SOCIETY OF ACTUARIES®

Mortality and Morbidity Data Sources

March 2019

Mortality and Morbidity Data Sources

AUTHOR R. Jerome Holman, FSA, MAAA

Caveat and Disclaimer

This study is published by the Society of Actuaries (SOA) and contains information from a variety of sources. It may or may not reflect the experience of any individual company. The study is for informational purposes only and should not be construed as professional or financial advice. The SOA does not recommend or endorse any particular use of the information provided in this study. The SOA makes no warranty, express or implied, or representation whatsoever and assumes no liability in connection with the use or misuse of this study.

Copyright ©2019 All rights reserved by the Society of Actuaries

TABLE OF CONTENTS

© 2019 Society of Actuaries

Section 1: Introduction

Many readers are likely to be familiar with mortality and morbidity data sources such as the CDC's Wide-ranging Online Data for Epidemiologic Research (CDC WONDER), the Human Mortality Database and the Social Security Administration. This report provides information on other mortality and morbidity data sources to consider. Some of these data sources are rich and complex. This report provides an overview of the data sources' characteristics and a roadmap to delving into their details.

1. British Doctors Study
2. National Health Interview Survey
3. Study of Seventh-Day Adventists
4. National Health and Nutrition Examination Survey
5. Health and Retirement Study
6. University of WA – Burden of Cardiovascular Diseases among US States 1990-2016

Section 2 of the report provides a brief description of the data sources and the study or research that originated them. Section 3 lists attributes for the first five data sources. The sixth requested data source does not have data directly associated with it. All information in this report is publicly available on the internet.

© 2019 Society of Actuaries

Section 2: Data Source Overview

This section gives a brief overview of each data source. You can jump to the detail in Section 3 for each one, except for Section 1.6, by holding the Ctrl key and clicking on the first hyperlink shown in each subsection.

1.1 British Doctors Study

The British Doctors Study, which began in 1951, was the world's first large prospective study of the effects of smoking to establish a convincing linkage between tobacco smoking and cause-specific mortality, and demonstrated prospectively the risk of death from lung cancer (1954) and myocardial infarction and chronic obstructive pulmonary disease (1956).

Its results, then and subsequently (particularly the 10-year, 20-year, 40-year and 50-year results), have substantially influenced national and personal decisions about quitting, as they assessed the lifelong effects of smoking and then of stopping.

Control of the data is maintained by the Nuffield Department of Population Health (NDPH) at the University of Oxford. The NDPH performs research on a wide range of mortality and morbidity topics. All data associated with the British Doctors Study and other research is subject to restricted access to researchers outside of the NDPH. Use of the data by external researchers requires following a rigorous approval process and fees, which depend upon the scope of the request.

1.2 National Health Interview Survey

The National Health Interview Survey (NHIS) provides information on the health of the US civilian noninstitutionalized population through confidential interviews conducted in households by the National Center for Health Statistics (NCHS). The NHIS is one of the nation's largest in-person household health surveys. It provides data for analyzing health trends and tracking progress towards achieving national health objectives.

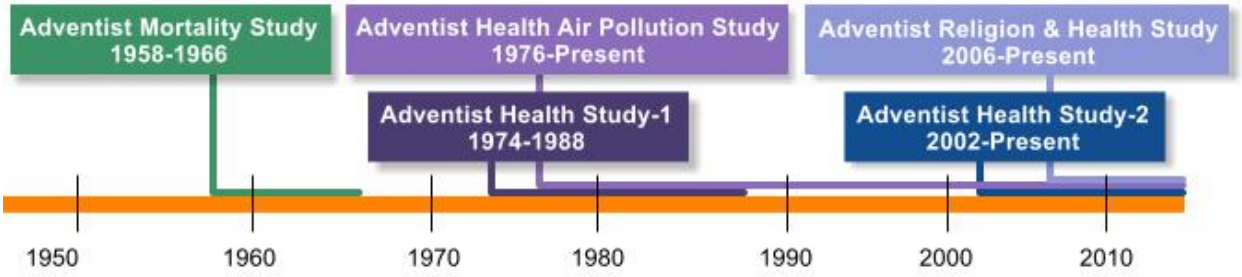
These data, continuously collected throughout the year, are also used for epidemiological and policy analysis, such as characterizing those with various health conditions, determining barriers to accessing and using appropriate health care, and evaluating federal health programs.

© 2019 Society of Actuaries

1.3 Study of Seventh-Day Adventists

Seventh-day Adventists (Adventists) have increasingly become the objects of epidemiologic studies, both because they tend to be far more homogeneous in many lifestyle choices and because they are more heterogeneous in nutritional habits than the general population. Certain lifestyle characteristics, such as heavy cigarette smoking, consumption of alcohol, and diets heavy in fats may confound or modify the effects of other factors, making it difficult to study members of the general population.

In the Adventist population, these potentially distorting characteristics are largely absent, making other factors more easily observed. Perhaps even more importantly, the wide range of dietary habits, from strict vegetarianism to a normal American diet, greatly enhances the ability of investigators. A timeline of the studies is shown below. The Adventist Mortality Study and the Adventist Health Study-1 (AHS-1) are closed and the other studies are ongoing. The most significant ongoing study, AHS-2, and all others shown below are managed by the Loma Linda University.



1.4 National Health and Nutrition Examination Survey

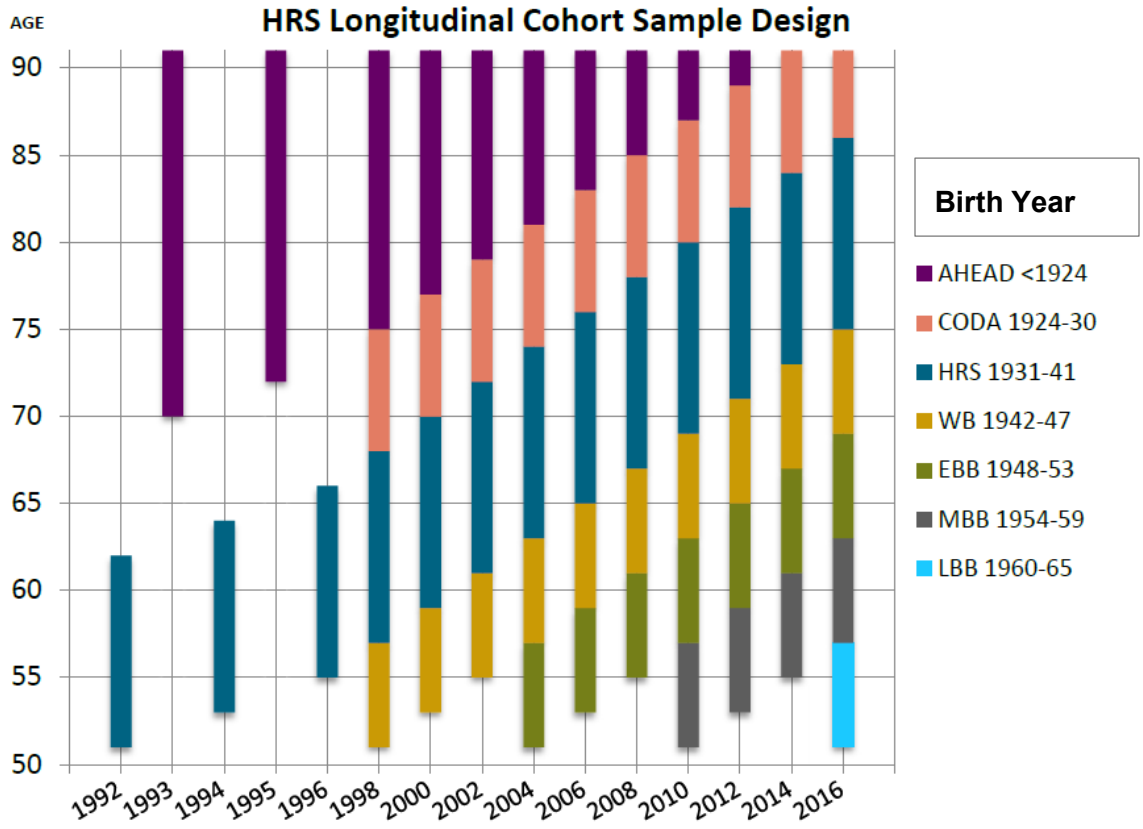
The National Health and Nutrition Examination Survey (NHANES) is a program of studies managed by the NCHS that is designed to assess the health and nutritional status of adults and children in the US. The survey is unique in that it combines interviews and physical examinations.

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.

© 2019 Society of Actuaries

1.5 The Health and Retirement Study

The Health and Retirement Study (HRS), managed by the University of Michigan, was the first longitudinal study of older people to include detailed economic and health information in the same survey. It was created by an act of Congress in 1990 and now is the largest and most comprehensive nationally representative multi-disciplinary panel study of Americans over the age of 50. Different cohorts have been added to the study at various points in time per the graph below.



In its original conceptualization, the HRS study was designed to follow age-eligible individuals and their spouses as they made the transition from active worker into retirement; the *Asset and Health Dynamics Among the Oldest Old* (AHEAD) study was designed to examine the dynamic interactions between health, family, and economic variables in the post-retirement period at the end of life. The HRS study spanned three waves of data collection: 1992, 1994, and 1996. The AHEAD study included two waves: 1993 and 1995. The HRS and AHEAD studies were merged in 1998. At that time, and in subsequent years, additional cohorts were added.

HRS data is very rich and complex. While substantial portions of it are easily accessible online following a simple registration process, the data is highly dispersed in different compartments that require SAS, SPSS, or STATA to process. The data can also be cross-linked to data from the Centers for Medicaid and Medicare, the Social Security Administration, or other sources, including NHANES, but access to data external to HRS requires additional approvals outside of HRS.

© 2019 Society of Actuaries

1.6 University of WA – Burden of Cardiovascular Diseases among US States 1990-2016

There is not a single database or source associated with this research. The data used is widely dispersed and not all publicly available. The data includes, but is not limited to, the following sources: the US Census Bureau (population counts), the National Center for Health Statistics (death certificate data), NHANES, and inpatient and outpatient claims data from private and public insurance schemes. The research, *The Burden of Cardiovascular Diseases Among US States, 1990-2016*, is posted on the Journal of the American Medical Association's (JAMA) website, https://healthmetrics.heart.org/wp-content/uploads/2018/04/The-Burden-of-Cardiovascular-Diseases-Among-U.S.-States-1990-2016.pdf. The primary author of the report is Gregory A. Roth, MD, MPH at the University of Washington.

There is a Cardiovascular Health Research Unit (CHRU) at the University of Washington. It is an amalgamation of the efforts of other areas including the Department of Medicine. The University of Washington offers to collaborate with external researchers, but they do not have a unique database like the British Doctors or Adventists.

Note that because there isn't an accessible block of data from the University of Washington associated with this work, no data detail is provided relative to this research in Section 3. The JAMA article cited above gives a complete description of data used in that research.

© 2019 Society of Actuaries

Section 3: Data Source Detail

3.1 British Doctors Study

Characteristic	Description
Sponsor	UK Medical Research Council, Cancer Research UK, and British Heart Foundation
Managing Organization	The University of Oxford, Nuffield Department of Population Health (NDPH)
Purpose	Study of the effects of smoking on cause-specific mortality that was initiated in 1951.
Coverage	All British doctors were sent a survey about smoking habits in October 1951. Of the 59,600 questionnaires mailed, 41,024 replies were received and 40,701 (34,494 males and 6,207 females) were sufficiently complete to be included in the follow-up. Because of the limited sample size and limited tobacco consumption, females were excluded from most reports, and the study focused on males. The 1978 questionnaire sought information from all male doctors born in the 20th century about a wider range of characteristics (including alcohol consumption and self-reported body mass index) and invited them to participate in a randomized trial of prophylactic daily aspirin to prevent death from stroke, myocardial infarction, or other vascular conditions. This trial has engendered others, some much larger, and has contributed to statistically stable meta-analyses of all trials.
Frequency	Follow-up questionnaires about changes in smoking habits were sent in 1957, 1966, 1971, 1978, 1991, 1998 and 2001. The span of gathered data was expanded (as noted in Coverage) in 1978.
Timing	The last study was in 2001. No other data is currently being gathered.
Data Source	Data on habits and health characteristics was gathered by survey. Data on the incidence of cancer and of cause-specific mortality among the male doctors is still continuing. All study participants are registered with the Office of National Statistics (ONS) to obtain mortality data (date of and cause of death) are provided by NHS Digital on behalf of ONS and is sourced from Civil Registration Data. Study participants are also registered with the National Cancer Registration and Analysis Service (NCRAS) to obtain cancer diagnoses.
Data Content	There is no manual of field descriptions, but a description of the survey and how data was classified indicates likely data specifics. Key aspects of the survey and follow-up process are described here: https://embryo.asu.edu/pages/british-doctors-study-1951-2001. While the data is based on individual responses, personal identifying information has been removed. 1. The short questionnaire included demographic information, smoking status, and follow-up questions, including amount of tobacco consumed and age at which one started smoking, for those who classified themselves as smokers. 2. The doctors were grouped by their ages at the beginning of the study: 34 years and younger, 35 to 44, 45 to 54, 55 to 64, 65 to 74, 75 to 84, and 85 and older. Doll and

© 2019 Society of Actuaries

	Hill also grouped the doctors by their estimated daily tobacco consumption using four categories: none, 1 to 14 grams, 15 to 24 grams, and 25 or more grams. The researchers also separated the doctors by their method of smoking: cigarettes, pipes, or both. 3. With access to the doctors' medical histories, the researchers collected data on lung cancer, cancers outside of the lungs, non-cancer respiratory diseases and general illness, coronary thrombosis, later called myocardial infarction and commonly known as heart attack, cardiovascular diseases other than coronary thrombosis, and all other diseases.
Data Availability	**Data use is restricted.** The British Doctors Study database is not shared with organizations outside of the University of Oxford and only accredited researchers in the Clinical Trial Service Unit & Epidemiological Studies Unit (CTSU), University of Oxford, have access to the data. Anonymous results of the study may be made available to collaborators and relevant bona fide researchers according to the Nuffield Depart of Population Health (NDPH), University of Oxford data-sharing policy: https://www.ndph.ox.ac.uk/about/data-access-policy.
Data Access	The Data Access and Sharing Policy for the Nuffield Department of Population Health, University of Oxford, defines policies and procedures within the department to ensure adherence to the Research Council's UK and Expert Advisory Group on Data Access common principles on data policy (1-3) and to allow appropriate data sharing for scientific research. The policy covers raw data, summary tables, and analyses, which are not released in publications or online from all studies held by the NDPH, regardless of the original study location or the source of funding. Study-specific policies should adhere to the same principles and must not conflict with this policy. It does not cover Freedom of Information requests or data subject access requests under the Data Protection Act. **Key Aspects:** 1. Those requesting access are asked to apply through the Richard Doll Centenary Archive preliminary inquiry form available from: Richard Doll Archive Data Access Preliminary Enquiry Form. Requesters should be employees of a recognized academic institution, health service organization, or commercial research organization with experience in medical research; and should be able to demonstrate, through their peer-reviewed publications in the area of interest, their ability to carry out the proposed study. 2. During the preliminary application stage, the requester and the Custodian have the opportunity for initial discussions before a full application is submitted using the Richard Doll Archive's full application form, which asks for details of the proposed study and required data/samples. Requesters are asked to provide their curriculum vitae and details of their current affiliation. The Richard Doll Centenary Archive reserves the right to contact the requester's institution as part of the process of confirming the requester's status. 3. Accepted (for both active and archived studies) proposals will be reviewed at the next appropriate Oversight Committee meeting as part of their monitoring of data access and sharing requests. Requests that are refused will also be reviewed by the Committee and requesters may appeal to the Oversight Committee if they disagree with the Custodian's refusal. For inactive studies, provisional agreement between the

© 2019 Society of Actuaries

	designated Custodian and the requester will be reviewed by the Oversight Committee in person, by telephone, or by email and a final decision made on approval or further action. 4. It is the responsibility of the Requestor to obtain approval from Local Research Governance entities within the University, e.g. Ethics Committee and the Research Ethics Committee associated with a specific study. Linked data may not be available without application to the provider. The Requestor is responsible for obtaining data use approvals. 5. When considering a request for data sharing, the likely time commitment for NDPH staff needs to be realistically assessed and the Requester will need to confirm that funding is available to cover reasonable NDPH costs.
Allowable Use	The data collections will be used for the purposes of medical research and education only and within the constraints of the consent (and any other legal basis) under which the data were originally gathered, and of any contractual agreements between the study from which data are requested and its funders or external data sources.
Other	The Nuffield Department of Population Health (NDPH) at the University of Oxford also conducts ongoing morbidity and mortality research in areas unrelated to the British Doctors Study. These studies are listed here: https://www.ndph.ox.ac.uk/about/richard-doll-centenary-archive. The philosophy of research at NDPH is explained by the department head, Rory Collins, https://soundcloud.com/uffieldepartmentofopulationealth/kat-arney-interviewsrory-collins-09-oct-17. Specific datasets associated with various studies are listed below. Each of these datasets follows the convention of data access and allowable uses described above. CHINA KADOORIE BIOBANK This is a prospective study of 500,000 men and women living in 10 regions of China, with data access procedures available here. The China Kadoorie Biobank (CKB), known previously as the Kadoorie Study of Chronic Disease in China (KSCDC), is set up to investigate the main genetic and environmental causes of common chronic diseases in the Chinese population. MILLION WOMEN STUDY This study contains a prospective cohort of 1.3M women living in the UK, with data access procedures available here. While the initial stimulus was to study the risk of breast cancer and other conditions associated with the use of menopausal hormones, most of the women who joined the study had reached adulthood in the 1960s and had considerably different lifestyles to those of previous generations. For example, large proportions had begun smoking and using oral contraceptives as teenagers and young adults. The prevalence of obesity was also increasing and there were claims that other behaviors, such as diet, had important effects on health. To answer questions about the effects of these factors on health reliably requires large scale population-based evidence. The Million Women Study is, therefore, investigating the short-term and long-term effects of these and many other factors on women's risk of developing or dying from different types of cancer, heart disease, stroke, dementia, and other mental and neurodegenerative disorders, and many other conditions in middle and old age. EPIC OXFORD This study contains a prospective cohort of 65,000 men and women living in the UK, with data access procedures available here. The Oxford component of the European Prospective

© 2019 Society of Actuaries

	Investigation into Cancer and Nutrition (EPIC) is a prospective cohort of 65,000 men and women living in the UK, many of whom are vegetarian. NATIONAL REGISTRY OF CHILDHOOD TUMORS The UK National Registry of Childhood Tumors (NRCT) Epidemiological Studies Research Database is a legacy research dataset, formerly held by the Childhood Cancer Research Group (now closed) at the University of Oxford. The NRCT research database includes information on 98,000 children diagnosed with cancer in the UK between 1962 and 2013, and on 67,000 children without cancer. The RDCA Data Access Oversight Committee has responsibility for access to data held in the NRCT database.

© 2019 Society of Actuaries

3.2 National Health Interview Survey

Characteristic	Description
Sponsor	US Government, primarily the NCHS, and through a variety of agencies that allocate funds to the National Health Interview Survey (NHIS) for the purposes of ongoing or periodic supplements to the survey.
Managing Organization	Center for Disease Control National Center for Health Statistics (CDC NCHS)
Purpose	The NHIS provides data for analyzing health trends and tracking progress towards achieving national health objectives. These data are also used for epidemiological and policy analysis, such as characterizing those with various health conditions, determining barriers to accessing and using appropriate health care, and evaluating federal health programs.
Coverage	The NHIS is based on a small sample of randomly-selected US households. Information in the Family Core questionnaire is collected on all household members. A household adult reports information for a randomly selected child less than 18 years of age in the Child Core questionnaire, and a randomly selected adult reports for him/herself in the Adult Core questionnaire. This format has been in place since 1997. The survey is being redesigned for 2019. Some of the family core questions are being migrated to the adult and child sections. The information gathered will not be identical to the 1997 design. It is being changed to adapt to the changing needs of what is deemed the most useful information to surveil.
Frequency	Data is gathered continuously throughout the year.
Timing	Generally, final files are released in late June following the survey year. Some data is available on an early release basis that can precede the final release by up to nine months. Early release data involves two formats. Available online: selected estimates based on NHIS data, health insurance coverage, wireless substitution (estimated landline vs. wireless use in households), and other selected topics produced on an as needed basis. This link lists online 2018 early release data, which are in the form of charts and graphs: https://www.cdc.gov/nchs/nhis/releases/released201812.htm#4. Available through the Research Development Center (RDC): data access of preliminary microdata files through an RDC (on site or remotely) ***is subject to a required approval process***. This data is unweighted and may not be scrubbed for personal data, which may be the driver of required approval for its use: https://www.cdc.gov/rdc/.
Data Source	Data is gathered from the US civilian noninstitutionalized population through confidential interviews conducted in households.
Data Content	The data content changes with the periodic redesign of the survey. The last redesign was in 1997. These topics are covered: limitation of activity, injuries, health insurance, access to

© 2019 Society of Actuaries

	health care, health care utilization, conditions (regarding health), behaviors (alcohol, tobacco, and physical activity), immunizations, and AIDS: https://www.cdc.gov/nchs/nhis/nhis_redesign.htm. Each year has an array of files. As an example, the 2017 NHIS data release consists of seven questionnaires, six core data files, a paradata file, a Family Disability Questions file, and a Functioning and Disability file. The 2017 NHIS is divided into 9 data files, each available in two formats: a column-delimited text (ASCII) format, and a comma-separated values (CSV) format. Each format is released in its own compressed ZIP file. This link gives more details on the 2017 file array: ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NHIS/2017/readme.pdf. The survey is being redesigned for first use in 2019. Pilot tests were done in 2018 where the questions have been modified to have annual core, rotating core, and sustaining and periodic supplements: https://www.cdc.gov/nchs/nhis/2019_quest_redesign.htm.
Data Availability	The annual final data and early release information are publicly available. See timing for restrictions on early release microdata files.
Data Access	Publicly available microdata can be downloaded. This is the link to the 2017 files: https://www.cdc.gov/nchs/nhis/nhis_2017_data_release.htm. Publicly available early release data summaries (graphs and charts) can be viewed online (see Timing).
Allowable Use	The Public Health Service Act (Section 308 (d)) provides that the data collected by the National Center for Health Statistics (NCHS) may be used only for the purposes of health statistical reporting and analysis. Any effort to determine the identity of any reported case is prohibited by this law. NCHS does all it can to assure that the identity of data subjects cannot be disclosed. All direct identifiers, as well as any characteristics that might lead to identification, are omitted from the data files. Any intentional identification or disclosure of a person or establishment violates the assurances of confidentiality given to the providers of the information. Therefore, users will: *1. Use the data in these data files for statistical reporting and analysis only.* *2. Make no use of the identity of any person or establishment discovered inadvertently and advise the Director, NCHS, of any such discovery (301-458-4500).* *3. Not link these data files with individually identifiable data from other NCHS or non-NCHS data files.* By using these data, you signify your agreement to comply with the above-stated statutorily-based requirements.
Other	An alternative access point is to use the Integrated Public Use Microdata Series (IPUMS) data aggregator facility that is part of the Institute for Social Research and Data Innovation at the University of Minnesota. The advantage of this data is that variables, all sourced from the original NHIS microdata files, are coded consistently across years where there have been

© 2019 Society of Actuaries

	changes and are accessible in groups of years, but not all variables in the original NHIS files are carried in these files. Access to the documentation and to IPUMS NHIS data analysis using the online tabulator is freely available. To get access to the data for downloading a customized data extract, users must agree to specified conditions of responsible use, which are like the conditions for using the NHIS public use files. For purposes of internal recordkeeping, and to provide the IPUMS NHIS staff with a clear sense of the user constituency (to improve outreach and better serve users), registration also requires users to provide some information about themselves, such as their discipline, academic or non-academic status, and institutional affiliation. Registered users are automatically added to the IPUMS NHIS e-mail list and receive occasional newsletters reporting data releases and new website features. To register for access to the data, go to the IPUMS NHIS registration webpage. This is the link for the NHIS IPUMS data. As an interesting side note, IPUMS aggregates many types of data, e.g. census data. Their mission is not limited to the NHIS. https://www.ipums.org/healthsurveys.shtml

© 2019 Society of Actuaries

3.3 Study of Seventh-Day Adventists

Characteristic	Description
Sponsor	Loma Linda University has been the main sponsor for the Adventist studies. AHS-2 has also received funding from the National Cancer Institute, World Cancer Research Fund, and the US Department of Agriculture. The EPA provides funding for the Adventist Health Air Pollution Study that is ongoing and a sub-study of AHS-2.
Managing Organization	Loma Linda University, School of Public Health
Purpose	Adventist Health Studies are long-term cohort studies exploring the links between lifestyle, diet, and disease among Seventh-day Adventists. They afford an opportunity to study the effects of heterogenous diets within the backdrop of relatively homogenous lifestyles where cigarette smoking and consumption of alcohol are largely absent. The earliest studies were focused on cancer and were later expanded to other diseases. Generally, they compared outcomes between Adventists and non-Adventists. The current Adventist Health Study-2 (AHS-2) is focused on outcomes related to varying diet within the Adventist community.
Coverage	These studies began in 1958 with the Adventist Mortality Study that was a cohort or prospective study of 22,940 California Adventists. Organized at Loma Linda University, it entailed an intensive 5-year follow-up and a more informal 25-year follow-up. More than 96,000 church members from the US and Canada are participating in the current study, AHS-2, conducted by researchers at the Loma Linda University School of Public Health.
Frequency	AHS-2 is ongoing with hospitalization forms mailed out every two years to the 96,000 participating church members in the US and Canada.
Timing	The AHS-2 study is ongoing and could last for 15-20 more years.
Data Source	Study data has been gathered through surveys. A participant fills out a survey detailing lifestyle and medical history. Hospitalization surveys are sent every other year. The study is closed to new participants.
Data Content	Specific data for the studies is not published. The surveys provide an indication of the type of data that may be available regarding AHS-2. The link below lists the enrollment and follow-up forms. https://publichealth.llu.edu/adventist-health-studies/researchers/questionnaire-and-survey-archives Survey forms for earlier Adventist studies are not published.
Data Availability	Data use is restricted, but AHS-2 is committed to furthering scientific investigation through collaboration with external researchers. The breadth of the data collected offers interested researchers a variety of interesting avenues of research in nutritional epidemiology and

© 2019 Society of Actuaries

	beyond. They welcome proposals from external collaborators who wish to make use of their data. AHS-2 data from the following sources may be available for approved collaborative research projects. More detailed descriptions of the available data variables will be provided upon approval of the letter of intent (see Data Access below) for help in crafting the project proposal. 1. Baseline questionnaire data 2. Calibration study data 3. Biomarkers 4. Outcome data (cancer incidence, mortality)
Data Access	Access to data occurs as part of an approved project proposal. Project proposal guidelines are described here: https://publichealth.llu.edu/sites/publichealth.llu.edu/files/docs/AHS-2%20Guidelines%20for%20external%20collaborators.pdf. **Key Aspects:** Letter of intent – Brief summary of the proposal that can be submitted at any time. Feasibility is judged with respect to scientific interest that does not duplicate other work that is either ongoing or the subject of a grant in the approval process. Responses to letters generally take 30 days. Proposal – A 3-5 page proposal with more rigor and detail than the letter of intent. Proposal deadlines are May 15, July 15, and November 15. Collaboration – A Loma Linda AHS-2 Investigator will be assigned to work with the external collaborator, i.e. the requestor. The Investigator helps facilitate and shape the data request. There is internal sharing with other AHS-2 investigators during the project and computer code must be reviewed and approved by an AHS-2 epidemiologist. Cost – The AHS-2 Investigator can run about 5%-10% FTE per year depending on the project complexity. Some projects might require an internal statistician. Data Site: All data must remain on Loma Linda computers. Access on a remote basis is feasible. Some onsite training may be needed. Most data is accessed and processed with SAS and R. Co-authoring – At least one member of the AHS-2 investigative team will be a co-author with sign-off rights on any manuscript. All manuscripts are subject to department head review.
Allowable Use	Data use must be in harmony with the original AHS-2 purpose and informed consent. Generally, study participants have given consent to data use on a personally non-identifiable basis, but all research involving human subjects requires a higher level of approval from the Loma Linda Institutional Review Board (IRB).
Other	Although not related to Loma Linda research, studies of Adventists have been done in other parts of the world, including Norway, Holland, Poland, Denmark, and Japan, but they, like the original Adventist mortality study, had a fundamental weakness: nonfatal events for various diseases were not measured.

© 2019 Society of Actuaries

3.4 National Health and Nutrition Examination Survey

This link is the gateway to detailed information about the National Health and Nutrition Examination Survey (NHANES): https://www.cdc.gov/nchs/nhanes/index.htm. This is a comprehensive overview of the 1999-2010 survey cycles: https://www.cdc.gov/nchs/data/series/sr_01/sr01_056.pdf. Because subsequent cycles are similar in structure and operation, the 1999-2010 survey document gives a good summary of the main aspects of current NHANES surveys.

Characteristic	Description
Sponsor	US Government
Managing Organization	Center for Disease Control National Center for Health Statistics (CDC NCHS)
Purpose	Findings from this survey are used to determine the prevalence of major diseases and risk factors for diseases. Information is used to assess nutritional status and its association with health promotion and disease prevention. NHANES findings are also the basis for national standards for such measurements as height, weight, and blood pressure. Data from this survey is in epidemiological studies and health sciences research, which help develop sound public health policy, direct and design health programs and services, and expand the health knowledge for the US.
Coverage	With respect to the 1999-2010 period, the survey used in-person face-to-face interviews and physical examinations for data collection. Approximately 5,000 people per year participated in NHANES. The 5,000 people surveyed each year are representative of the civilian non-institutionalized US population. Each two-year cycle had about 10,000 participants. In 2015-2016, 15,327 persons were selected for NHANES from 30 different survey locations. Of those selected, 9,971 completed the interview and 9,544 were examined.
Frequency	Data is gathered on a two-year collection cycle. The most recent cycle is 2017-2018. The survey became continuous in 1999. Prior to that, it was done periodically starting in 1959.
Timing	It is anticipated that an initial data release will occur approximately nine months after the completion of each two-year data collection cycle, and intermittent releases will follow as remaining data are processed, until all releasable data are available for public use.
Data Source	Data is gathered from the US civilian noninstitutionalized population through confidential interviews conducted in households and physical exams in mobile exam centers (MEC) and, in some instances, post MEC data collection.
Data Content	The data content changes with the NHANES biennial cycle. The link below is a summary of the available data from the 1999-2016 two-year cycles. The data is shown in sub-groups of questionnaire topics, post-examination follow-up questionnaire components, examination components, and laboratory components. https://wwwn.cdc.gov/nchs/data/nhanes/survey_contents.pdf Each biennial cycle during 1999-2018 and prior periodic surveys have their data grouped under an access point. This link for 2015-2016 is an example of the data content (2017-2018 content is not fully populated yet). This entry point opens to separate data compartments for

© 2019 Society of Actuaries

	demographics, dietary, examination, laboratory, questionnaire, and limited access data. Detail on how the data was gathered and advisements on its use are also described. https://wwwn.cdc.gov/nchs/nhanes/continuousnhanes/default.aspx?BeginYear=2015
Data Availability	The Division of Health Examination Surveys (DHANES) disseminates NHANES data and outlines dissemination procedures. The DHANES policy is consistent with the CDC and NCHS policies, including the guiding principles of making high quality data available: 1. As widely as practicable; 2. As soon as possible after data collection; 3. In as much detail as possible; and 4. While maintaining survey participant confidentiality. Data is both publicly and non-publicly available. Publicly available data is described in Data Content. Data that has not been publicly released can include restricted variable components of publicly released data or non-public data on sera, DNA, and imaging studies. Gaining access to data that has not been publicly released requires a research proposal to the NCHS Research Development Center. If requested data are not currently collected or available in NHANES, a proposal to obtain (and fund) the new data items can be submitted via email to the NHANES Biospecimen Program. The NHANES Project Officer and a technical panel evaluate all proposals for scientific merit. The NCHS Human Subject Contact and Ethics Review Board (ERB) then review the proposal for any potential human subject concerns and the NCHS Confidentiality Officer for disclosure risk.
Data Access	Publicly available microdata can be downloaded. Data is widely dispersed in the compartments described in Data Content. As an example, this link goes to the 2015-2016 demographic data: NHANES 2015 Demographic Data. NHANES data files are stored in SAS transport format. There are many files for each NHANES cycle to reduce their download time and for size management. They can be read and processed directly in SAS or can be converted with a free SAS viewer to .csv formatted files. This link provides a series of tutorials for the use of NHANES data: https://wwwn.cdc.gov/nchs/nhanes/tutorials/default.aspx.
Allowable Use	The Public Health Service Act (Section 308 (d)) provides that the data collected by the National Center for Health Statistics (NCHS) may be used only for the purpose of health statistical reporting and analysis. Any effort to determine the identity of any reported case is prohibited by this law. NCHS does all it can to assure that the identity of data subjects cannot be disclosed. All direct identifiers, as well as any characteristics that might lead to identification, are omitted from the data files. Any intentional identification or disclosure of a person or establishment violates the assurances of confidentiality given to the providers of the information. Therefore, users will: *1. Use the data in these data files for statistical reporting and analysis only.* *2. Make no use of the identity of any person or establishment discovered inadvertently and advise the Director, NCHS, of any such discovery (301-458-4500).* *3. Not link these data files with individually identifiable data from other NCHS or non-NCHS data files.*

© 2019 Society of Actuaries

	By using these data, you signify your agreement to comply with the above-stated statutorily based requirements.
Other	NHANES collects biospecimens from participants (sera, plasma, urine, and DNA) and makes them available for public research. When the research produces data that can be shared, it is added to available public data. All research on biospecimens requires approval of the proposed research. This link gives an overview of the biospecimen program: https://www.cdc.gov/nchs/nhanes/biospecimens/biospecimens.htm.

© 2019 Society of Actuaries

3.5 The Health and Retirement Study

Characteristic	Description
Sponsor	The study is supported by a cooperative agreement with the National Institute on Aging and funding from the Social Security Administration.
Managing Organization	University of Michigan
Purpose	The Health and Retirement Study (HRS) is a longitudinal study of older people that includes detailed economic and health information. The HRS study includes the separate study concepts of the two earliest studies that were merged in 1998. The combined studies, now deemed the HRS study, follow age-eligible individuals and their spouses as they make the transition from active worker into retirement, and they examine the dynamic interactions between health, family, and economic variables in the post-retirement period at the end of life. This study serves to support the development of policies and programs for aging Americans over the age of 50.
Coverage	The main part of the data is based on a 90-minute to three-hour core survey that takes place every two years. The sample size ranges from 18,000-23,000 for any given wave. Experimental modules (approximately ten per wave) are done at the end of the core survey with about 1,500 participants per module. Enhanced face-to-face interviews are done every other year per the timing described below in Frequency, and supplemental studies linked to the core on a range of topics are done in the "off year" with sample sizes of 3,000-7,000.
Frequency	Data is gathered in even years. Periodically, new cohorts are added to the study. The Late Baby Boomers born between 1960-1965 are the newest cohort that was added in 2016. After an initial survey, participants are re-surveyed every two years. A more in-depth survey is done in four-year intervals. After the participant's initial survey, half of them do the more comprehensive survey the first time in two years and the other half four years after the initial survey.
Timing	Scheduled release dates do not appear to be published or fixed with respect to a survey cycle (aka wave). As examples, the most recent wave (2016-2017) released an early core file in July 2018. Final exit and post-exit files (version 1.0) were released in December 2018. In the next earlier wave (2014-2015), the final core file, version 2.0, was released in December 2017.
Data Source	Data is initially gathered for new cohorts from a representative sample of the non-institutionalized US population. If respondents subsequently transfer to an institution, they continue to be tracked. To allow independent analysis of key subgroups, African-Americans, Hispanics, and residents of the state of Florida were oversampled. Survey weights are provided to adjust for this.
Data Content	The core content includes data on health, health services, labor force, economic status, family structure, expectations, and other supplemental modules covering a wide range of topics. The following link shows types of data collected through 2015. Each of the data programs shown has its own link that gives an overview of the study, a description of its data, and the associated questionnaire used to gather the information: https://hrs.isr.umich.edu/data-

© 2019 Society of Actuaries

	products/collection-path?_ga=2.94268381.1968439725.1549564241-1733566983.1548799743.
Data Availability	Publicly released HRS data is available for free to researchers and analysts. Access is conditional on registering with HRS.
Data Access	**Public Data** access requires a simple online registration. The bulk of the data is from the core of the biennial core survey. As an example, this link describes the content and structure of the 2014 core data: http://hrsonline.isr.umich.edu/modules/meta/2014/core/desc/h14dd.pdf. There are also several "off year" studies on a variety of topics. Each one of those has its own data and package of materials to interpret it. An example of this is the Consumption and Activities Study (CAMS), which has been done on an ongoing basis and, since 2001, on a biennial basis. This is a link to the 2017 CAMS data description: http://hrsonline.isr.umich.edu/modules/meta/2017/cams/desc/2017CAMS_DD.pdf. **Non-Public Data** access requires a supplemental registration process. These studies involve sensitive health data. Example of this are the 2016 venous blood study, biennial biomarker data since 2006, and various genetic material releases. This is the link to a brief data description of the 2014 biomarker data: http://hrsonline.isr.umich.edu/modules/meta/bio2014/desc/Biomarker2014DD.pdf. Non-public data also includes restricted data fields within otherwise pubic data, e.g. geographic information, birthdate, and date of death. A third form of data involves administrative linkages of HRS data to outside agencies. Access to this data also requires the supplemental registration with HRS and, in some cases, additional layers of approval outside of HRS to access the data. HRS data is linked with data from the Centers for Medicare and Medicaid (CMS), the Social Security Administration, Veterans Affairs, and other miscellaneous external entities. This link lists the data sources: https://hrs.isr.umich.edu/data-products/restricted-data/available-products. HRS data files require SAS, SPSS, or STATA to be utilized. Generally, all public data can be downloaded. Restricted data may require access through HRS systems or licensing the data if it is processed outside of HRS systems.
Allowable Use	By receiving the data, which have been freely provided, you agree to use them for research and statistical purposes only and to make no effort to identify the respondents. In addition, you agree to send us a copy of any publications you produce based on the data. By registering with HRS, you agree to the Conditions of Use governing access to Health and Retirement public release data. You must agree to: 1. not attempt to identify respondents, 2. not transfer data to third parties except as specified, 3. not share your username and password, 4. include specified citations in work based on HRS data, 5. provide information to us about publications based on HRS data, 6. report apparent errors in the HRS data or documentation files, and 7. notify us of changes in your contact information.
Other	The HRS database is very rich and complex. There are two entities that aggregate HRS data to create user-friendly data. The primary resource is the RAND Center for the Study of Aging. They have aggregated HRS data in ways that simplify its use but, per their caution, users

© 2019 Society of Actuaries

	should rely on data documents published by HRS for the use of their data to make effective use of the RAND version of it: https://www.rand.org/well-being/social-and-behavioral-policy/centers/aging/dataprod.html. Data is also aggregated by the University of Southern California Dornsife Center for Economic and Social Research (CESR). They have added expanded upon RAND work on HRS data by including some additional variables. Their work also extends to an international focus of studying aging in other countries where, similar to working with HRS, they have aggregated data from other primary sources.

© 2019 Society of Actuaries

About The Society of Actuaries

The Society of Actuaries (SOA), formed in 1949, is one of the largest actuarial professional organizations in the world dedicated to serving 30,000 actuarial members and the public in the United States, Canada and worldwide. In line with the SOA Vision Statement, actuaries act as business leaders who develop and use mathematical models to measure and manage risk in support of financial security for individuals, organizations and the public.

The SOA supports actuaries and advances knowledge through research and education. As part of its work, the SOA seeks to inform public policy development and public understanding through research. The SOA aspires to be a trusted source of objective, data-driven research and analysis with an actuarial perspective for its members, industry, policymakers and the public. This distinct perspective comes from the SOA as an association of actuaries, who have a rigorous formal education and direct experience as practitioners as they perform applied research. The SOA also welcomes the opportunity to partner with other organizations in our work where appropriate.

The SOA has a history of working with public policymakers and regulators in developing historical experience studies and projection techniques as well as individual reports on health care, retirement, and other topics. The SOA's research is intended to aid the work of policymakers and regulators and follow certain core principles:

Objectivity: The SOA's research informs and provides analysis that can be relied upon by other individuals or organizations involved in public policy discussions. The SOA does not take advocacy positions or lobby specific policy proposals.

Quality: The SOA aspires to the highest ethical and quality standards in all of its research and analysis. Our research process is overseen by experienced actuaries and non-actuaries from a range of industry sectors and organizations. A rigorous peer-review process ensures the quality and integrity of our work.

Relevance: The SOA provides timely research on public policy issues. Our research advances actuarial knowledge while providing critical insights on key policy issues, and thereby provides value to stakeholders and decision makers.

Quantification: The SOA leverages the diverse skill sets of actuaries to provide research and findings that are driven by the best available data and methods. Actuaries use detailed modeling to analyze financial risk and provide distinct insight and quantification. Further, actuarial standards require transparency and the disclosure of the assumptions and analytic approach underlying the work.

Society of Actuaries
475 N. Martingale Road, Suite 600
Schaumburg, Illinois 60173
www.SOA.org

© 2019 Society of Actuaries