

National Health and Nutrition Examination Survey III (NHANES) DNA Data: Guidelines for Secondary Data Analysis of NHANES III Genetic Data

SUMMARY

The National Health and Nutrition Examination Survey (NHANES) is a program of periodic surveys conducted by the National Center for Health Statistics (NCHS) of the Centers for Disease Control and Prevention (CDC). Examination surveys conducted since 1960 by NCHS have provided national estimates of the health and nutritional status of the U.S. civilian non-institutionalized population. To add to the large amount of information collected for the purpose of describing the health of the population, blood lymphocytes were collected in NHANES III in anticipation of advances in genetic research.

There have been three announcements for use of these specimens announced (Tuesday, June 1, 1999 [64 FR 29321] August 8, 2002 [67FR 51585]) and (January 13, 2006[71FR 2248]). NCHS is announcing the availability of genotyping data files [attach web address for list of genes/SNPS] for secondary data analysis. There are two types of genotyping data available; genotyping data with a limited number of associated variables that have been anonymized so the data can no longer be linked to the public use files; and genotyping data that can be linked to all the NHANES public use data in the NCHS Research Data Center (RDC).

SUPPLEMENTARY BACKGROUND INFORMATION

The goals of NHANES are (1) to estimate the number and percentage of people in the U.S. population and designated subgroups with selected diseases and risk factors for those diseases; (2) to monitor trends in the prevalence, awareness, treatment and control of selected diseases; (3) to monitor trends in risk behaviors and environmental exposures; (4) to analyze risk factors for selected diseases; (5) to study the relation among diet, nutrition and health; (6) to explore emerging public health issues and new technologies; (7) to establish and maintain a national probability sample of baseline information on health and nutritional status.

The Third National Health and Nutrition Examination Survey (NHANES III) began in the Fall of 1988 and ended in the Fall of 1994. Survey data were collected and can be analyzed from two phases: Phase-1 was conducted from October, 1988 to October, 1991, and Phase-2 was conducted from October, 1991 to October, 1994. Both phases are nationally representative samples. For details of the sampling design see¹
<http://www.cdc.gov/nchs/about/major/nhanes/nh3data.htm>.

Blood specimens were collected from participants as a part of NHANES III. Lymphocytes were isolated from the blood collected from participants aged 12 years and older and stored in liquid nitrogen or as cell cultures immortalized with Epstein-Barr virus and frozen at the Molecular Biology Branch of DLS, NCEH, CDC Atlanta, GA. DNA in the form of crude cell lysates was made available for approved research projects from phase 2 (1991-1994) participants only.

Health information collected in the NHANES III is kept in strictest confidence. During the informed consent process, survey participants are assured that data collected will be used only for stated purposes and will not be disclosed or released to others without the consent of the individual or the establishment in accordance with section 308(d) of the Public Health Service Act (42 U.S.C. 242m). Although the consent form was signed by participants in the survey, and participants consented to storing specimens of their blood for future research, specific mention of genetic research was not included during NHANES III. Nevertheless, given the scientific importance of this resource, the CDC/NCHS Ethics Review Board (ERB) initially approved use of the DNA only if the samples were anonymized.

The anonymization requirements proved to be restrictive and difficult to implement, therefore, in August, 2001 the CDC/NCHS ERB approved a revised plan for using these specimens based on the guidelines in the August, 1999 National Bioethics Advisory Commission (NBAC) report on the use of stored biological materials for research. This revised plan included a process that gives researchers the ability to obtain more information associated with specimens for protocols that are determined by the ERB to have minimal risk for harm to the participant. For those protocols that cannot be conducted under unlinked (or anonymous) conditions, but have been determined to involve minimal risk, the revised plan allowed for linking of the genetic laboratory results to the NHANES data through the NCHS Research Data Center. This process ensured that confidentiality of the subjects' identity was maintained.

PROCEDURES FOR SUBMITTING PROPOSALS FOR ANALYSIS OF AVAILABLE GENOTYPING DATA

The cover page of the research proposal should contain the title of the research project, the name, address phone number and E-mail address of the lead investigator along with the name of the institution. Office for Human Research Protections (OHRP) assurance number for the institutions included in the research project should be included. CDC investigators need to include their Scientific Ethics Verification Number. E-mail submission of the proposal to NHANESgenetics@cdc.gov is encouraged.

The proposals should be a maximum of 1-2 single-spaced typed pages using twelve cpi type density.

Proposals will be reviewed by the NHANES program to assure the proposed research meets the restraints of the original consent and the approved plan for use of the NHANES III DNA. The ERB review will be conducted, even if investigator's proposal has been reviewed by their home institution.

Specific content of proposal:

The proposal title page should include the title of the research proposal; a list of the investigators and institutions; OHRP assurance number for the institutions included in the research project; address, phone number and E-mail address of lead investigator. CDC investigators need to include their Scientific Ethics Verification Number. The proposal should contain, and will be evaluated according to, the following elements:

- (1) **Specific Aims:** List the broad objectives, describe concisely what the research is intended to accomplish and state the specific hypotheses to be tested.
- (2) **Additional data sets:** List any additional data sets other than NHANES public use files that would be required.
- (3) **Expected Use:** Briefly describe how the research results would be used? (presentations, publications etc)?
- (4) **Participant Confidentiality:** Include a statement that the data will only be accessed in the RDC and that RDC procedures applied to these analyses provide protections with regards to data access as well as restrictions on analytic output pertaining to confidentiality risks. *Suggested text can be found below.*
- (5) **Period of performance:** Specify the project period. The period may be up to three years

Suggested Proposal Submission Text for the Section on Participant Confidentiality

The NHANES RDC provides a mechanism whereby researchers can access detailed NHANES data files in a secure environment, without jeopardizing the confidentiality of respondents. Currently, all analysis of NHANES genetic data must be performed at the NCHS' Research Data Center (RDC) in Hyattsville, MD since these data have been categorized as having the potential for disclosure risk. Certain restrictions have been put in place in the RDC to ensure that users cannot remove linked data or output that has not been subjected to a review for confidentiality. After analysis of the genetic data with other NHANES data in the RDC, researchers may take the results of their analyses off-site only after disclosure review by NHANES staff. Disclosure review consists of looking for tabular cells less than five including tables with multiple genetic variations linked to demographic variables, tables with genetic variations associated with multiple non-genetic variables, tables with variables that provide a high degree of disclosure risk such as geographic variables. Line listings of data are also suppressed during disclosure review. In general, disclosure review of NHANES genetic data results is consistent with the guidelines published in the NCHS Staff Manual on Confidentiality (see **Appendix II**, Requirements for the Release of NCHS Micro Data Files). Data linkage of NHANES genetic data to other sensitive data sets, such as Mortality and CMS data, will be require additional approvals beyond approval of this proposal.

Submission of Proposals:

Proposals will be accepted twice a year during the months of October and in May. Electronic submission of proposals is encouraged. Please submit proposals to: NHANESgenetics@cdc.gov

RESEARCH DATA CENTER COST SCHEDULE

Currently, all secondary data analysis of NHANES genetic data must be performed at the NCHS' Research Data Center (RDC) in Hyattsville, MD. There is a cost associated with using this center and these fees can be found at <http://www.cdc.gov/nchs/data/GuidelinesRDC11-8-05.pdf> under the heading "Costs for Using the RDC". Please inquire to NHANESgenetics@cdc.gov if you have any additional questions.

AGENCY AGREEMENT

Individuals requesting anonymized data sets must enter in a data use agreement with DHANES/NCHS stating that they will not transfer these data to a third party and will return the data at the termination of the project.

PROGRESS REPORTS

CDC/NCHS ERB continuation reports are required annually.

SEND REQUESTS FOR INFORMATION: NHANES genetic program E-mail address

1. Plan and Operation of the Third National Health and Nutrition Examination Survey, 1988-94. National Center for Health Statistics. Vital Health Stat (32) 1994.