

A Mixed Method Approach to Sample Design

Most users of social survey data give only a cursory look at the particulars of sample design, if they look at all. These details can be important, however, especially for researchers interested in community and neighborhood effects. Literally hundreds of such studies have been published, on a wide variety of topics, but it is relatively rare to find a critical discussion of neighborhood or community context in relation to sample design or units of measure. Ideally, theoretical and operational definitions would mesh, but this is seldom the case. This paper proposes a mixed method approach to sample design for personal interview surveys to insure that the boundaries of sampling units “make sense” in their local context, drawing on recent experience with the National Children’s Study (NCS).

The Children’s Health Act of 2000 authorized NICHD “to conduct a longitudinal study of environmental influences (including physical, chemical, biological, and psychosocial) on children’s health and development.” A key debate in the subsequent design of the study revolved around the sample, specifically whether it would be based on a representative sample of births in the general population, or whether participating children would be recruited through medical clinics (as done, e.g., in the Women’s Health Initiative). There were good arguments on both sides, but finally in 2004, based on the recommendation of an expert panel constituted for this purpose, Duane Alexander (NICHD Director) announced that the sample would be representative. This feature of the NCS design was praised in a recent review of the study undertaken by the National Academy of Sciences.

Accordingly, a stratified multistage cluster sample was designed. The first stage of sample selection involved 105 counties. Seven were designated to be “vanguards” for the study; data collection in the vanguards is slated to begin in early 2009. The remaining counties were divided into three waves, to begin in 2010, 2011, and 2012 respectively. Based on a competitive procurement process, contracts have been awarded to research groups around the country to collaborate in the development of the study and to implement the data collection in specific study locations. UNC (Entwisle, PI) currently holds contracts for the Duplin County Vanguard Center and also for the Rockingham County and Burke County study locations, both of which are in Wave I. We draw on our experience in these counties to illustrate community involvement in sample design.

Each PSU (county) will contribute 1000 births to the NCS cohort over a period of four years. Within PSUs, secondary or tertiary sampling units called sample segments will be selected that collectively yield 250 births per year. The design of the segments is the joint responsibility of the coordinating center (Westat) and the NCS study centers. The specific procedures vary somewhat between larger and smaller PSU’s and have evolved from the vanguard to the Wave I sites. Nevertheless, the goal has been and continues to be the selection of

10 to 15 segments from a population of segments that are mutually exclusive, conterminous units yielding equal numbers of births.

The proposed paper will report on our experience in the design of NCS sample segments in Duplin County, Rockingham County, and Burke County, North Carolina. Discussions with long-term residents were held in June 2006 (Duplin), January 2008 (Rockingham), and November 2008 (Burke). The proposed paper will describe procedures as we adapted to changes introduced by the coordinating center between the vanguards and wave I study locations and as we adapted to local conditions as we moved from one county to the next. We will describe the acquisition, geocoding, and spatial analysis of birth certificates, the preparation of draft maps, the selection of long-term residents for group discussions, the particular issues raised, and the development of the segments.

Scientifically and operationally, it is important that segment boundaries correspond to meaningful neighborhood and community boundaries. One of the guiding hypotheses of the NCS is that neighborhoods and communities impact child health. Studies often use census tracts as proxies for neighborhoods, arguing that at least historically in metropolitan areas of the Northeast and midWest, these were developed in relation to sociospatial units that held meaning for local residents. The validity of such arguments elsewhere is an open question, perhaps especially in nonmetropolitan areas. The proposed paper will address this issue directly by comparing the segments developed using community input to those developed without this local knowledge, and to other units such as census tracts and block groups.

Boundaries need to make sense for operational reasons as well. Only a subset of births in the PSU's will be eligible for inclusion in the NCS cohort. Field staff will need to explain to residents why some are recruited and not others, and while ultimately the explanation involves principles of probability sampling, it makes more sense locally if sampling unit boundaries resonate with local understandings of community. Between a third and a half of the births occurring in Duplin, Rockingham, and Burke Counties will be eligible for the NCS, so the study will be very visible. Further, although most NCS participants will be recruited through household enumeration and follow-up, some will occur through prenatal clinics and hospitals. Segment definitions that make sense "on the ground" will be important to the identification of eligible births when mothers are recruited through these alternate venues. Given the complexity of the study and its longitudinal design, community engagement is a crucial component.