January 11, 2018

# What is a Small Population?[1]

Srinivasan and colleagues[2] (2015, p.1) provided their definition of a small population as one 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." They go on to note that "it is critical to ensure that all segments of the US population benefit from [health] research and from the latest technologic advances…"

Unfortunately, populations for which it is difficult to obtain adequate sample size are also likely to be expensive to study because dispersion and accessibility increase logistical costs. Hence, even if it is technically feasible to study a small population, it may not be easy to obtain funding for the study. This argues for increased efforts to document the needs, potential benefits, and methods for enhancing the efficiency of the study.

In other situations there may be no good sample frames because there is no agreed to definition of the population or a way to identify it. These, so called "hidden populations" are "small" by the above definition but raise more fundamental questions in health research and require additional data source work even to measure health disparities. Devers et al. (2013) provides some examples of this challenge and potential options for addressing it.

Much of the interest in studying health disparities for "small" populations was stimulated by the Department of Health and Human Service's Healthy People project[3] in 2010. The project, which set a vision and strategy for improved health outcomes by 2020, listed as one of their goals a desire to "achieve health equity, eliminate disparities, and improve the health of all groups." The elimination of these disparities and inequity is to be assessed across the following categories: race/ethnicity; gender; socioeconomic status; disability status; lesbian, gay bisexual, and transgender status; and geography. Several of these categories—in particular, some races, LGBT status, and some geographies—can be indicative of small populations.

The National Cancer Institute (NCI)[4] and Health Resources and Services Administration (HRSA)[5] both provide a wealth of information about measuring diversity of health outcomes and about measuring diversity in access to health care. Diversity is usually established by comparisons. A diversity measure of a small population may be compared to that of the U.S. population, or comparisons may be made among the diversity measures for its subpopulations. For example a small population of interest might be people whose work exposes them to a particular hazard,

---

[1] This is a living document prepared as background for the workshop by staff of the Committee on National Statistics, National Academy of Sciences, Engineering and Medicine with substantial input from sponsors, steering committee members and presenters.

[2] Authors are co-sponsors of this workshop.

[3] See, https://www.healthypeople.gov/sites/default/files/HP2020_brochure_with_LHI_508_FNL.pdf (December 2017).

[4] See, https://www.cancer.gov/about-nci/organization/crchd/about-health-disparities.

[5] See, https://bhw.hrsa.gov/shortage-designation/muap.

such as miners in a particular area or type of mine who experience poor health outcomes. Small populations can also occur from combinations of characteristics, e.g., members of American Indian and Alaska Native tribal groups who live in small distinct communities; immigrants who are “undocumented” by country of origin;  individuals at risk for HIV by category (men who have sex with men, sex workers, illicit drug users.) As such, the range of possibilities is large and contingent on a researcher’s specification of a research question and other covariates for control or study.

Small populations and the inference challenges associated with small sample size are inevitable due to variation in incidence of disease, prevalence of health-related behaviors, and heterogeneity in population characteristics. Promoting and strengthening research with small populations is of particular importance because substantial health disparities may arise from the combination of disparities in many small and distinct demographic groups. Lack of evidence concerning etiology of outcomes and most effective treatments for such groups may perpetuate disparities.

## Different Kinds of Health Research

Commonly used approaches in public health -- surveillance/epidemiological studies and intervention studies – involve different types of inferences.  Public health *surveillance* and *epidemiological* studies are generally accomplished through observational studies of the health status and health needs of population groups, either using existing data sources or designing surveys to collect needed information from a target population. These studies are strictly observational, with no attempt by the researcher to affect the outcome.  On the other hand, intervention studies examine the effect of a treatment, behavior modification or treatment delivery option on an outcome.  Examining the effects of an *intervention* requires a carefully designed study that may be referred to as an intervention, prevention, behavioral study, implementation study or clinical trial.

### *Surveillance/Epidemiological Study*

The goal of observational studies for *surveillance* is typically descriptive: to estimate the percentage of some target population or subpopulation within a geography, or the percentage of that population with health disparities or certain health outcomes. *Epidemiological* observational studies are analytical, seeking to estimate associations, risk factors, odds ratios, or relative risks. They may include cohort studies, case control studies, or cross-sectional analysis. Typically, available survey data, administrative records, registries, electronic health records, and other data may be used for surveillance studies or as a guide to selecting qualified individuals to survey/enroll in epidemiological studies.

For small populations, these studies may be challenged to find available data. The question of finding people for rare population research has been well addressed by the survey research community.  In this type of survey research, “small” might depend on how the population is perceived in relation to either a larger group, such as the rest of the US population or partitioned into subgroups by features such as race/ethnicity, geography or socioeconomic status. In his introductory chapter to the monograph on hard-to-survey populations, Tourangeau noted that “problems [in sampling] arise when a target population represents a small fraction of the frame population.” Tourangeau et al. (2014, p. 4).

Recently there has been considerable interest in conducting surveillance on so-called “hidden” populations—those that are not easily identifiable from administrative records and household/self-report surveys because individuals therein are reluctant to self-identify. In statistical terms, for these populations there is no sampling frame. Examples include the homeless, migrant workers, immigrants and various gender preference minorities. While survey research traditionally relies on probability sampling from a frame to make sure estimates that are derived are unbiased and generalizable to the target population, that method cannot be used to sample from hidden populations because information about the identify of individuals in the group is either not available or cannot be reliably ascertained from survey respondent’s reports. There are non-probability sampling approaches for reaching them, such as respondent-driven sampling, venue-based sampling, and on-line sampling; however, these techniques raise questions about potential bias and lack of generalizability if used to estimate a population size or disparity level. Lack of sample frames is a problem that requires more fundamental methods and data source work to be done, even to understand the basic issues about the population and its health.

In addition to issues with size, dispersion, and accessibility, small populations may also be hard-to-reach because their unwillingness to participate in research studies, or because of negative histories with social institutions and with past research. Tourangeau et al. (2014) listed many examples of the challenges with surveying these groups and possible methods that may be used. In general, this monograph provides a useful taxonomy for distinguishing and surveying hard-to-survey populations.

### *Intervention Study*

For the purpose of this workshop “intervention study” is meant to define any study with a goal to establish a causal effect of a treatment applied to individuals. The randomized controlled trial (RCT) is the gold standard in health intervention research. Other examples of intervention studies are prevention, behavioral, and implementation studies with different approaches to randomization. Clinical trials to establish drug efficacy are the most well-known examples of intervention studies and frequently use randomized controlled trials (RCT). The challenge, especially with small populations or small samples, include a number of logistical and ethical issues that can arise, along with inefficiencies in how the RCT makes use of information that can result in low power and low external validity. An RCT is also impractical in some real world settings such as dissemination and implementation studies.  Recent alternative designs (stepped wedge, interrupted time series, regression discontinuity, and dynamic waitlist) make use of optimization strategies to more efficiently use available information to maximize power with modest sample size. All of these new approaches have their strengths and weaknesses that should be carefully considered in any given situation.

IOM (2001) lists the following situations that might warrant a *small* clinical trial: rare diseases, unique study populations, individually tailored therapies, environments that are isolated, emergency situations, and public health urgency. The same situations might result in small samples for any type of intervention trial. IOM (2001) also provides a summary of statistical issues, designs and analysis approaches that might be useful for small clinical trials and provides the following recommendations for researchers designing such studies: define the research question; tailor the design; clarify methods of reporting of results; perform corroborative statistical analysis; and exercise caution in interpretation. IOM (2001, p 10) also recommended more federal funding of research on alternative designs for small sample studies.

The need to address methodological challenges concerning small populations is predicated on determining whether or not a “small” population is meaningfully different and should be studied. Answers to this question arise from concerns of populations, funders and researchers. These may differ. For the researchers and funders, answers to the “meaningfully different” question may arise from the significant amount of prior analysis including surveillance studies, epidemiological studies, laboratory studies, etc. that occur prior to funding. IOM (2010, pp. 2, 3, and 5-9) proposes that the L.E.A.D framework (for Locate Evidence, Evaluate Evidence, Assemble Evidence, and Inform Decisions) be followed when designing a study. It goes on to specify that first the researcher must identify the question to be answered by the study. Then the following steps should be followed: (1) Locate all the types of evidence that could be useful in answering the question. (2) Evaluate the quality of the evidence, especially its level of certainty (internal validity) and generalizability. (3) Develop a transparent and comprehensive summary of the evidence related to why an action should be taken, what that action should be, and how it should be taken. If evidence is limited, examine the potential for blending it with theory, professional experience, and local wisdom. (4) Use the summary to inform the decision-making process. Explicitly or implicitly, plans and proposals for research on small populations must address the meaningfully different criteria and consider the different needs of the relevant stakeholders. Lack of clarity concerning this issue may well be the source of some of the frustration felt over funding patterns addressing health in small populations.

If evidence is limited other options might be more initial data source work, accumulation of data over multiple small studies or efforts to understand mechanism in biological studies. Accumulation of data over multiple small studies might be challenging because slow accumulations of results require data harmonization and may be subject to secular change. Identification of appropriate biomarkers or intermediate endpoints may allow studies with larger or more easily obtainable outcomes.

In specifying the target population of the study, researchers may consider whether the research question would support the combination of the “small population” with others to make the study more manageable. A population may not be small if it can be combined with similar groups from other geographic areas. For example, developing interventions for Nicaraguan immigrants in the US might prove difficult because of the small population size. However, if relevant social, psychological and biological research suggested that the intervention approach was appropriate for Central American immigrants generally then the scope of the study might be expanded. Even in this case, whether or not to combine groups is a difficult question. There are huge differences in the environments in different Central American countries and immigrants in different parts of the U.S. experience health care differently. Proposals to group demographic subgroups in order to strengthen etiological or intervention studies require considerable discussion and sensitivity, especially where there has been history of injustice and health disparities.

In some instances, however, researchers may be even more challenged when subgroups of the small target population are ethno-culturally distinct. This cultural distinctiveness can require adapted or culturally grounded interventions for the subgroups, requiring small sample intervention research to test if the new intervention is effective by subgroup. This is typically the case for subpopulations, for example, for an American Indian, Alaska Native, and Native Hawaiian group, or for a rural Scotch-Irish Appalachian population.

A further challenge involves contextual variables such as toxic exposures from a single factory or busy roadway or neighborhood characteristics such as lead levels in homes or neighborhood poverty. Because such exposures may occur over small areas, they may either define a small population of interest in itself (e.g., residents near a specific factory or mine) or they may result in confounding or other statistical issues for studies of small populations.

However small populations are defined, intervention studies with these groups will likely necessitate small sample research. Study designs are often underpowered due to their sample sizes. As noted by Fok et al. (2015),

> “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.”

The goals of this work shop include clarifying the overall taxonomy of small population challenges and articulating opportunities and gaps in efforts to address them.

## References

Devers, K., Gray, B., Ramos, C., Shah, A., Blavin, F., and Waidmann, T. (2013). The Feasibility of Using Electronic Health Records (EHRs) and other Electronic Health Data for Research on Small Populations”, A report by the Urban Institute, delivered to the Office of the Assistant Secretary for Planning and Evaluation at HHS. https://www.urban.org/research/publication/feasibility-using-electronic-health-data-research-small-populations.

Fok, C.C., Henry, D. and Allen, J. (2015). “Maybe small is too small a term: Introduction to advancing small sample prevention science.” Prevention Sciences. Oct; 16 (7):943-949. Springer.

Institute of Medicine (2001), “Small Clinical Trials: Issues and Challenges. Committee on Strategies for Small Number Participant Clinical Research Trials. Evans, C. H. and Ildstad, S. T. (editors). National Academy Press.

Institute of Medicine (2010), “Bridging the Evidence Gap in Obesity Prevention: A Framework to Inform Decision Making.” Washington D.C. The National Academy Press.

Tourangeau, R., Edwards, B., Johnson, T.P., Wolter, K.M, and Bates, N. (editors) (2014). “Hard-to-Survey Populations.” Cambridge University Press.

Srinivasan, S., Moser, R. Willis, G., Riley, W., Alexander, M., Berrigan, D., and Kobrin S. (2015).”Small is Essential: Importance of Subpopulation Research in Cancer Control.” Editorial, American Journal of Public Health, April 23, 2015.