

Health Disparities. Irene Dankwa-Mullan is also a guest editor for this supplement issue.

Correspondence should be sent to Irene Dankwa-Mullan, Acting Deputy Director, Division of Extramural Scientific Programs, National Institute on Minority Health and Health Disparities, 6707 Democracy Blvd, Bethesda, Maryland 20817 (e-mail: dankwamullani@mail.nih.gov). Reprints can be ordered at <http://www.ajph.org> by clicking the "Reprints" link.

This article was accepted May 4, 2015.

doi:10.2105/AJPH.2015.302756

Note. The findings and conclusions in this editorial are those of the authors and should not be construed to represent the views of the National Institutes of Health or any of the sponsoring organizations and agencies or the US government.

Small Is Essential: Importance of Subpopulation Research in Cancer Control

Contributors

I. Dankwa-Mullan led the writing and conceptualized the issues outlined.

Y. T. Maddox provided the vision for the science vision and added editorial comments. Both authors contributed significantly to the editorial.

References

1. Heckler MM. *The Heckler Report. Report of the Secretary's Task Force on Black and Minority Health*. Department of Health and Human Services, Office of Minority Health. 1985. Available at: <http://archive.org/stream/reportofsecretar00usde#page/n1/mode/2up>. Accessed April 28, 2015.
2. US Department of Health and Human Services, Office of Disease Prevention and Health Promotion. Healthy

People 2020. Available at: <http://www.healthypeople.gov>. Accessed May 5, 2015.

3. Institute of Medicine. *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care*. Washington, DC: The National Academies Press; 2002.
4. Shonkoff JP. Leveraging the biology of adversity to address the roots of disparities in health and development. *Proc Natl Acad Sci U S A*. 2012;109(suppl 2):17302–17307.
5. Provencal, N. Booij, L. and Tremblay RE. The developmental origins of chronic physical aggression: biological pathways triggered by early life adversity. *J Exp Biol*. 2015;218(pt 1):123–133.

6. Shannon KE, Beauchaine TP, Brenner SL, Neuhaus E, Gatzke-Kopp L. Familial and temperamental predictors of resilience in children at risk for conduct disorder and depression. *Dev Psychopathol*. 2007;19(3):701–727.
7. van der Werff SJA, van den Berg SM, Pannekoek JN, Elzinga BM, van der Wee NJ. Neuroimaging resilience to stress: a review. *Front Behav Neurosci*. 2013;7:39.
8. Hertzman C. Putting the concept of biological embedding in historical perspective. *Proc Natl Acad Sci U S A*. 2012;109(suppl 2):17160–17167.
9. McEwen BS. Brain on stress: how the social environment gets under the skin. *Proc Natl Acad Sci U S A*. 2012;109(suppl 2):17180–17185.

The ability to harness the benefits of “big data” has had a revolutionary impact on science, with its focus on the volume and variety of data sources, and application of both traditional and innovative analytic methods appropriate for large, aggregated data sets. We are concerned, however, about the opposite: “small data,” for which the size, dispersion, or accessibility of the population of interest makes it difficult to obtain adequate sample sizes to test specific research questions. Examples include racial or ethnic subpopulations (e.g., Honduran Latin Americans), populations occurring in specific geographic areas (e.g., reservations), and populations that have relatively rare characteristics (e.g., transgender persons). A great challenge is determining when a small group is of practical or theoretical interest (Figure 1). We define “practical and theoretical interest” broadly to include issues involving social justice, biological or geographic factors, and disease burden.¹ Ultimately, it is critical to ensure that all segments of the US population benefit from this

research and from the latest technologic advances in cancer care services and delivery.

INCLUDING UNDERREPRESENTED GROUPS IN RESEARCH

An example of the potential negative ramifications of not including underrepresented groups in research—or inappropriately aggregating them across groups—comes from the study of racial and ethnic health disparities and issues of equity in the United States. Intervention research often does not include a wide range of racial/ethnic subgroups; so it is not feasible to test whether an intervention created specifically for the majority group is also efficacious for the subgroups. Likewise, the ability to test whether an intervention can be altered for a particular subgroup is also often not possible. Epidemiological and surveillance research usually involves the inclusion of “minority or underserved populations” in addition to White or non-Hispanic White (NHW) groups. While this has allowed for a better understanding of these

smaller populations and provides some progress toward addressing health inequities, there remain pockets of communities that are severely underrepresented within the broader “minority and underserved populations.”^{2–6}

As a further example, although Asian Americans as a whole have high incomes and good health outcomes overall when compared with NHWs, Hispanics, African Americans, and American Indian/Alaska Natives, this generalized statistic masks the fact that subgroups of Asian Americans, such as the Cambodians and Hmong, lag severely behind other Asian Americans.^{3,4,7,8} Even within the NHW population there are communities that have long been disadvantaged (such as those living in Appalachian states), with low levels of income, literacy, and health outcomes.^{9–11} These subgroups have generally been omitted or excluded from the research process because of challenges with identification and recruitment. Through this commentary, we hope to encourage research in subpopulations; we recommend both the development of new methods and the innovative use of existing

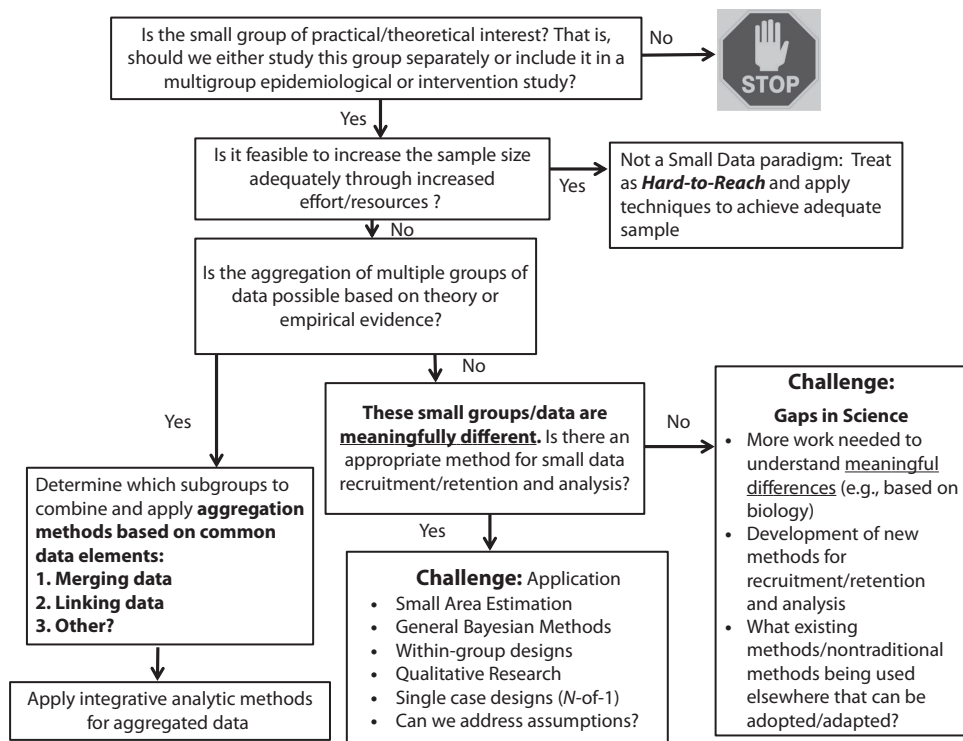


FIGURE 1—Research with small data: identifying challenges.

methodological and analytic strategies across both intervention and epidemiological research.

ALTERNATIVE STUDY DESIGNS AND ANALYSIS PROCEDURES

There is growing recognition that to implement interventions in small populations, it may be necessary to consider alternative study designs, such as the use of single-case designs attributing propensity scores, and randomized group designs. In 2013, many of the studies submitted to the Division of Cancer Control and Population Sciences at the National Cancer Institute (NCI) on subpopulation research that did not score well in peer review received comments that the

randomized clinical trial design was not appropriate because the sample size was insufficient to detect changes in the effect of the intervention. This criticism raises the question of whether these studies would be better suited for alternatives to the standard randomized control trial design, such as single-case, within-subject controls, and a variety of quasi-experimental designs.

One solution for testing interventions in small samples is to focus on within-rather than between-group designs. Because a within-group design uses the sample as its own control, there is no need for a separate control group, reducing by up to half the sample size required for accurate statistical comparisons. Among group designs, there are a number

of quasi-experimental approaches that could be considered, including interrupted time series¹²⁻¹⁵ and stepped wedge designs, the latter being particularly useful for studies in which there are distinct and dispersed cohorts or communities in which the intervention can be rolled out in a staggered manner.^{16,17} Single-case studies involving a series of *N-of-1* trials could be used to test intervention adaptations in an iterative manner, and Bayesian estimates can be produced from this series of trials to evaluate the potential generalizability of the findings to the subpopulation.¹⁸ Within-subject designs require more intensive longitudinal data than typically obtained through between-subject designs, but the advent of technologies for capturing temporally

dense data, such as ecological momentary assessment and passive sensor technologies, makes these approaches more viable. Such data could also be used in conjunction with multilevel analyses of behavior across different spatial areas. This kind of study design can be statistically powerful, even with modest numbers of samples per geographic unit.¹⁹

For epidemiological research, innovative recruitment methods may be very useful. For example, respondent-driven sampling^{20,21} has been successfully employed to identify and recruit groups for studies in which there is no existing sample frame, such as drug addicts or ethnic subgroups. Innovative analytic approaches, such as integrative data analysis,²² could be employed where independent data sets are combined together and analyzed as a whole to produce adequate representation and sample sizes. Integrative data analysis can also be used to combine data across multiple iterations of the same national survey where any one sample does not constitute an adequate sample size.

ADDRESSING THE CHALLENGE OF SMALL DATA

The National Institutes of Health (NIH)—and by extension the NCI—has an obligation to conduct research to improve the health of all Americans, not just the health of the majority population or those who are easy to identify. We therefore recommend the development and the use of methodological and analytic procedures to allow subpopulations to be meaningfully included in research. Figure 1 illustrates a model for determining when a “small”

group is of research interest. However, it is also clear that other entities need to be involved in identifying populations of interest and in developing initiatives to address these groups, not just those who are responsible for grant funding decisions. For example, at the NIH, training for peer reviewers in study sections may be needed to ensure that they are knowledgeable about these innovative methods so that sound, rigorous scientific applications that employ them are understood and scored appropriately.

In addressing the above issues, NCI is planning a workshop to address three areas related to small populations:

- (1) identification, recruitment, and retention strategies;
- (2) epidemiological design and analytic approaches for small samples; and
- (3) intervention design and analytic approaches for subpopulations.

Based on the products of this workshop and responses to this editorial, the NCI will explore next steps to strengthen subpopulation research. ■

Shobha Srinivasan, PhD
Richard P. Moser, PhD
Gordon Willis, PhD
William Riley, PhD
Mark Alexander, MSc
David Berrigan, PhD, MPH
Sarah Kobrin, PhD, MPH

About the Authors

All of the authors are with the Division of Cancer Control and Population Sciences, National Cancer Institute, Rockville, MD.

Correspondence should be sent to Shobha Srinivasan, Health Disparities Research Coordinator National Cancer Institute, Division of Cancer Control and Population Sciences, 9609 Medical Center Drive, Room 4E432, MSC 9764, Rockville, Maryland 20850 (e-mail: ss688k@nih.gov). Reprints can be ordered at <http://www.ajph.org> by clicking the "Reprints" link.

This editorial was accepted August 12, 2015.

doi:10.2105/AJPH.2014.302267

Contributors

All authors contributed equally to this editorial.

Acknowledgments

The authors would like to thank and acknowledge the guidance and support of Robert T. Croyle, PhD. The authors would also like to thank Kathleen Ulan- day, MPH, for assistance with the cita- tions.

References

1. Waller LA, Gotway CA. *Applied Spatial Statistics for Public Health Data*. Hoboken, NJ: John Wiley & Sons; 2004.
2. Yancey AK, Ortega AN, Kumanyika SK. Effective recruitment and retention of minority research participants. *Annu Rev Public Health*. 2006;27:1–28.
3. Islam NS, Khan S, Kwon S, Jang D, Ro M, Trinh-Shevrin C. Methodological issues in the collection, analysis, and reporting of granular data in Asian American populations: historical chal- lenges and potential solutions. *J Health Care Poor Underserved*. 2010;21(4):1354–1381.
4. Ghosh C. Healthy People 2010 and Asian Americans/Pacific Islanders: defining a baseline of information. *Am J Public Health*. 2003;93(12):2093–2098.

5. Institute of Medicine. *The Health of Lesbian, Gay, Bisexual, and Transgender Aging: Biographical Approaches for Inclusive Care and Support*. London, UK: Jessica Kingsley; 2012.
6. Fredriksen-Goldsen KI, Kim HJ, Barkan SE, Muraco A, Hoy-Ellis CP. Health disparities among lesbian, gay, and bisexual older adults: results from a population-based study. *Am J Public Health*. 2013;103(10):1802–1809.
7. Chen MS Jr, Hawks BL. A debunking of the myth of healthy Asian Americans and Pacific Islanders. *Am J Health Promot*. 1995;9(4):261–268.
8. National Heart, Lung, and Blood Institute. Addressing Cardiovascular Health in Asian American and Pacific Islanders: a Background Report. 2000. Available at: <http://www.nhlbi.nih.gov/health/prof/heart/other/aapibkgd/aapibkgd.pdf>. Accessed June 25, 2014.
9. Behringer B, Friedell GH. Appalachia: where place matters in health. *Prev Chronic Dis*. 2006;3:A113.
10. Appalachian Regional Commission. Underlying Socioeconomic Factors Influencing Health Disparities in the Appalachian Region. Final Report. 2008. Available at: http://www.arc.gov/assets/research_reports/SocioeconomicFactorsInfluencingHealthDisparitiesinAppalachianRegion5.pdf. Accessed June 25, 2014.
11. Vanderpool RC, Gainor SJ, Conn ME, Spencer C, Allen AR, Kennedy S. Adapting and implementing evidence-based cancer education interventions in rural Appalachia: real-world experiences and challenges. *Rural Remote Health*. 2011;11(4):1807.
12. Cook TD, Campbell DT. *Quasi-Experimentation: Design & Analysis Issues for Field Settings*. Boston, MA: Houghton Mifflin Company; 1979.
13. Hartmann DP, Gottman JM, Jones RR, Gardner W, Kazdin AE, Vaught RS. Interrupted time-series analysis and its application to behavioral data.

- J Appl Behav Anal*. 1980;13(4):543–559.
14. Gillings D, Makuc D, Siegel E. Analysis of interrupted time series mortality trends: an example to evaluate regionalized perinatal care. *Am J Public Health*. 1981;71(1):38–46.
15. Biglan A, Ary D, Wagenaar AC. The value of interrupted time-series experiments for community interven- tion research. *Prev Sci*. 2000;1(1):31–49.
16. Mdege ND, Man MS, Taylor Nee Brown CA, Torgerson DJ. Systematic review of stepped wedge cluster ran- domized trials shows that design is particularly used to evaluate interven- tions during routine implementation. *J Clin Epidemiol*. 2011;64(9):936–948.
17. Woertman W, de Hoop E, Moerbeek M, Zuidema SU, Gerritsen DL, Teerenstra S. Stepped wedge designs could reduce the required sample size in cluster randomized trials. *J Clin Epidemiol*. 2013;66(7):752–758.
18. Zucker DR, Ruthazer R, Schmid CH. Individual (N-of-1) trials can be combined to give population com- parative treatment effect estimates: methodologic considerations. *J Clin Epidemiol*. 2010;63(12):1312–1323.
19. Raudenbush SW, Bryk AS. *Hierarchical Linear Models*. 2nd ed. Thousand Oaks, CA: Sage Publications; 2002.
20. Heckathorn DD. Respondent-driven sampling: a new approach to the study of hidden populations. *Soc Probl*. 1997;44(2):174–199.
21. Salganik MJ, Heckathorn DD. Sampling and estimation in hidden populations using respondent-driven sampling. *Sociol Methodol*. 2004;34:193–239.
22. Curran PJ, Hussong AM. Integrative data analysis: the simultaneous analysis of multiple data sets. *Psychol Methods*. 2009;14(2):81–100.