

SUPPLEMENTARY METHODS

About NHANES III

The National Health and Nutrition Examination Survey (NHANES) is a periodic survey conducted by NCHS. The NHANES program began in the early 1960s and has been conducted as a series of surveys focusing on different population groups or health topics. In 1999, the survey became a continuous program that has a changing focus on a variety of health and nutrition measurements to meet emerging needs. The survey examines a nationally representative sample of about 5,000 persons each year. These persons are located in counties across the country, 15 of which are visited each year (<https://www.cdc.gov/nchs/nhanes/index.htm>).

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. The NHANES interview includes demographic, socioeconomic, dietary, and health-related questions. The examination component consists of medical, dental, and physiological measurements, as well as laboratory tests administered by highly trained medical personnel. Institutional Review Board (IRB) approval and documented consent were obtained from participants. The data are de-identified and freely accessible to the public.

As the data are de-identified and freely accessible to the public, we downloaded the relevant data sets from the website (<https://wwwn.cdc.gov/nchs/nhanes/nhanes3/>). In preparing a data set for analysis, other data files must be merged with either or both of these files to obtain many important analytic variables. 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. 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.

The National Center for Health Statistics (NCHS) has linked data collected from several NCHS population surveys with death certificate records from the National Death Index (NDI). The public-use linked mortality files (LMF) include a limited set of variables for adult participants only. The files provide mortality follow-up data from the date of survey participation through December 31, 2019. In the public-use LMF, each survey participant who is linkage eligible for mortality follow-up is assigned a vital status code (0=assumed alive, 1=assumed deceased). The leading causes of death include diseases of the heart, malignant neoplasms, chronic lower respiratory diseases, accidents, cerebrovascular diseases, Alzheimer's disease, diabetes mellitus, influenza and pneumonia, nephritis, nephrotic syndrome and nephrosis (National Center for Health Statistics Division of Analysis and Epidemiology. NHANES III Public-use Linked Mortality Files, 2019. Hyattsville, Maryland. Available from:<https://www.cdc.gov/nchs/data-linkage/mortality-public.htm>. doi:10.15620/cdc:117141).

About AIC

Akaike Information Criterion (AIC) is a metric used in statistical model selection to balance model fit and complexity. Introduced by Hirotugu Akaike in 1973, it is calculated as $AIC=2k-2\ln(L)$, where k is the number of parameters in the model and L is the maximum likelihood. The AIC balances goodness of fit, represented by $2\ln(L)$, with model complexity, penalized by $2k$. A lower AIC value indicates a better trade-off between accuracy and simplicity, making it the preferred model. This criterion helps prevent overfitting by ensuring that adding more parameters to the model is justified by a significantly better fit.¹

REFERENCE

1. Akaike H. A new look at the statistical model identification. IEEE Trans Automat Contr 1974;19:716-723.