# Executive Summary: Characterizing Anticipated Mobility of the NCS Cohort

by

Jeff Lehman, Jyothi Nagaraja, and Warren Strauss

For the U.S. population, the Census Bureau estimates that roughly 46 percent of all persons aged five years or older have moved at least once between 1995 and 2000, with just over half of these movers (54 percent) relocating within their original county of residence and the remaining movers (46 percent) relocating to a different county or state. Considering these estimates of population mobility, it is reasonable to assume that over the course of the NCS a significant percentage of the initial participants can be expected to relocate. Thus, the impact of this cohort mobility on the financial, logistical, and scientific objectives of the study must be considered. For example, there may be increased costs associated with data collection and tracking of participants that move a significant distance away from their original geographic area, and there may be implementation difficulties, such as standardization of measures, associated with the data collection activities for these mobile study subjects (in particular for those that move to a new location that is not in close proximity to any of the NCS data collection centers). Alternatively, if tracking and collecting data on study subjects that relocate is very difficult, or is deemed too costly in relation to the information gained, it may be preferable in certain cases to employ alternative data collection methods that would help to mitigate these additional costs (e.g., telephone interview in place of an in-person interview, mailing samplers rather than technician visits, etc.), or in some extreme cases to simply allow the subject to drop from the study. In this case, there would likely be scientific impacts in that mobility may increase study subject attrition or decrease the amount of information available resulting in smaller sample sizes and/or less accurate data for statistical analyses of interest in the NCS.

In this report, data sources describing county-to-county flows of the US population are used to construct transition matrices that allow us to project the amount of subject mobility for the NCS. This projection is applied to designs that initiate with a finite number of geographic regions (50, 100 and 250 clusters/counties) in order to determine the extent to which designs with a larger number of initial clusters help to minimize the negative impact of population mobility. More particularly, given a candidate design we estimate the number of individuals that move away from the original counties represented in this design and/or move to counties that are a significant distance (e.g., greater than 25, 50, or 100 miles away) from these original counties over the course of the study. Additionally, we estimate the number of counties, and the number of people living in these counties, that are a significant distance from the original counties and that have a relatively large number of study subjects. These types of summary statistics can be used to evaluate the expected number of individuals for which it may be necessary to implement alternative data collection procedures, and to compare, in terms of mobility, different approaches to selecting the geographically clustered design.

Table 1 displays a portion of these summary statistics for a 50, 100, and 250-county design in which the counties were selected with probability proportional to size sampling from eight strata consisting of four regions of the United States and rural versus urban counties. To provide

expected mobility rates for this type of probabilistic sampling, the results displayed in the table are average mobility rates for 50 realizations of each type of design. The results indicate that a significant portion of the cohort will move away from the initial set of counties selected in the NCS. For example, at Year 5 of the study approximately 16,500 study subjects will move away from the initial set of counties in a 50-county design, approximately 15,000 subjects will move away from the initial set of counties in a 100-county design, and approximately 11,500 subjects will move away from the initial set of counties in a 250-county design. By increasing the catchment area of the original set of counties (e.g., regions consisting of all counties within 100 miles of an original county), the impact of mobility, while still significant is much less severe.

**Table 1. Impact of mobility for geographically clustered designs.**

| # of Original Counties | Year | # Subjects Living in Original Counties | # of Subjects in Counties w/in 25 Miles of Original Counties | # of Subjects Residing in Counties >25 Miles from Original Counties | | | # of Subjects Residing in Counties >100 Miles from Original Counties | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | All Counties [a] | Counties with >10 Subjects (# Counties)[b] | Counties with >25 Subjects (# Counties) [b] | All Counties[a] | Counties with >10 Subjects (# Counties) [b] | Counties with >25 Subjects (# Counties) [b] |
| 51 | 1 | 100,000 | 0 | 0 | 0 (0) | 0 (0) | 0 | 0 (0) | 0 (0) |
| | 5 | 83,445 | 3,349 | 13,206 | 9,030 (316) | 5,825 (114) | 4,590 | 2,607 (103) | 1,494 (32) |
| | 10 | 70,353 | 5,682 | 23,965 | 18,776 (519) | 14,319 (235) | 8,520 | 5,803 (183) | 4,115 (74) |
| | 15 | 59,927 | 7,282 | 32,791 | 27,051 (666) | 21,503 (313) | 11,912 | 8,719 (246) | 6,495 (103) |
| | 20 | 51,567 | 8,356 | 40,077 | 33,929 (776) | 27,625 (377) | 14,857 | 11,299 (297) | 8,677 (129) |
| 100 | 1 | 100,000 | 0 | 0 | 0 (0) | 0 (0) | 0 | 0 (0) | 0 (0) |
| | 5 | 85,056 | 3,705 | 11,239 | 6,954 (280) | 3,986 (88) | 2,493 | 1,177 (54) | 559 (13) |
| | 10 | 73,250 | 6,363 | 20,387 | 15,024 (478) | 10,576 (195) | 4,609 | 2,775 (102) | 1,744 (36) |
| | 15 | 63,856 | 8,257 | 27,887 | 21,951 (621) | 16,485 (273) | 6,421 | 4,257 (140) | 2,901 (53) |
| | 20 | 56,330 | 9,594 | 34,076 | 27,788 (733) | 21,496 (333) | 7,982 | 5,582 (172) | 3,966 (68) |
| 250 | 1 | 100,000 | 0 | 0 | 0 (0) | 0 (0) | 0 | 0 (0) | 0 (0) |
| | 5 | 88,500 | 3,713 | 7,787 | 3,588 (183) | 1,433 (40) | 835 | 252 (14) | 79 (2) |
| | 10 | 79,457 | 6,492 | 14,051 | 8,754 (365) | 4,888 (113) | 1,509 | 709 (34) | 321 (8) |
| | 15 | 72,292 | 8,572 | 19,137 | 13,303 (498) | 8,251 (172) | 2,064 | 1,126 (48) | 578 (13) |
| | 20 | 66,575 | 10,129 | 23,296 | 17,149 (600) | 11,204 (220) | 2,525 | 1,500 (60) | 859 (19) |

[a] # of subjects residing in the indicated counties (e.g., counties greater than 25 miles from original counties).
[b] # of subjects residing in counties that are "far" from the original counties and that contain a large number of study subjects.

Based on these results, there will clearly need to be procedures and tracking mechanisms to maintain contact with and to collect data from subjects that move (in particular for subjects that move away from the region in which they were originally recruited). For example, one reasonable mechanism may be to refer subjects that move away from one data collection center but within the catchment area of an already established data collection center to those centers for the necessary data collection. Alternatively, all study subjects that move into regions that are not covered by an established collection center might be referred to a single organization capable of following these individuals across the nation (or a mobile data collection center may be useful for particular regions that coincide with a significant portion of the cohort population). Tracking of study subjects, on the other hand, may include frequent contact with study subjects via telephone or postcard mailings, and/or other tracking procedures, such as searching for individuals via public databases. Of course, as with any study component, the costs associated with tracking and collecting data from participants who moved should be weighed against the overall benefits of continuing to follow that subject, the investment already incurred by the NCS for that family, and the timing of when the child leaves the study because of a move.

In terms of comparing designs that initiate with different numbers of selected counties, these results demonstrate that designs with more regions will have greater coverage of the mobile cohort over time. Thus, if cohort mobility were the only consideration for choosing a design,

then designs with fewer geographic regions would be less optimal. Taking into consideration factors such as the financial cost of tracking and collecting data from study subjects that move as well as the costs of starting and maintaining data collection operations in each of the initially selected counties, the optimal design becomes less apparent. For example, suppose that an established data collection center would be able to track and collect data from those study subjects that resided within a 100-mile radius, and suppose that we have a nominal goal of covering 95 percent of the original cohort with local or mobile data collection centers (such as the trailers used in NHANES). For a 250-County design there would be no need for mobile data collection centers at any time over the course of the study since the portion of the population that resides in counties that are greater than 100 miles away from the original counties is always less than 5,000 individuals. On the other hand, for the 50- and 100-County designs it would be necessary to establish mobile data collection centers over the course of the study in order to maintain coverage of at least 95,000 study subjects. In particular, for the 100-County design mobile data collection centers would be needed beginning prior to year 15 of the study, and for the 50-County design, mobile data collection centers would be needed prior to year 10 of the study. Assuming that the mobile units can cover a number of regions (e.g., 10 to 20) in a year, it would appear that for both the 50 and 100-County designs a large portion of the cohort can be covered with the data collection centers in the original counties plus a small number of additional mobile data collection centers (e.g., judging by the number of counties that are greater than 100 miles from the original counties but have a relatively large number of study subjects). Additionally, it is clear that much of the work for the mobile collection centers would only be necessary in the later years of the study.

This simple example suggests that designs initiating with on the order of 50 to 100 counties, each having a data collection center, and adding mobile data collection centers over time in order to maintain coverage of at least 95 percent of the cohort population, may be more cost-efficient than a design that initiates with 250 counties since they would likely require a much smaller number of data collection centers over the course of the study. A useful follow-on analysis to these results would be to provide an accurate assessment of the number of regions that would need to be covered by mobile centers in order for the 50- and 100-county designs to maintain coverage of at least 95 percent (or some other significant proportion) of the cohort. Presumably, by strategically locating these mobile centers, coverage of the desired portion of the cohort population could be accomplished in an efficient manner. In any case, based on these analyses there does not appear to be a strong argument that a 250-county design would be more optimal in terms of cohort mobility. In fact, it appears that a 50- or 100-county design, with additional mobile collection centers added over the course of the study, would be more cost-efficient.

Finally, it should be noted that the methods utilized in projecting the geographic dispersal of a candidate NCS cohort may be most useful if applied once the actual NCS cohort (or the geographic regions of the NCS cohort) has been selected. This would allow a more accurate portrayal of the expected cohort mobility for the selected design, and provide study leaders important information that could be useful in planning where, when, and to what degree the cohort will disperse across the United States.