

The Third National Health and Nutrition Examination Survey (NHANES III), 1988-94, Series 11, No. 10A (Hepatitis C Virus Genotype Data and Documentation) Data Release.

The National Health and Nutrition Examination Survey (NHANES) is a periodic survey conducted by NCHS. The third National Health and Nutrition Examination Survey (NHANES III), conducted from 1988 through 1994, was the seventh in a series of these surveys based on a complex, multi-stage sample plan. It was designed to provide national estimates of the health and nutritional status of the United States' civilian, noninstitutionalized population aged two months and older.

This data release, Series 11 No. 10A, contains the NHANES III Hepatitis C virus genotype data file and documentation. This data release does not replace the previous NHANES III data releases. The released data is in flat text format. A SAS program is provided to read the flat data. The user is advised to have someone who knows SAS data step to adapt the provided SAS program to his or her specific SAS environment.

Background information on the procedures, survey components, questionnaires, examination and laboratory methods, and statistical analysis guidelines is available on the NHANES III Questionnaires and Reports & Reference Manuals (CD-ROM). All data users are strongly encouraged to review these reference materials and reports before analyzing NHANES III data.

Guidelines for Data Users

- o NHANES III survey design and demographic variables are found on the Household Adult Data File, Household Youth Data File, the Laboratory Data File and the Examination Data File. In preparing a data set for analysis, other data files should be merged with either or both of the Adult Household Data File or the Youth Household Data File to obtain many important analytic variables.
- o All of the NHANES III public use data files are linked with the common survey participant identification number (SEQN). Merging information from multiple NHANES III data files using this variable ensures that the appropriate information for each survey participant is linked correctly.
- o NHANES III public use data files do not have the same number of records on each file. The Household Questionnaire Files (divided into two files, Adult and Youth) contain more records than the Examination Data File because not everyone who was interviewed completed the examination. The Laboratory Data File contains data only for persons aged one year and older. The Individual Foods Data File based on the dietary recall, the Prescription Medication Data File, and The Vitamin and Minerals Data File all have multiple records for each person rather than the one record per sample person contained in the other data files.
- o For each data file, SAS program code with standard variable names and labels is provided as separate text files on the CD-ROM that contains the data files. This SAS program code can be used to create a SAS data set from the data file.

- o Modifications were made to items in the questionnaires, laboratory, and examination components over the course of the survey; as a result, data may not be available for certain variables for the full six years. In addition, variables may differ by phase since some changes were implemented between phases. Users are encouraged to read the Notes sections of the file documentation carefully for information about changes.
- o Extremely high and low values have been verified whenever possible, and numerous consistency checks have been performed. Nonetheless, users should examine the range and frequency of values before analyzing data.
- o Some data were not ready for release at the time of this publication due to continued processing of the data or analysis of laboratory specimens. A listing of those data are available in the general information section of each data file.
- o Confidential and administrative data are not available or released to the public. Additionally, some variables have been recoded to protect the confidentiality of the survey participants. For example, all age-related variables were recoded to 90+ years for persons who were 90 years of age or older.
- o Some variable names may differ from those used in the Phase 1 NHANES III Provisional Data Release and some variables included in the Phase 1 provisional release may not appear on these files. Do not use the Phase 1 provisional release; use the current (six-year) release.
- o Although the data files have been edited carefully, it is possible that errors may still exist. Please notify NCHS staff (301-458-4636) of any suspected errors in the data file or the documentation. Refer to the NCHS website at <http://www.cdc.gov/nchs/nhanes.htm> for updates to these data files.

Analytic Considerations

- o NHANES III (1988-94) was designed so that the survey's first three years, 1988-91, its last three years, 1991-94, and the entire six years were national probability samples. Analysts are encouraged to use all six years of survey results.
- o Sample weights are available for analyzing NHANES III data. One of the following three sample weights will be appropriate for nearly all analyses: interviewed sample final weight (WTPFQX6), examined sample final weight (WTPFEX6), and mobile examination center (MEC)- and home-examined sample final weight (WTPFHX6). Choosing which of these sample weights to use in any analysis depends on the variables being used. A good rule of thumb is to use "the least common denominator" approach. In this approach, the user checks the variables of interest. The variable that was collected on the smallest number of persons is the "least common denominator," and the sample weight that applies to that variable is the appropriate one to use for that analysis. For more detailed information, see the Analytic and Reporting Guidelines for NHANES III (U.S. DHHS, 1996).

Referencing or Citing NHANES III Data

- o In publications, please acknowledge NCHS as the original data source. For instance, the reference for this NHANES III Hepatitis C Virus Genotype Data File in this release is: U.S. Department of Health and Human Services (DHHS). National Center for Health Statistics. Third National Health and Nutrition Examination Survey, 1988-1994, NHANES III Hepatitis C Virus Genotype data file (Series 11, No. 10A). Hyattsville, MD: Centers for Disease Control and Prevention, 2001.

Problems Using the Data

NHANES III is a wonderfully rich source of data and NCHS encourages you to use the data for research and analysis. However, the dataset is large and complex and familiarity with data file manipulation and analysis is required. NCHS does not have the personnel resources to perform analyses, check results, debug programs or do literature review for your work. Thorough review of the extensive documentation on the planning of the survey, analytic guidelines and individual datasets should resolve most questions. If you still have questions after careful review of the documentation, please contact the Data Dissemination Branch at (301)458-4636.

NHANES III Hepatitis C Virus Genotype Data File Index

Description	Variable Name	Positions
GENERAL INFORMATION		
Respondent identification number	SEQN	1-5
GENOTYPE	HCPGENTP	6
RNA POSITIVE	H CPRNA	7

NHANES III Hepatitis C Virus Genotype Data File

Positions SAS name	Counts	Item description and code	Notes
1-5 SEQN	364	Sample person identification number 00110-53490	
6 HCPGENTP	142	GENOTYPE 1 1A	See note
	73	2 1B	
	8	3 2A	
	27	4 2B	
	17	5 3A	
	3	6 4	
	5	7 6	
	89	Blank	
7 H CPRNA	283	RNA POSITIVE 1 Positive	See note
	81	2 Negative	

Special Notes

Blank Result Field

These laboratory tests were performed on 364 samples out of the 402 positive specimens for anti-HCV. If there was not sufficient sera for the assays or HCP was not equal to 1 then the sample person will have a blank in the result field.

Laboratory Tests

HCPRNA:

Hepatitis C RNA: Testing for HCV RNA by reverse-transcriptase-polymerase-chain reaction (RT-PCR) amplification of the 5' noncoding region was performed on anti-HCV positive samples. Samples found to be negative for HCV RNA were extracted a second time by the same procedure with an additional incubation at 50 degrees Celsius for 45 minutes with 25 units of reverse transcriptase (Boehringer Mannheim, Indianapolis) and 10 units of RNasin (Boehringer Mannheim).

NOTE: See the following references for more details on laboratory methods:

Nainan OV, Cromeans TL, Margolis HS. Sequence-specific, single primer amplification and detection of PCR products for identification of hepatitis viruses. *J Virol Methods* 1996;61:127-34.

Alter, MJ, Kruszon-Moran D, Nainan OV, McQuillan GM, et. al. The prevalence of hepatitis C virus in the United States, 1988 through 1994. *N Engl J Med* 1999;341:556-62.

HCPGENTP:

Hepatitis C genotype. The genotype of HCV RNA positive samples were determined by sequencing of 300 nucleotides in the NS5b region.

Note; See the following references for more details on laboratory methods:

Simmonds P, Holmes EC, Cha TA, et al. Classification of hepatitis C virus into six major genotypes and a series of subtypes of phylogenetic analysis of the NS-5 region. *J Gen Virol* 1993;75:2391-9.

Alter, MJ, Kruszon-Moran D, Nainan OV, McQuillan GM, et. al. The prevalence of hepatitis C virus in the United States, 1988 through 1994. *N Engl J Med* 1999;341:556-62.