EPA 600/R-05/017

# Interim Report
# Develop a Sampling Strategy: Prepare to Implement a Cohort Study of Children's Environmental Health

***Prepared for:***
**Pauline Mendola, Ph.D.**
**U.S. Environmental Protection Agency**
**National Health and Environmental Effects Research Laboratory**

**Prepared by RTI International under contract 68-D-02-069:**
**Sharon Myers, MPH**

**RTI Work Assignment Leader: Sharon Myers, MPH**
RTI International
3040 Cornwallis Road
RTP, NC 27709
Telephone: (919) 541-7316
Fax: (919) 541-7250
E-mail: sharonm@rti.org

**Disclaimer:**

The information in this document has been funded wholly or in part by the U.S. Environmental Protection Agency under contract 68-D-02-069 to RTI International. It has been subjected to review by the National Health and Environmental Effects Research Laboratory and approved for publication. Approval does not signify that the contents reflect the views of the Agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

EPA 600/R-05/017

**Contract No. 68-D-02-069** **August 25, 2004**

# Deliverable: Interim Report for Task 2, Develop a Sampling Strategy

# Prepare to Implement a Cohort Study of Children's Environmental Health

Work Assignment Number: 02-10

RTI Project No. 08601.001.010

***Prepared for:***

**EPA Work Assignment Manager: Pauline Mendola, Ph.D.**

Attn: Mr. James Kempf
**U.S. Environmental Protection Agency**
**RTP-Procurement Operations Division (D143-01)**
**4930 Old Page Road**
**Durham, NC 27709**

**Prepared by:**

**RTI Work Assignment Leader: Sharon Myers, M.P.H.**
RTI International
3040 Cornwallis Road
Research Triangle Park, NC 27709
Telephone: (919) 541-7316
Fax: (919) 541-7250
E-mail: sharonm@rti.org

**RTI Contracting Officer: Kelly Koeller-Anna**
**Office of Research Contracts**
**RTI International**
**P.O. Box 12194**
**Research Triangle Park, NC 27709-2194**
**Telephone: (919) 316-3746**
**Fax: (919) 541-6624**
**E-mail: kellyk@rti.org**

RTI INTERNATIONAL

## 1.0 Overview

The goal of the National Children's Study (NCS) is to collect information on environmental and social factors in the lives of children, starting before birth and continuing until age 21, to inform researchers about their effects on the health and development of children. In this document, we propose sample design options to be considered for implementation in a North Carolina-based longitudinal cohort study (LCS). The purpose of the report is to present "within-primary sampling unit (PSU)" procedures for sampling, recruiting and collecting data for that 5,000-children cohort. EPA proposed in the work assignment that initiated this work that RTI consider the sampling and design approaches currently under consideration for the NCS, and develop sampling scheme options for the North Carolina-based pilot study. Considerable work has already been done for the NCS by Battelle (Strauss et al, 2004) and Westat (Westat, 2002) in developing sampling scheme options. This report builds on the sample design options already proposed and focuses on the possible implementation of the sample design schemes in North Carolina counties.

The first-stage sample design for the NCS involves the selection of PSUs in the U.S. in which the study will be conducted. The goal of the within-PSU sample selection is to obtain approximately 25% of the children for the study from women who are at a pre-conception stage and the remaining 75% of the children from women who are pregnant at recruitment, ideally in their first trimester. Two sample designs are evaluated as possible options for the within-PSU sample selection using a probability-based sampling scheme. The first option is based entirely on an area household sampling frame, as requested by the WAM. The second option is a dual frame design that has the potential to provide equivalent inferential support at reduced cost by incorporating area household sampling as well as sampling in the offices of prenatal care providers. The dual frame design was initially proposed in the Work Plan as RTI's main sampling strategy, and this approach had already been developed when we received feedback from the WAM. Thus, it is included here for consideration.

## 2.0 Selection of PSUs for LCS

Because the LCS is a pilot test of the within-PSU sampling, recruiting and data collection procedures, the PSUs to be used in the LCS may be purposively selected counties to ensure a range of population characteristics and study implementation situations. Some of the factors to take into consideration in selecting the PSUs are:

- Urban/Rural status
- Proximity to Research Triangle Park (RTP)
- Socioeconomic status
- Hispanic population.

Choosing PSUs with different urbanicity, ethnic and socioeconomic characteristics will provide an opportunity to test the study procedures in various settings that may be encountered in the areas selected for the NCS. Because RTI is located in RTP and has strong connections to other local research institutions, it will probably be more efficient for us to gain cooperation from residents and obstetricians in the RTP area as opposed to the counties in the eastern and western portions of NC which may not be as familiar with RTI. PSUs may be purposively chosen for the LCS and will cover the range of characteristics specified above in order to test the study implementation procedures in situations that may be expected to occur in the NCS.

If the goal of the LCS is simply to test the study implementation procedures, purposive selection of the PSUs will be the preferred selection method. However, if one of the goals of the LCS is to produce state-level estimates for North Carolina, it is recommended that a sample of 30 PSUs be selected using a probability proportional to size (PPS) sampling scheme. The list of NC counties could be stratified by some of the above mentioned characteristics in order to test the recruitment and data collection procedures in areas with different characteristics.

The number of PSUs to choose for the LCS will be determined partly by the length of the recruitment period. We understand the NCS plans to recruit women over a period lasting between 3 and 5 years. Since this is a pilot study, we recommend a shorter recruitment period in order to test the procedures prior to the implementation of the full-scale NCS. For the purposes of providing sample size estimates, we have assumed a 2-year recruitment period. Depending on the number of PSUs chosen in which to implement the LCS, some NC counties, especially those in rural areas, may not have sufficiently large populations to support the sample requirements. When county-level statistics on births are below the sizes needed within the PSUs, adjacent counties will be combined to form PSUs. When combining counties to form PSUs, it is desirable to combine counties that are as homogeneous as possible with respect to urbanicity, sociodemographic and socioeconomic characteristics and total population size. County level sociodemographic and socioeconomic statistics from Census 2000 data are provided in ***Exhibit 1***, along with birth and pregnancy estimates from data provided in the National Vital Statistics Reports (Martin et al, 2003). As shown in ***Exhibit 2***, some of the counties in NC are geographically clustered with respect to racial, ethnic and socioeconomic characteristics. These clusters of adjacent counties will aide in the task of combining relatively homogeneous counties to form PSUs.

### 3.1 Sample Design Option 1: Area Household Sample

The area household sampling approach can be used to select a probability-based sample of pregnant women in the selected PSUs, along with another probability-based sample of women of child-bearing age who are not pregnant. According to the National Vital Statistics Reports (Martin et al, 2003),

over 97% of live births occurred to mothers between the ages of 15 and 39. As a result, we will define women of child-bearing age as women who are between 15 and 39 years old.

Each selected PSU will be subdivided into area segments so that recruitment and data collection efforts can be restricted to small areas as opposed to being spread out over the whole county or combination of counties. The area segments can be defined by either Census block groups or postal carrier routes (sub-5-digit zip code areas). Population sociodemographic estimates for each Census block group are readily available through the 2000 Census data products. Population sociodemographic estimates for the postal carrier routes are available from commercial marketing businesses. The choice of which type of area segments to use will depend on the preferred method of constructing the area segment sampling frame. Once the area segments are defined, a Probability-Proportional-to-Size (PPS) sample of area segments will be selected based on the estimated number of births in the area segment, along with any sociodemographic characteristics which are to be oversampled.

The sampling frame for each selected area segment within each selected PSU will consist of a list of all residential housing units in the selected area segment. Construction of a list of all residential housing units can be accomplished through either counting-and-listing or commercial address lists. Counting-and-listing involves sending study staff to the selected area segments to record all residential housing units. Commercial address lists are available from marketing businesses which purchase updated postal delivery files directly from the U.S. Postal Service and are able to distinguish between residential, business and post office box addresses. The counting-and-listing method is considered to be the most accurate in terms of coverage of the specified area segments, but requires a lot of resources in the form of field staff labor hours. Commercial address lists have been used in other area household sample studies (Iannacchione et al, 2003) in which the address lists have been found to provide coverage of at least 97.5% of all households in the metropolitan study area. The main disadvantage in the use of commercial address lists is that they cannot identify the physical location of households in rural areas or college campuses which use postal mail boxes instead of local postal carrier delivery to the individual residences. RTI staff have developed methods for the National Survey of Drug Use and Health on how to perform counting-and-listing methods for group quarters (e.g. college dormitories), civilian residences on military bases, and gated communities.

Some marketing companies provide mailing lists which target households with certain characteristics, such as age of residents, languages spoken in the home and household income. These targeted mailing lists could be merged with the postal service lists to create strata containing households that are more likely to contain women of child-bearing age. Counting-and-listing methods can also be performed on subsamples of the selected households in the area segments as a means of assessing the

undercoverage of the commercial address lists in order to make appropriate adjustments to the final analysis weights.

If the counting-and-listing method is chosen as the preferred method for constructing the area segment household sampling frames, Census block groups would be the preferred classification of area segments since block groups are well-defined and about the right size, and reliable population estimates are easily obtained from the Census Bureau. If the commercial address list method is chosen for constructing the area segment household sampling frames, postal carrier routes would be the preferred classification of area segments since postal carrier routes are similar in size to Census block groups and the carrier route designations are easily available on the commercial address lists. If the commercial address list method is chosen for the sampling frame construction but Census block groups are the preferred classification method for the area segments, the addresses on the list can be geocoded to assign each household to its corresponding Census block group. The disadvantage to relying on geocoding addresses to determine the Census block group to which the addresses belong is that newer addresses, i.e. addresses located in new subdivisions, may not appear in the most recent GIS databases and the location of those newer addresses would have to be estimated based on the zip code +4 values, thereby reducing the coverage of the sampling frame within the Census block group.

Once the sampling frames of the households within each selected area segment have been constructed, a systematic random sample of households will be chosen for screening. Study staff will attempt to conduct a short screening interview at each selected household to determine the number of child-bearing age women and the number of pregnant women in the household. The screening interview results will then be used to stratify the households to allow for oversampling of certain characteristics or to improve the efficiency of the sample by allocating higher probabilities of selection to households in which there is a higher chance of at least one of the female residents becoming pregnant. Pregnant women identified by the household screening interview will be selected for the study, but subsampling may be used to give priority to recruiting pregnant women who are in the first trimester of their pregnancy. A portion of the women of child-bearing age who do not report currently being pregnant at the time of the household screener interview will be selected for the study to obtain pre-conception measures in the hopes that some of these women will become pregnant during the study recruitment period. We propose developing a model to predict the probability a woman will become pregnant during the recruitment period and oversampling those women who are estimated to have a higher chance of becoming pregnant based on the results of the household screening interviews.

The non-pregnant women selected for the study will be provided with menstrual diaries and pregnancy test kits to be updated by the selected women on a monthly basis. The women will be asked to call the study staff if they discover that they are pregnant. Periodic telephone follow-ups will be

conducted with the non-pregnant sampled women to check on their pregnancy status and update any changes to their contact information.

### 3.2 Sample Design Option 2: Multiframe Sample Design

Area household sampling is necessary for complete population coverage, but sampling at prenatal care providers may be able to reduce total costs without sacrificing probability-based sampling methods. The use of prenatal care providers to recruit patients presents two major difficulties: (1) getting prenatal care practices to participate and (2) getting participating practices to implement the recruitment protocols with sufficient diligence that probabilities of selection can be calculated and the validity of the sample can be validated. The area household sample would consist of a probability-based sample of households in the PSU in order to screen and recruit women of child-bearing age who are not pregnant at the time of screening, along with recruiting women who are pregnant and have not yet been recruited through a prenatal care provider. The prenatal care provider sample would consist of a sample of women residing in the selected PSUs who present at the prenatal care provider during their first trimester.

As an alternative to the large number of households which will need to be screened in order to recruit pregnant women, we propose that the NCS planners consider an alternate sample design approach that combines recruitment through prenatal care providers and recruitment through an area household sample. To gain the cooperation of prenatal care providers in the selected PSUs, we propose a top-down approach in which hospitals where births occur are recruited first, and then the prenatal care providers who are considered members of the hospital staff will be recruited. Due to fears of malpractice suits, in general, only persons certified to deliver babies are willing to provide prenatal care. According to a National Vital Statistics Report for 2002 (Martin et al, 2003), 99.1% of live births in the U.S. occurred in a hospital and 92.1% of hospital births were attended by a physician. Because of the vast majority of births occurring in hospitals and the direct links between hospital staff members, birth attendants and prenatal care providers, the recruitment of prenatal care providers through their associated hospitals will be the most efficient method. Based on previous studies, physicians can be difficult to recruit for participation in research studies. We hypothesize that by obtaining the endorsement of the associated hospital, along with the endorsement of relevant medical associations, physicians may be more likely to agree to participate in the LCS.

In preparation for recruiting prenatal care providers to assist with the sample selection, a list of the number of births in hospitals serving women who reside in the selected PSUs will be obtained from the North Carolina State Center for Health Statistics for the most recently available year of vital statistics data. The hospitals identified as birth locations for the female residents of the selected PSUs will be recruited for the study. Any stand alone birth clinics identified as locations for a sufficient number of

births in the selected PSUs may also be approached for participation in the study. After gaining the approval of the hospital administrators, the hospitals will be asked to provide a list of their staff members who are approved birth attendants. Some obstetrics practices and hospitals retain midwives on their staff, so this hospital-based recruitment method of prenatal care providers will also include some midwife-attended births. The 2002 National Vital Statistics Report (Martin et al, 2003) shows that 93.2% of births attended by a midwife occurred in a hospital, while only 4.0% of midwife attended births occurred at a residence and 2.6% occurred at a freestanding birthing center. Even though women who choose to give birth at home with the assistance of a midwife or who do not pursue prenatal care will not be recruited for the study through the prenatal care practices, women in these categories will have a chance to enter the study through the area household sample.

In the participating physicians' offices, all women presenting during the first trimester of the pregnancy will be selected for the study. If desired, women who present for their first prenatal visit in the later stages of the pregnancy will also be selected. Selected women who choose to participate in the study will be asked to fill out a contact information card and sign an informed consent allowing study staff to contact the women directly. Due to HIPAA privacy restrictions, the study staff will not be allowed to obtain any identifying information on a prenatal provider's patient without the patient's direct consent. Methods will need to be developed to allow physicians' office staff to quickly send the contact information to the study staff for immediate follow-up in order to try to get the initial set of measurements while the women are still in their first trimester. Since the first prenatal visit usually occurs after at least 8 weeks gestation, there will not be a lot of time to obtain the environmental samples prior to the end of the first trimester (13 weeks gestation). The recruitment of pregnant women through the prenatal care providers will rely on the providers' staff to accurately adhere to the study sample selection and record-keeping protocols, in addition to performing their regular duties in the providers' offices. Because the staff in a prenatal care provider's office may already be overworked with respect to their regular office duties, any sampling and recruitment methods implemented would need to be the least burdensome as possible on the office staff. One option to reduce the burden on prenatal provider office staff is to employ the use of a study representative to assist with the study recruitment. However, based on focus groups conducted with obstetrics medical providers in the RTP area, office staff would prefer not to have an outside person in the office for purposes of study recruitment (Dimitropolous, 2004). The calculation of the probabilities of selection and analysis weights after data collection would rely heavily on the providers' staff keeping accurate records on a daily basis of the number of pregnant women selected and the number of selected women who choose to participate in the study. Over a 2-year recruitment period, there would most likely be a number of occasions when the providers' staff may neglect to give the study

materials to an eligible patient, or neglect to accurately record the total number of women who were selected or who chose to participate.

In order to remove bias associated with the exclusion of certain types of pregnant women by recruiting only through a prenatal care provider, a supplemental area household sample also is essential. The use of two sampling frames to select samples of women for the study will require the use of multiframe estimation methods to appropriately calculate the sample probabilities and analysis weights. Lessler and Kalsbeek (1992), Lessler (1981) and Casady and Sirkin (1980) present methods for the use of data collected from studies using more than one sampling frame.

## 4.0 Validation Samples

If subsamples are needed to validate low burden or inexpensive (surrogate) measures versus the higher burden or more costly (true) measures, the size of the validation subsample will depend on the strength of the correlation between the surrogate and true measures. If the correlation between the measures is high, the power of statistical tests based on the surrogate measure may be nearly as good as if the true measures were obtained for the entire sample. If the correlation between the measurement types is not high, very little additional power may be gained and the validation sample may not be cost effective. Subsamples of the LCS participants may be used to estimate the correlation between the surrogate and true measures in order to inform decisions on the use of validation measure in the full NCS study. The time and budget for this work assignment were not sufficient to investigate the utility of validation subsamples for the LCS.

## 5.0 Sample Size Requirements

Whether state-level estimates will be made from the collected data has a large effect on the number of PSUs to be chosen for the study, and as a result, the number of pregnant women needed to recruit for the study within each selected PSU in order to obtain a sample of 5,000 children. Exhibits 3 – 6 present sample size requirements for the area household sample and multiframe sample design options for the option of 8 purposively selected PSUs. Exhibits 7 – 10 present sample size requirements for the area household sample and multiframe sample design options for the option of 30 randomly selected PSUs.

## 6.0 Summary

This report has considered several sampling design options for selecting pregnant women and non-pregnant women of childbearing age (nominally, 15 to 39) for the NC longitudinal cohort study (LCS). The primary options we have considered are selection of the sample entirely through an area

household sampling design versus using a dual frame design that also incorporates probability-based samples of women presenting for prenatal care at their medical care providers' offices. In either case, all women of childbearing age in the sample households would be monitored for pregnancy during the 2 years of recruitment. All resulting pregnancies would be included in the LCS sample.

The number of sample PSUs to be selected for the LCS depends on the inferential requirements of the LCS. If inferences to the State of NC are not necessary, a relatively small purposively selected sample of PSUs will be sufficient (e.g., eight) for testing the NCS procedures in a variety of settings. However, if state-level statistical inferences are desired, a probability-based sample of a larger number of PSUs (e.g., 30) would be preferable.

The minimum PSU size is directly affected by the above choice; smaller PSUs being sufficient for recruiting 5,000 participating children with a larger sample of PSUs. The minimum PSU size also is affected by the length of the recruitment period. We suggest a relatively short recruitment period (e.g., 2 years) so that testing of NCS methodology can begin as soon as possible.

Two options for defining area sample segments within sample PSUs have been considered: Census block groups and postal carrier routes. Postal carrier routes can greatly reduce costs with little, if any, loss of population coverage. Moreover, postal addresses can be merged with commercial lists of households expected to contain women of childbearing age to produce a more efficient sampling stratum.

RTI looks forward to the opportunity to refine these options and proceed with sample selection in consultation with the U.S. EPA.

## 7.0 References

Casady, R.J. and M.G. Sirkin (1980). "A Multiplicity Estimator for Multiple Frame Sampling." Proceedings of the *American Statistical Association Section on Survey Research Methods*, pp. 601-605.

Dimitropolous, L (2004) "Final Report: National Children's Study Focus Groups." Project Report prepared for the U.S. Environmental Protection Agency, RTI project number 08601.000.001.007.

Iannacchione, V. G., J. M. Staab and D.T. Redden (2003) "Evaluating the use of residential mailing addresses in a metropolitan household survey." *Public Opinion Quarterly*. Volume 67 (2), 202-210.

Lessler, J.T. and W.D. Kalsbeek (1992). *Nonsampling Error in Surveys*. New York, NY: Wiley (Section 5.2)

Lessler, J.T. (1981). "Multiplicity Estimators with Multiple Counting Rules for Multistage Sample Surveys." Proceedings of the *American Statistical Association Section on Social Statistics*, pp. 12-16.

Martin, J.A., B.E. Hamilton, P.D. Sutton, S.J. Ventura, F. Menacker and M.L. Munson (2003). "Births: Final Data for 2002." National Vital Statistics Reports, vol. 52, no. 7. National Center for Health Statistics, Hyattsville, MD.

Strauss, W., J. Lehman, J. Menkedick, L. Ryan, T. Pivetz, N. McMillian, B. Pierce and S. Rust (2004). "White Paper on Evaluation of Sampling Design Options for the National Children's Study." Prepared by Battelle for the National Children's Study Program Office.

Westat (2002). "Sampling Strategies for the Proposed National Children's Study." Prepared for the National Center for Health Statistics.

**Exhibit 1. Racial, Ethnic, Socioeconomic and Size Characteristics of North Carolina Counties**

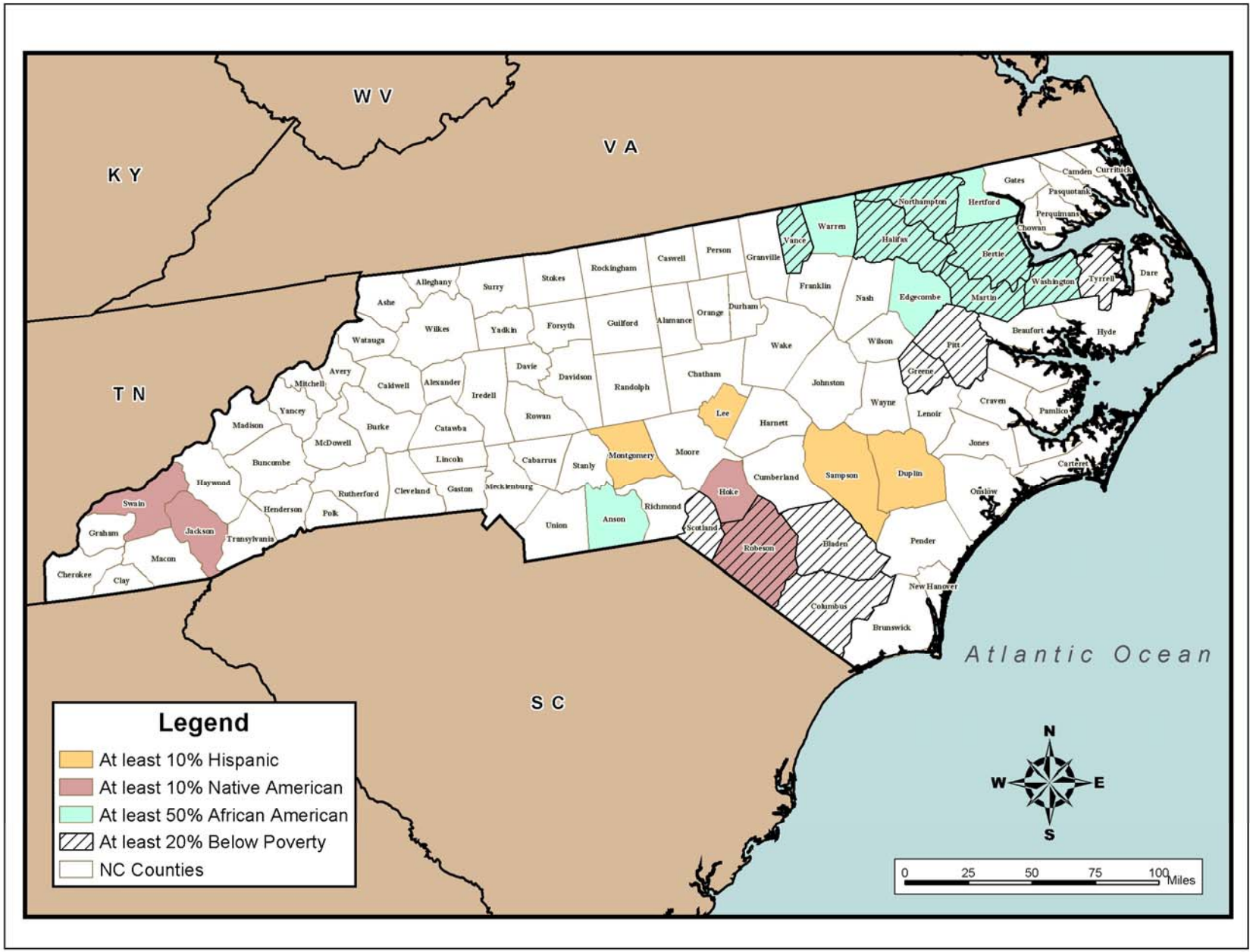
| County Name | # Births | # Pregnancies | Median Income | % Below Poverty | % Occupied Households | Total Households | % White | % Black | % Amer Indian | % Hispanic |
|---|---|---|---|---|---|---|---|---|---|---|
| Alamance County | 1,854 | 2,944 | 39,168 | 11.14 | 93.01 | 55,463 | 72.75 | 20.57 | 0.40 | 6.75 |
| Alexander County | 444 | 660 | 38,684 | 8.48 | 93.18 | 14,098 | 92.14 | 4.31 | 0.05 | 2.50 |
| Alleghany County | 113 | 166 | 29,244 | 17.15 | 71.63 | 6,412 | 95.96 | 0.89 | 0.14 | 4.96 |
| Anson County | 311 | 559 | 29,849 | 17.77 | 90.05 | 10,221 | 43.81 | 54.23 | 0.37 | 0.83 |
| Ashe County | 270 | 395 | 28,824 | 13.45 | 78.47 | 13,268 | 97.73 | 0.57 | 0.31 | 2.42 |
| Avery County | 201 | 295 | 30,627 | 15.26 | 54.84 | 11,911 | 97.69 | 0.47 | 0.23 | 2.41 |
| Beaufort County | 528 | 881 | 31,066 | 19.47 | 82.75 | 22,139 | 63.33 | 33.52 | 0.23 | 3.24 |
| Bertie County | 238 | 446 | 25,177 | 23.46 | 85.56 | 9,050 | 31.39 | 67.24 | 0.53 | 0.99 |
| Bladen County | 404 | 689 | 26,877 | 20.98 | 84.21 | 15,316 | 54.46 | 39.58 | 2.36 | 3.71 |
| Brunswick County | 798 | 1,254 | 35,888 | 12.58 | 59.18 | 51,431 | 78.15 | 17.95 | 0.99 | 2.68 |
| Buncombe County | 2,693 | 4,070 | 36,666 | 11.44 | 91.28 | 93,973 | 88.11 | 7.84 | 0.46 | 2.78 |
| Burke County | 1,111 | 1,667 | 35,629 | 10.68 | 92.25 | 37,427 | 86.33 | 5.62 | 0.37 | 3.57 |
| Cabarrus County | 1,819 | 2,823 | 46,140 | 7.06 | 93.70 | 52,848 | 80.74 | 13.97 | 0.43 | 5.05 |
| Caldwell County | 1,012 | 1,513 | 35,739 | 10.71 | 92.04 | 33,430 | 91.63 | 5.62 | 0.23 | 2.49 |
| Camden County | 80 | 125 | 39,493 | 10.06 | 89.54 | 2,973 | 80.04 | 17.36 | 0.46 | 0.71 |
| Carteret County | 633 | 958 | 38,344 | 10.71 | 61.55 | 40,947 | 88.17 | 8.21 | 0.58 | 1.74 |
| Caswell County | 269 | 451 | 35,018 | 14.40 | 90.30 | 9,601 | 61.81 | 36.00 | 0.14 | 1.77 |
| Catawba County | 1,925 | 2,934 | 40,536 | 9.08 | 92.68 | 59,919 | 83.05 | 9.26 | 0.31 | 5.57 |
| Chatham County | 616 | 968 | 42,851 | 9.72 | 92.43 | 21,358 | 72.31 | 17.20 | 0.52 | 9.62 |
| Cherokee County | 257 | 379 | 27,992 | 15.34 | 76.57 | 13,499 | 94.42 | 1.76 | 1.85 | 1.25 |
| Chowan County | 174 | 301 | 30,928 | 17.60 | 86.61 | 6,443 | 54.68 | 42.19 | 0.65 | 1.51 |
| Clay County | 81 | 119 | 31,397 | 11.44 | 70.91 | 5,425 | 97.15 | 1.38 | 0.55 | 0.83 |
| Cleveland County | 1,295 | 2,076 | 35,283 | 13.27 | 91.89 | 40,317 | 74.04 | 23.36 | 0.17 | 1.49 |
| Columbus County | 690 | 1,151 | 26,805 | 22.70 | 88.56 | 24,060 | 61.04 | 32.82 | 3.43 | 2.32 |
| Craven County | 1,212 | 1,978 | 35,966 | 13.08 | 90.65 | 38,150 | 66.71 | 27.21 | 0.50 | 4.02 |
| Cumberland County | 4,827 | 8,186 | 37,466 | 12.79 | 90.65 | 118,425 | 52.04 | 36.98 | 1.66 | 6.90 |
| Currituck County | 218 | 328 | 40,822 | 10.68 | 64.58 | 10,687 | 89.71 | 7.15 | 0.40 | 1.43 |
| Dare County | 339 | 502 | 42,411 | 8.03 | 47.58 | 26,671 | 93.66 | 3.05 | 0.39 | 2.22 |
| Davidson County | 1,949 | 2,975 | 38,640 | 10.07 | 93.15 | 62,432 | 84.92 | 10.40 | 0.49 | 3.24 |
| Davie County | 420 | 632 | 40,174 | 8.60 | 91.95 | 14,953 | 89.73 | 6.95 | 0.20 | 3.47 |

| County Name | # Births | # Preg-nancies | Median Income | % Below Poverty | % Occupied Households | Total Households | % White | % Black | % Amer Indian | % Hispanic |
|---|---|---|---|---|---|---|---|---|---|---|
| Duplin County | 642 | 1,062 | 29,890 | 19.44 | 89.02 | 20,520 | 55.09 | 29.89 | 0.22 | 15.14 |
| Durham County | 3,993 | 6,900 | 43,337 | 13.37 | 93.26 | 95,452 | 46.85 | 42.12 | 0.29 | 7.63 |
| Edgecombe County | 739 | 1,360 | 30,983 | 19.59 | 84.96 | 24,002 | 34.81 | 62.21 | 0.36 | 2.79 |
| Forsyth County | 4,459 | 7,369 | 42,097 | 11.05 | 93.06 | 133,093 | 62.90 | 29.84 | 0.31 | 6.40 |
| Franklin County | 638 | 1,051 | 38,968 | 12.59 | 87.62 | 20,364 | 65.05 | 30.49 | 0.55 | 4.44 |
| Gaston County | 2,636 | 4,105 | 39,482 | 10.86 | 93.78 | 78,842 | 80.76 | 15.74 | 0.33 | 3.00 |
| Gates County | 119 | 205 | 35,647 | 17.03 | 88.88 | 4,389 | 56.57 | 41.36 | 0.65 | 0.77 |
| Graham County | 88 | 129 | 26,645 | 19.51 | 65.97 | 5,084 | 89.77 | 0.17 | 8.61 | 0.75 |
| Granville County | 579 | 966 | 39,965 | 11.67 | 93.06 | 17,896 | 61.63 | 34.00 | 0.42 | 4.02 |
| Greene County | 232 | 404 | 32,074 | 20.23 | 90.88 | 7,368 | 47.44 | 44.18 | 0.20 | 7.96 |
| Guilford County | 6,575 | 11,145 | 42,618 | 10.59 | 93.50 | 180,391 | 57.62 | 35.66 | 0.50 | 3.80 |
| Halifax County | 702 | 1,278 | 26,459 | 23.90 | 87.41 | 25,309 | 36.81 | 57.61 | 3.76 | 1.01 |
| Harnett County | 1,395 | 2,236 | 35,105 | 14.90 | 87.55 | 38,605 | 70.36 | 23.01 | 0.79 | 5.86 |
| Haywood County | 610 | 895 | 33,922 | 11.51 | 80.66 | 28,640 | 96.12 | 1.31 | 0.59 | 1.41 |
| Henderson County | 958 | 1,425 | 38,109 | 9.72 | 87.02 | 42,996 | 90.10 | 3.83 | 0.41 | 5.47 |
| Hertford County | 274 | 513 | 26,422 | 18.33 | 92.07 | 9,724 | 31.11 | 66.17 | 0.97 | 1.57 |
| Hoke County | 522 | 885 | 33,230 | 17.68 | 90.85 | 12,518 | 44.65 | 36.45 | 11.72 | 7.18 |
| Hyde County | 56 | 93 | 28,444 | 15.44 | 66.17 | 3,302 | 66.19 | 30.67 | 0.13 | 2.25 |
| Iredell County | 1,617 | 2,521 | 41,920 | 8.16 | 91.22 | 51,918 | 79.59 | 15.63 | 0.35 | 3.41 |
| Jackson County | 509 | 758 | 32,552 | 15.12 | 68.38 | 19,291 | 84.85 | 2.48 | 10.12 | 1.74 |
| Johnston County | 1,776 | 2,770 | 40,872 | 12.81 | 92.83 | 50,196 | 77.42 | 16.03 | 0.49 | 7.74 |
| Jones County | 112 | 189 | 30,882 | 16.91 | 86.79 | 4,679 | 59.53 | 37.30 | 0.26 | 2.72 |
| Lee County | 640 | 1,031 | 38,900 | 12.84 | 92.75 | 19,909 | 66.28 | 22.14 | 0.52 | 11.65 |
| Lenoir County | 729 | 1,270 | 31,191 | 16.57 | 87.78 | 27,184 | 51.56 | 45.01 | 0.28 | 3.17 |
| Lincoln County | 844 | 1,270 | 41,421 | 9.17 | 93.48 | 25,717 | 89.39 | 6.92 | 0.24 | 5.73 |
| McDowell County | 532 | 790 | 32,396 | 11.60 | 90.35 | 18,377 | 92.99 | 3.12 | 0.31 | 2.88 |
| Macon County | 288 | 422 | 32,139 | 12.62 | 61.83 | 20,746 | 96.47 | 1.39 | 0.46 | 1.52 |
| Madison County | 255 | 374 | 30,985 | 15.39 | 82.29 | 9,722 | 97.89 | 0.56 | 0.25 | 1.35 |
| Martin County | 312 | 554 | 28,793 | 20.19 | 91.67 | 10,930 | 47.46 | 50.08 | 0.39 | 2.06 |
| Mecklenburg County | 11,329 | 18,790 | 50,579 | 9.20 | 93.39 | 292,780 | 59.87 | 31.60 | 0.37 | 6.45 |
| Mitchell County | 171 | 250 | 30,508 | 13.83 | 82.73 | 7,919 | 97.26 | 0.18 | 0.71 | 1.98 |

| County Name | # Births | # Preg-nancies | Median Income | % Below Poverty | % Occupied Households | Total Households | % White | % Black | % Amer Indian | % Hispanic |
|---|---|---|---|---|---|---|---|---|---|---|
| Montgomery County | 326 | 526 | 32,903 | 15.36 | 69.62 | 14,145 | 64.75 | 22.88 | 0.43 | 10.43 |
| Moore County | 811 | 1,284 | 41,240 | 11.37 | 87.37 | 35,151 | 74.49 | 19.24 | 0.99 | 3.99 |
| Nash County | 1,160 | 1,979 | 37,147 | 13.44 | 90.80 | 37,051 | 55.94 | 39.47 | 0.50 | 3.36 |
| New Hanover County | 2,482 | 3,880 | 40,172 | 13.05 | 85.64 | 79,616 | 79.41 | 16.91 | 0.47 | 2.04 |
| Northampton County | 238 | 440 | 26,652 | 21.26 | 83.13 | 10,455 | 34.60 | 64.14 | 0.31 | 0.73 |
| Onslow County | 2,398 | 3,802 | 33,756 | 12.91 | 86.35 | 55,726 | 70.88 | 18.95 | 0.81 | 7.25 |
| Orange County | 2,447 | 3,810 | 42,372 | 14.13 | 93.05 | 49,289 | 77.25 | 13.60 | 0.40 | 4.46 |
| Pamlico County | 121 | 195 | 34,084 | 15.33 | 76.36 | 6,781 | 71.20 | 26.20 | 0.84 | 1.32 |
| Pasquotank County | 489 | 864 | 30,444 | 18.40 | 90.33 | 14,289 | 49.98 | 47.13 | 0.29 | 1.23 |
| Pender County | 479 | 768 | 35,902 | 13.59 | 77.19 | 20,798 | 72.35 | 23.75 | 0.49 | 3.64 |
| Perquimans County | 121 | 198 | 29,538 | 17.94 | 76.87 | 6,043 | 68.25 | 30.76 | 0.18 | 0.60 |
| Person County | 447 | 734 | 37,159 | 11.99 | 90.85 | 15,504 | 66.85 | 30.14 | 0.53 | 2.09 |
| Pitt County | 2,535 | 4,217 | 32,868 | 20.34 | 89.95 | 58,408 | 62.90 | 32.50 | 0.36 | 3.15 |
| Polk County | 178 | 267 | 36,259 | 10.07 | 86.03 | 9,192 | 90.70 | 6.65 | 0.26 | 3.01 |
| Randolph County | 1,750 | 2,628 | 38,348 | 9.15 | 93.09 | 54,422 | 87.83 | 6.05 | 0.51 | 6.63 |
| Richmond County | 601 | 1,005 | 28,830 | 19.56 | 89.88 | 19,886 | 60.22 | 34.37 | 2.04 | 2.83 |
| Robeson County | 1,689 | 2,832 | 28,202 | 22.81 | 91.41 | 47,779 | 28.90 | 25.85 | 41.21 | 4.86 |
| Rockingham County | 1,162 | 1,852 | 33,784 | 12.79 | 91.99 | 40,208 | 74.73 | 21.76 | 0.35 | 3.07 |
| Rowan County | 1,728 | 2,730 | 37,494 | 10.59 | 92.52 | 53,980 | 76.95 | 18.16 | 0.35 | 4.12 |
| Rutherford County | 792 | 1,221 | 31,122 | 13.92 | 85.29 | 29,535 | 84.60 | 13.24 | 0.17 | 1.81 |
| Sampson County | 801 | 1,331 | 31,793 | 17.55 | 88.59 | 25,142 | 57.09 | 31.04 | 2.00 | 10.77 |
| Scotland County | 500 | 858 | 31,010 | 20.56 | 91.19 | 14,693 | 48.54 | 39.07 | 9.83 | 1.18 |
| Stanly County | 730 | 1,126 | 36,898 | 10.71 | 90.40 | 24,582 | 82.78 | 12.67 | 0.27 | 2.13 |
| Stokes County | 587 | 872 | 38,808 | 9.07 | 91.26 | 19,262 | 93.66 | 4.34 | 0.33 | 1.87 |
| Surry County | 897 | 1,336 | 33,046 | 12.42 | 91.54 | 31,033 | 89.56 | 4.09 | 0.28 | 6.49 |
| Swain County | 148 | 224 | 28,608 | 18.31 | 72.30 | 7,105 | 60.34 | 2.63 | 33.65 | 1.47 |
| Transylvania County | 309 | 463 | 38,587 | 9.48 | 79.21 | 15,553 | 91.42 | 5.33 | 0.42 | 1.02 |
| Tyrrell County | 41 | 70 | 25,684 | 23.32 | 75.64 | 2,032 | 55.69 | 38.00 | 0.00 | 3.62 |
| Union County | 1,721 | 2,660 | 50,638 | 8.14 | 94.96 | 45,695 | 81.25 | 13.54 | 0.50 | 6.17 |
| Vance County | 588 | 1,054 | 31,301 | 20.50 | 89.03 | 18,196 | 42.13 | 53.75 | 0.35 | 4.56 |
| Wake County | 10,258 | 16,462 | 54,988 | 7.82 | 93.47 | 258,953 | 69.16 | 22.31 | 0.38 | 5.41 |

| County Name | # Births | # Preg-nancies | Median Income | % Below Poverty | % Occupied Households | Total Households | % White | % Black | % Amer Indian | % Hispanic |
|---|---|---|---|---|---|---|---|---|---|---|
| Warren County | 206 | 377 | 28,351 | 19.45 | 73.08 | 10,548 | 33.56 | 59.10 | 5.63 | 1.59 |
| Washington County | 158 | 286 | 28,865 | 21.76 | 86.93 | 6,174 | 40.38 | 56.25 | 0.05 | 2.27 |
| Watauga County | 832 | 1,227 | 32,611 | 17.91 | 71.43 | 23,155 | 95.86 | 2.06 | 0.29 | 1.46 |
| Wayne County | 1,547 | 2,591 | 33,942 | 13.84 | 90.06 | 47,313 | 58.96 | 34.23 | 0.42 | 4.94 |
| Wilkes County | 828 | 1,229 | 34,258 | 11.85 | 91.08 | 29,261 | 92.83 | 3.94 | 0.12 | 3.45 |
| Wilson County | 990 | 1,721 | 33,116 | 18.47 | 93.11 | 30,729 | 50.11 | 44.65 | 0.31 | 6.04 |
| Yadkin County | 467 | 692 | 36,660 | 9.97 | 91.68 | 15,821 | 91.87 | 3.43 | 0.15 | 6.48 |
| Yancey County | 196 | 287 | 29,674 | 15.76 | 76.80 | 9,729 | 98.04 | 0.55 | 0.39 | 2.69 |

**Exhibit 2. Map of North Carolina Counties with Racial, Ethnic and Socioeconomic Characteristics**



## Exhibit 3. Sample Size Estimates for Area Household Samples with 8 PSUs

| Estimates per PSU per year | Estimated Rate | Estimated Sample Size |
|---|---|---|
| Number continuing in the LCS after birth | 0.95 | 313 |
| Number of births, accounting for multiple births | 1.033[a] | 329 |
| Number of reported pregnancies that go full term | 0.75 | 319 |
| Number of pregnancies that are reported to LCS | 0.95 | 425 |
| Annual number getting pregnant | 0.0786[a] | 448 |
| Number of women that agree to participate in the LCS | 0.7 | 5695 |
| Average number of women aged 15 to 39 per household | 1.113[b] | 8136 |
| Number of households selected for the LCS | 1 | 7310 |
| Number of households with at least one woman aged 15 to 39 | 0.383[b] | 7310 |
| Number of households that participate in the household screening interview | 0.85 | 19086 |
| Number of listed addresses that are occupied housing units | 0.89[c] | 22454 |
| Number of sample addresses needed | | 25230 |
| | | |
| Number of area segments | | 100 |
| Number of sample addresses needed per area segment | | 252 |
| Number of participating births needed per segment | | 3.1 |

Data Sources:
a. National Vital Statistics Reports, vol. 52, no 10
b. Public Use Microdata (PUMS) for the American Community Survey, 2002
c. Census 2000 Summary File 1 (SF1)

## Exhibit 4. Allocation of Sample Sizes to Area and Prenatal Care Samples as Part of Multiframe Design

| Estimates | Estimated Sample Size |
|---|---|
| Number of NC LCS participants | 5000 |
| Number of years of recruitment | 2 |
| Number of PSUs (counties or sets of adjacent counties) | 8 |
| Average number of participants per year, per PSU | 313 |
| Area household sample participants per year, per PSU (25% pre-conception) | 78 |
| Obstetrician sample participants per year, per PSU (75% in first trimester) | 235 |

## Exhibit 5. Sample Size Estimates for Area Household Sample with 8 PSUs as Part of Multiframe Design

| Estimates per PSU per year | Estimated Rate | Estimated Sample Size |
|---|---|---|
| Number continuing in the LCS after birth | 0.95 | 78 |
| Number of births, accounting for multiple births | 1.033[a] | 82 |
| Number of reported pregnancies that go full term | 0.75 | 79 |
| Number of pregnancies that are reported to LCS | 0.95 | 106 |
| Annual number getting pregnant | 0.0786[a] | 112 |
| Number of women that agree to participate in the LCS | 0.7 | 1419 |
| Average number of women aged 15 to 39 per household | 1.113[b] | 2028 |
| Number of households selected for the LCS | 1 | 1822 |
| Number of households with at least one woman aged 15 to 39 | 0.383[b] | 1822 |
| Number of households that participate in the household screening interview | 0.85 | 4756 |
| Number of listed addresses that are occupied housing units | 0.89[c] | 5596 |
| Number of sample addresses needed | | 6287 |
| | | |
| Number of area segments | | 30 |
| Number of sample addresses needed per area segment | | 210 |
| Number of participating births needed per segment | | 2.6 |

Data Sources:
a. National Vital Statistics Reports, vol. 52, no 10
b. Public Use Microdata (PUMS) for the American Community Survey, 2002
c. Census 2000 Summary File 1 (SF1)

## Exhibit 6. Sample Size Estimates for Prenatal Care Providers with 8 PSUs as Part of Multiframe Design

| Estimates per PSU per year | Estimated Rate | Estimated Sample Size |
|---|---|---|
| Number continuing in the LCS after birth | 0.95 | 235 |
| Number of births, accounting for multiple births | 1.033[a] | 247 |
| Number of reported pregnancies that go full term | 0.75 | 239 |
| Number of women that agree to participate in the LCS | 0.7 | 319 |
| Number of women not already recruited through the area household sample | 0.8 | 456 |
| Number of women successfully contacted by LCS recruiters | 0.9 | 570 |
| Number of women who complete the contact consent form | 0.6 | 634 |
| Number of women given the contact consent form by their obstetrician | 0.85 | 1056 |
| Number of women selected for the LCS | 1 | 1242 |
| Number of women reporting for prenatal care in their first trimester | 0.837[a] | 1242 |
| Number of pregnancies among women served by participating obstetricians | 0.75 | 1484 |
| Number of pregnancies needed in the PSU | | 1979 |

Data Sources:
a. National Vital Statistics Reports, vol. 52, no 10
b. Public Use Microdata (PUMS) for the American Community Survey, 2002
c. Census 2000 Summary File 1 (SF1)

## Exhibit 7. Sample Size Estimates for Area Household Samples with 30 PSUs

| Estimates per PSU per year | Estimated Rate | Estimated Sample Size |
|---|---|---|
| Number continuing in the LCS after birth | 0.95 | 83 |
| Number of births, accounting for multiple births | 1.033[a] | 87 |
| Number of reported pregnancies that go full term | 0.75 | 85 |
| Number of pregnancies that are reported to LCS | 0.95 | 113 |
| Annual number getting pregnant | 0.0786[a] | 119 |
| Number of women that agree to participate in the LCS | 0.7 | 1510 |
| Average number of women aged 15 to 39 per household | 1.113[b] | 2157 |
| Number of households selected for the LCS | 1 | 1938 |
| Number of households with at least one woman aged 15 to 39 | 0.383[b] | 1938 |
| Number of households that participate in the household screening interview | 0.85 | 5061 |
| Number of listed addresses that are occupied housing units | 0.89[c] | 5954 |
| Number of sample addresses needed | | 6690 |
| | | |
| Number of area segments | | 30 |
| Number of sample addresses needed per area segment | | 223 |
| Number of participating births needed per segment | | 2.8 |

Data Sources:
a. National Vital Statistics Reports, vol. 52, no 10
b. Public Use Microdata (PUMS) for the American Community Survey, 2002
c. Census 2000 Summary File 1 (SF1)

## Exhibit 8. Allocation of Sample Sizes to Area and Prenatal Care Provider Samples for Multiframe Design

| | Sample Size Estimates |
|---|---|
| Number of NC LCS participants | 5000 |
| Number of years of recruitment | 2 |
| Number of PSUs (counties or sets of adjacent counties) | 30 |
| Average number of participants per year, per PSU | 83 |
| Area household sample participants per year, per PSU (25% pre-conception) | 21 |
| Obstetrician sample participants per year, per PSU (75% in first trimester) | 62 |

## Exhibit 9. Sample Size Estimates for Area Household Sample with 30 PSUs as Part of Multiframe Design

| Estimates per PSU per year | Estimated Rate | Estimated Sample Size |
|---|---|---|
| Number continuing in the LCS after birth | 0.95 | 21 |
| Number of births, accounting for multiple births | 1.033[a] | 22 |
| Number of reported pregnancies that go full term | 0.75 | 21 |
| Number of pregnancies that are reported to LCS | 0.95 | 29 |
| Annual number getting pregnant | 0.0786[a] | 30 |
| Number of women that agree to participate in the LCS | 0.7 | 382 |
| Average number of women aged 15 to 39 per household | 1.113[b] | 546 |
| Number of households selected for the LCS | 1 | 490 |
| Number of households with at least one woman aged 15 to 39 | 0.383[b] | 490 |
| Number of households that participate in the household screening interview | 0.85 | 1281 |
| Number of listed addresses that are occupied housing units | 0.89[c] | 1507 |
| Number of sample addresses needed | | 1693 |
| | | |
| Number of area segments | | 10 |
| Number of sample addresses needed per area segment | | 169 |
| Number of participating births needed per segment | | 2.1 |

Data Sources:
a. National Vital Statistics Reports, vol. 52, no 10
b. Public Use Microdata (PUMS) for the American Community Survey, 2002
c. Census 2000 Summary File 1 (SF1)

## Exhibit 10. Sample Size Estimates for Prenatal Care Providers with 30 PSUs as Part of Multiframe Design

| Estimates per PSU per year | Estimated Rate | Estimated Sample Size |
|---|---|---|
| Number continuing in the LCS after birth | 0.95 | 62 |
| Number of births, accounting for multiple births | 1.033[a] | 65 |
| Number of reported pregnancies that go full term | 0.75 | 63 |
| Number of women that agree to participate in the LCS | 0.7 | 84 |
| Number of women not already recruited through the area household sample | 0.8 | 120 |
| Number of women successfully contacted by LCS recruiters | 0.9 | 150 |
| Number of women who complete the contact consent form | 0.6 | 167 |
| Number of women given the contact consent form by their obstetrician | 0.85 | 279 |
| Number of women selected for the LCS | 1 | 328 |
| Number of women reporting for prenatal care in their first trimester | 0.837[a] | 328 |
| Number of pregnancies among women served by participating obstetricians | 0.75 | 392 |
| Number of pregnancies needed in the PSU | | 522 |

Data Sources:
a. National Vital Statistics Reports, vol. 52, no 10
b. Public Use Microdata (PUMS) for the American Community Survey, 2002
c. Census 2000 Summary File 1 (SF1)