevention researchers with even very large prevention trials may face small numbers of actual degrees of freedom in an analysis. In addition, research questions of large importance to prevention science, such as in health disparities and ethnic minority health research, may involve small samples because of the size of the distinct population that is affected. Accordingly, prevention researchers engaged in potentially important studies of these types often receive negative journal or grant reviews that describe their designs as underpowered due to their sample sizes.

It is therefore tempting to define “small” merely in terms of statistical power. However, lack of power may result from weak effects as much as from sample size. A sample size that is adequate for a medication study with strong effects may be insufficient for a psychosocial prevention trial with more modest effect sizes. Other definitions of “small” may focus on additional indicators that include instability in analyses, overfitting in parameterization and models specification, and violations of statistical assumptions.

Common to these many definitions of “small” is a shared concern regarding the extent to which individual observations have influence on the results of an analysis. Randomness can be defined as the degree to which individual observations in a sample may be replaced with other observations drawn from the same population, without reducing the representativeness of the sample. In small samples randomness is impacted; the potential influence of outlying observation is amplified, reducing representativeness. The nature of the influence of individual observations will also vary across methods. An outlying observation can exert leverage on an ordinary least squares (OLS) regression line, whereas a teacher who misreads instructions may influence an entire classroom or school that is one “subject” in the level of a MLM analysis of a school-based prevention trial. Finally, a single aberrant correlation or covariance between indicators may compromise an entire SEM analysis.

Frequentist (classical) statistical methods are based on the assumption that sampling distributions provide unbiased estimates of population parameters. Whatever the precise meaning of influence for the particular technique being used, as a general rule, in frequentist methods, as samples decrease in size, the potential for single observations to bias parameter estimates increases.

A simple simulation using OLS regression to estimate the relation between sample size, effect size, and influence (i.e., the extent to which the removal of a single case changes the effect estimate) illustrates this issue. This simulation calculates 100 bivariate regression models for every combination of five effect sizes ($r = .1, .3, .5, .7, .9$) and five sample sizes ($N=10, 20, 50, 100, 200$). Influence was significantly and negatively related to both effect size ($B = -.11, t(2496)=15.45, p < .01$) and sample size ($B = -.001, t(2496)=25.46, p < .01$). The interaction between sample size and effect size was also significant and positive ($B = .006, t(2496)=8.36, p < .01$), suggesting that larger effect sizes to some extent mitigate the effect of small sample size upon influence. However, examination of the data reveals a likely threshold in this effect. The maximum influence value extracted from each simulation is presented in Figure 1, where each panel represents an effect size with sample size on the X-axis and maximum influence produced on the Y-axis. Samples of 50 or larger seldom return observations with high influence, however samples of 10 and 20 do so frequently. Surprisingly, though it is somewhat diminished, this threshold effect appears to continue to exist even with very strong effect sizes ($r = .7$ or $r = .9$), raising the question of similar influence in research using this and other statistical techniques with sample sizes of 10 or 20.

We propose it is this issue of influence that is at the core of the meaning of “small.” Hoyle (1999) notes that statistical techniques are differentially influenced by small samples. Samples smaller than 50 may negatively impact multiple imputation, but bootstrapping procedures appear to work well with samples as small as 20, while confirmatory factor analysis typically requires samples of at least 100. These differences across techniques and methods themselves stem from method differences in their sensitivity to influence by individual observations. This issue of influence is particularly a concern for parametric techniques, is less of an issue for nonparametric techniques, and is not an issue for Bayesian approaches. Accordingly, each paper in this Special Section can be understood through the strategies it proposes to manage this issue of influence.

The strategies presented in the “Advancing Small Sample Prevention Science” Special Section represent efforts to cope with situations where sample size threatens current assumptions upon which the analyses of prevention trials are based. The goals of the Special Section are to (1) suggest some defining features of small sample research and identify some of the key research design and analytic issues that arise in prevention research with small samples, (2) develop applied, problem-oriented, and methodologically innovative solutions to these design and analytic issues, and (3) evaluate the potential role of these innovative methods in describing phenomena, testing theory, and evaluating interventions. Through these efforts, we hope to promote broader application of these methodological innovations to more general problems in design and analysis both within and beyond small sample research.

Directions for Advancing Small Sample Prevention Science

This Special Section aims to propose innovative applications of statistical methodologies for the design and analysis of small sample data. Specific targeted areas for innovation include: (1) research designs and analytic methods that can maximize statistical power for analyses of interventions conducted with small, culturally distinct samples—examples include dynamic wait list research designs, Bayesian approaches, matching, and imputation; (2) strategies for reducing error and bias in measures applied in studies with culturally distinct samples, and (3) use of qualitative methods and mixed methods combining qualitative and quantitative data.

A primary strategy in small sample research involves application of methods that increase statistical power, conventionally defined as one minus the probability of failing to reject the null hypothesis (H_0) given that H_0 is false. Concerns regarding power are particularly important in research exploring interventions of medium or small effect sizes, and especially important in cases when these changes constitute potentially important effects (Prentice & Miller, 1992), as is often the case in prevention research. The typical advice for increasing statistical power is to increase sample size. However, strategies are needed for maximizing power when increasing sample size is not feasible or possible. For example, dynamic wait list designs, which randomly assign participants to different start times, increase statistical power by increasing the number of time periods available for statistical analysis. Bayesian statistics (Kadane, this issue; Kaplan, 2014) incorporate prior information in the analysis, which can also increase power to detect an effect. These and other types of power maximization strategies are often poorly understood by applied researchers.

Imputation techniques enhance power by making the full use of the data provided by each participant when some of the data cells or time points are missing. Imputations procedures have become increasingly advanced in recent years (de Jong & Spiess, 2015; van Buuren, 2011); these advances hold special promise for small sample research, in which the data from each individual observation is of heightened importance. However, small samples studies often do not yield sufficient data to allow good estimation in many imputation approaches, and those state of the art imputation approaches potentially most useful in small samples analysis, are often difficult to use and not currently implemented in statistical packages.

A second potential small sample research strategy involves reducing error and bias in measures. Classical test theory (Allen & Yen, 2001) advocates adding items in order to increase the reliability of measures. Item response theory (IRT; Lord, 1980) instead focuses on evaluating the functioning of each individual item, often through evaluation of the item through its graded level of item difficulty (Samejima, 1996). This approach typically results in scales with more precise measurement qualities than longer measures of the same construct developed using classical approaches (Henry, Pavuluri, Youngstrom, & Birmaher, 2008). However, IRT methods were developed for use with large samples (Hoyle, 1999), and little guidance exists by way of simulations or examples of implementation with real data regarding the possibilities and limitations in the use of IRT approaches with small samples.

A third strategy for innovation with small sample size involves the use of qualitative data, including a number of recent advances in mixed-methods approaches. There have been notable examples of important new findings produced through small sample research (e.g., Castro & Coe, 2007) that merge qualitative and quantitative analytic strategies in ways that pool the unique strengths aligned with each approach. Qualitative inquiry can more completely inform basic findings in quantitative designs with insufficient power to explore interactions and multivariate relationships, and mixed methods affords opportunities to capitalize upon the strengths of both methodological traditions. What is currently lacking are systematic descriptions of approaches that provide integration of qualitative and quantitative data. One promising possibility involves the use of cluster analysis to assist qualitative researchers in deeper understanding of interrelations in their coded data (Henry, Tolan, & Gorman-Smith, 2005). Descriptions are lacking of the relative strengths of different clustering methods in application with the types of small samples typical in qualitative research.

Overview of the Special Section

The following articles aim to provide an overview of these potential strategies for small sample size in the three areas described above. The Special Section is organized into two subsections. The first subsection provides input on research design, while the second focuses on analysis. Each article within a subsection provides description of an innovative solution to one or more small samples challenges. A final article then summarizes the different approaches, with particular emphasis upon their applicability to testing for intervention effects in prevention research, identifying a set of themes arising across the approaches. These themes provide a set of conclusions regarding future directions in small sample research.

Small Sample Research Design

Hopkin, Hoyle, and Gottfredson (2015) open the first subsection by providing a definition of “small” based in pragmatism, meaning near the “lower bound” of the size required by a particular statistical model for it to perform satisfactorily. They then present a series of practical optimization strategies for “maximizing the yield” of data from small samples when sample size cannot be increased. The authors emphasize the maximization of statistical power, then also highlight approaches that go beyond power maximization alone.

These include strategies that can still extract valuable information from a study when statistical power is inadequate for hypothesis testing, and highlight the usefulness of graphical methods when sample size is too small for inferential statistics.

In the next article, Wyman, Henry, Knoblauch, and Brown (2015) propose two research designs, the dynamic wait-listed design (DWLD) and the regression point displacement design (RPDD), as research design alternatives to the randomized controlled trial (RCT) that have particular utility with small samples, including those cases when group-based intervention are involved. Both the DWLD and RPDD aim to increase efficiency and statistical power, while also allowing for flexibility that enables balancing community needs and research priorities. The DWLD extends the traditional wait-listed controlled design by increasing the number of time periods in which individuals or groups are randomized to receive intervention. The RPDD compares intervention units to their expected values. These expected values are obtained from archival or other existing data on a large number of non-intervention units prior to and following the time period of the intervention. Examples of the DWLD and RPDD are provided in the paper.

Fok, Henry, and Allen (2015) extend the work of Wyman and colleagues (2015) on the DWLD by discussing a related design, the stepped wedge design (SWD), and by introducing an additional small sample design alternative to RCT, the interrupted time-series design (ITSD). They discuss similarities and differences between the SWD and DWLD in their historical origin and application, along with differences in the statistical modeling of each design. The authors then describe the main design characteristics of the interrupted time series design (ITSD), as well as some of its strengths and limitations. Case examples are provided for both SWD and ITSD. Fok et al. conclude with a critical review of the ITSD, SWD, DWLD, and RPDD, along with a discussion of the types of contextual factors that prevention researchers working with small samples should consider in selecting an optimal research design.

Little attention is paid in prevention research to the ability of measures to accurately assess change; this property can be termed “responsiveness” or “sensitivity to change.” Fok and Henry (2015) review definitions and measures of responsiveness, and suggest five strategies for increasing sensitivity to change, with a central focus on prevention research with small samples. These strategies include (a) improving understandability and cultural validity, (b) assuring that the measure covers the full range of the latent construct being measured, (c) eliminating redundant items, (d) maximizing sensitivity of the device used to collect responses; and (e) asking directly about change. Examples from research with small samples are discussed.

Analysis of Small Samples Data

The second subsection of the Special Section aims at providing guidance for researchers in the analysis phase by offering both special considerations and alterations in conventional statistical procedures, and alternative statistical procedures for use with small samples. In this subsection, Hoyle and Gottfredson (2015) begin by discussing the role of sample size in two statistical techniques that are widely used in prevention research: multilevel modeling (MLM) and structural equation modeling (SEM). For both statistical techniques, the authors

draw from simulation studies to examine the minimum sample size needed to produce reliable results. Suggestions are provided to address the way in which prevention researchers could make the best use of these statistical approaches when the sample sizes are close to the minimum recommended sizes.

The next three articles propose alternative strategies for the analysis of small samples. First, prevention studies often involve outcomes consisting of ordered categories, or ordinal data, whose distributions violate the assumptions of normality central to most statistical techniques. In light of considerations associated with randomness discussed above, the impact of these violations is especially impactful in small samples research. Hedeker (2015), in the next paper, proposes a method that addresses this concern by allowing for the statistical modeling of multilevel ordinal data. The approach permits the analysis of clustered or longitudinal data by making use of proportional odds regression. Hedeker provides a description of the approach and several examples in application of the method. He concludes by discussing computational issues, including considerations in identifying the least biased estimation procedures, particularly in cases with small samples. This new approach has considerable promise in advancing sophisticated multilevel analyses of small sample intervention studies.

An alternative strategy for the analysis of data when samples prove too small for inferential statistics involves use of qualitative analysis. Recent developments in mixed-methods approaches combine the use of qualitative and quantitative data. Cluster analysis provides a potential mixed methods approach facilitating more full integration of qualitative and quantitative data. In their paper, Henry, Dymnicki, Mohatt, Kelly, and Allen (2015) discuss the use of cluster analysis in prevention studies. They describe how cluster analysis provides new insights into coded qualitative data obtained from interviews allowing tracking and understanding of the motives of intervention participants and of factors influencing their involvement in intervention activities. Three cluster methods—latent class analysis, hierarchical cluster analysis, and K-Means clustering—are used to group participants according to similar profiles of codes. Using simulation, the authors examine the functioning and accuracy level of these three methods with samples as small as 50, drawing from a real world application of the method to highlight implications of this approach for use in small sample prevention research.

Finally, Bayesian statistics offers a third strategy providing a promising alternative in analytic approaches for prevention science. Bayesian analysis has particular promise and notable strengths in application with small samples. In the final article, Kadane (2015) discusses the Bayesian approach, providing an application of the approach with a case example using an existing prevention data set. Through this case example, Kadane examines ways in which Bayesian techniques uniquely address a number of the issues that arise in prevention research with small samples.

Future Directions

Henry, Fok, and Allen (2015) conclude the Special Section with a summary of the innovations proposed, describing how they address specific areas of concern in small sample

research. They evaluate the significance of each approach in terms of both innovation and utility, and provide a set of conclusions regarding the various approaches. These research design and analytic techniques allow researchers to describe phenomena, test theory, and evaluate interventions in small sample research, with special emphasis on testing for intervention effects.

Increasingly, prevention researchers across multiple disciplines and fields working in health disparities areas are calling for methods that allow for rigorous, scientific inquiry to address these important societal concerns. Together, the articles and the final commentary in this Special Section distill the challenges and hopes by affording researchers a number of concrete solutions to problems in small sample research, and in doing so, provide promising future directions for prevention science.

Acknowledgments

This Special Section of the *Prevention Science* was supported through a grant from the National Institute on 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 also provided through grants from the National Institute on Drug Abuse, the National Institute on Alcohol Abuse and Alcoholism, the National Institute on Minority Health and Health Disparities, and the National Institute of General Medical Services [T32 DA037183, R21AA016098, RO1AA11446; R21AA016098; R24MD001626; P20RR061430].

References

- Allen, MJ.; Yen, WM. Introduction to measurement theory. Long Grove, IL: Waveland Press Inc; 2001.
- Anderson J. Unraveling health disparities: Examining the dimensions of hypertension and diabetes through community engagement. *Journal of Health Care for the Poor and Underserved*. 2005; 16:91–117. DOI: 10.1353/hpu.2005.0121 [PubMed: 16327099]
- Boomsma, A. On the robustness of LISREL (maximum likelihood estimation) against small sample size and nonnormality. Amsterdam: Sociometric Research Foundation; 1983.
- Castro F, Coe K. Traditions and alcohol use: A mixed-methods analysis. *Cultural Diversity and Ethnic Minority Psychology*. 2007; 13:269–84. DOI: 10.1037/1099-9809.13.4.269 [PubMed: 17967095]
- Centers for Disease Control and Prevention (CDC). CDC Health Disparities and Inequalities Report – United States. *MMWR*. 2011; 60(Supplement):1–114. 2011.
- de Jong, R.; Spiess, M. Robust multiple imputation. In: Engel, U.; Jann, B.; Lynn, P.; Scherpenzeel, A.; Sturgis, P., editors. *Improving survey methods: Lessons from recent research*. New York, NY: Routledge/Taylor & Francis Group; 2015. p. 397–411.
- Fok CCT, Henry D. Increasing the sensitivity of measures to change. *Prevention Science*. 2015; doi: 10.1007/s11121-015-0545-z
- Fok CCT, Henry D, Allen J. Research designs for intervention research with small samples II: Stepped wedge and interrupted time-series designs. *Prevention Science*. 2015; doi: 10.1007/s11121-015-0572-9
- Forero C, Maydeu-Olivares A. Estimation of IRT graded response models: Limited versus full information methods. *Psychological Methods*. 2009; 14:275–299. DOI: 10.1037/a0015825 [PubMed: 19719362]
- Giger JN, Davidhizar R. Promoting culturally appropriate interventions among vulnerable populations. *Annual Review of Nursing Research*. 2007; 25:293–316.

- Hedeker D. Methods for multilevel ordinal data in prevention research. *Prevention Science*. 2015; doi: 10.1007/s11121-014-0495-x
- Hedeker, D.; Gibbons, RD. *Longitudinal data analysis*. Hoboken, NJ: Wiley-Interscience; 2006.
- Henry D, Dymnicki AB, Mohatt N, Allen J, Kelly JG. Clustering methods with qualitative data: A mixed methods approach for prevention research with small samples. *Prevention Science*. 2015; doi: 10.1007/s11121-015-0561-z
- Henry D, Pavuluri MN, Youngstrom E, Birmaher B. Accuracy of brief and full forms of the Child Mania Rating Scale. *Journal of Clinical Psychology*. 2008; 64:1–14. DOI: 10.1002/jclp.20464 [PubMed: 18161043]
- Henry D, Tolan PH, Gorman-Smith D. Cluster analysis in family psychology research. *Journal of Family Psychology*. 2005; 19:121–132. DOI: 10.1002/jclp.20464 [PubMed: 15796658]
- 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*. 2015; doi: 10.1007/s11121-014-0542-7
- Hoyle, RH. *Statistical strategies for small sample research*. London: Sage Publications; 1999.
- Hoyle RH, Gottfredson NC. Sample size considerations in prevention research applications of multilevel modeling and structural equation modeling. *Prevention Science*. 2015; doi: 10.1007/s11121-014-0489-8
- Institute of Medicine(IOM). *Unequal treatment: Confronting racial and ethnic disparities in health care*. Washington, DC: The National Academies Press; 2002.
- Kadane JB. Bayesian methods for prevention research. *Prevention Science*. 2015; doi: 10.1007/s11121-014-0531-x
- Kaplan, D. *Bayesian statistics for the social sciences*. New York, NY: Guilford; 2014.
- Kenny DA, Mannetti L, Pierro A, Livi S, Kashy DA. The statistical analysis of data from small groups. *Journal of Personality and Social Psychology*. 2002; 83:126–37. DOI: 10.1037/gdn0000022 [PubMed: 12088122]
- Lord, FM. *Applications of item response theory to practical testing problems*. Mahwah, NJ: Lawrence Erlbaum Associates; 1980.
- Murray CJL, Kulkarni SC, Michaud C, Tomijima N, Bulzacchelli MT, Iandiorio TJ, Ezzati M. Eight Americas: Investigating mortality disparities across races, counties, and race-counties in the United States. *PLoS Med*. 2006; 3(12):e545.doi: 10.1371/journal.pmed.0030545
- Prentice D, Miller DT. When small effects are impressive. *Psychological Bulletin*. 1992; 112:160–164. DOI: 10.1037/0033-2909.112.1.160
- Samejima, F. Graded response model. In: van der Linden, WJ.; Hambleton, RK., editors. *Handbook of modern Item Response Theory*. New York, NY: Springer Publishing Company; 1996. p. 85-100.
- Tanaka JS. ‘How big is big enough?’: Sample size and goodness of fit in structural equation models with latent variables. *Child Development*. 1987; 58:134–46. DOI: 10.1111/j.1467-8624.1987.tb03495.x
- van Buuren, S. Multiple imputation of multilevel data. In: Hox, J.; Roberts, JK., editors. *Handbook for advanced multilevel analysis*. New York, NY: Routledge/Taylor & Francis Group; 2011. p. 173-196.
- Wyman PA, Henry D, Knoblauch S, Brown CH. Designs for testing group-based interventions with limited number of social units: The dynamic wait-listed and regression point displacement designs. *Prevention Science*. 2015; doi: 10.1007/s11121-014-0535-6

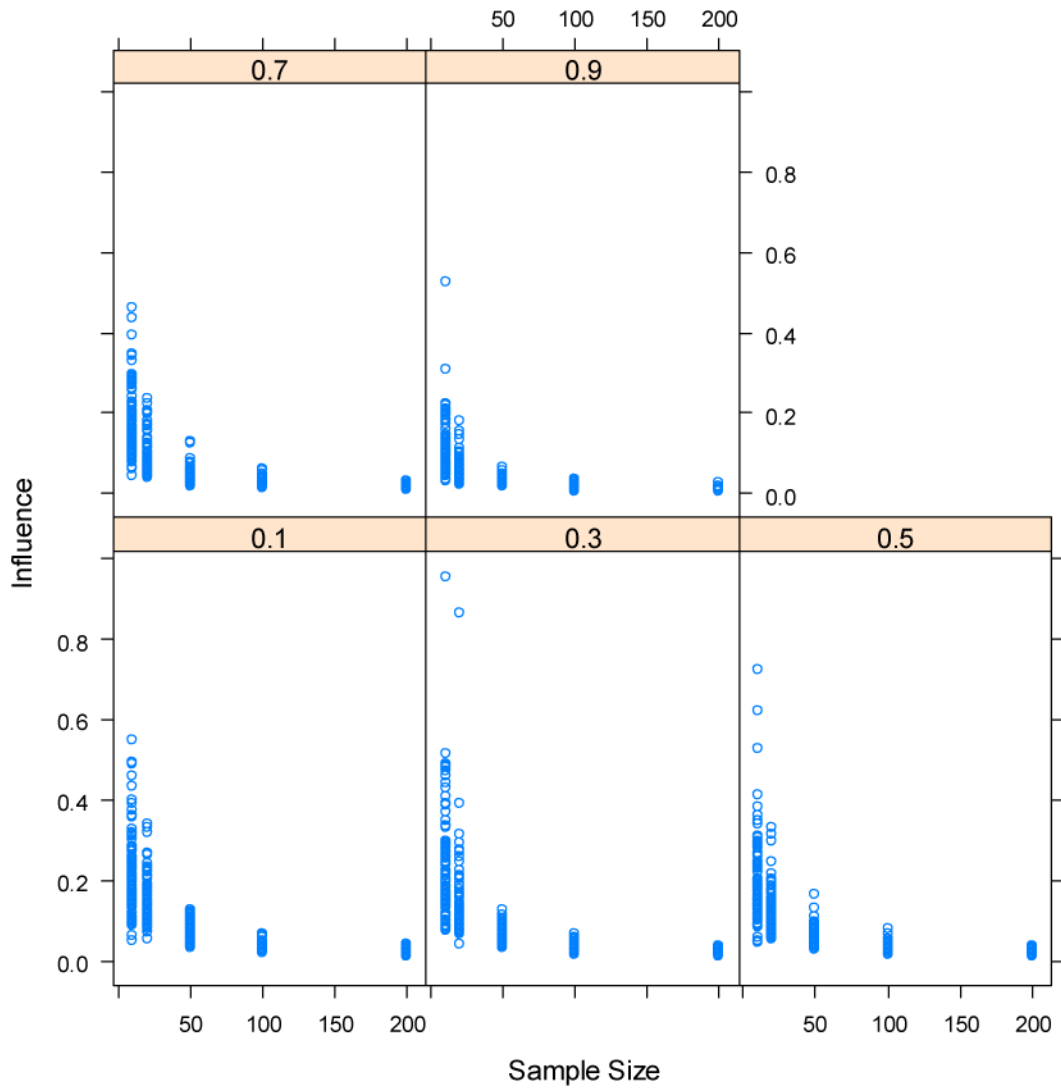


Figure 1. Influence of individual observations as a function of sample size (x -axis) and effect size r (panel headings). Maximum influence value produced in a simulation calculating 100 bivariate ordinary least squares regression models for every combination of five effect sizes ($r = .1, .3, .5, .7, .9$) and five sample sizes ($N = 10, 20, 50, 100, 200$). Influence is change in regression weight when an observation is removed (y -axis).