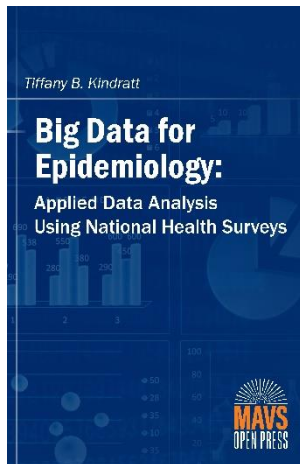# Chapter 2: Overview of National Health Surveys Lecture Slides Transcript

Slide 1

Hi everyone, my name is Tiffany Kindratt. This presentation covers "Chapter 2, Overview of National Health Surveys," in the textbook Big Data for Epidemiology: Applied Data Analysis using National Health Surveys.

Slide 2

Here is an outline of Chapter 2. The following sections will be covered in this presentation. Section 2.1 covers an introduction to the chapter. Section 2.2 covers the history of national health surveys. Section 2.3 covers an overview of the National Health Interview Survey or NHIS. Section 2.4 covers an overview of the Medical Expenditure Panel Survey or MEPS. Section 2.5 covers an overview of the Health Information National Trends Survey or HINTS. Section 2.6 covers an overview of the Behavioral Risk Factor Surveillance System or BRFSS. Section 2.7 covers an overview of the National Health and Nutrition Examination Survey or NHANES. Section 2.8 covers restricted data. Section 2.9 covers linked data. Section 2.10 provides a summary of the chapter. Section 2.11 includes the references.

Slide 3

Section 2.1 Introduction. This chapter provides an overview of the US national health surveys covered in the textbook. The surveys were just mentioned in the previous slide. It will include the NHIS, the MEPS, the HINTS, the BRFSS and NHANES. There will be a research case study with data that are publicly available from each national health survey used to answer research questions pertaining to how predisposing and enabling factors of individuals are associated with morbidity, mortality and health services use. An example of the predisposing factors that will be evaluated as part of the research case studies in this textbook are race and ethnicity. Examples of the enabling factors that are used in this textbook as independent variables are those describing interactions between health care providers and patients.

Slide 4

Section 2.2 History of National Health Surveys. National health surveys have been used in the United States since the 1920s. One of the first efforts came from the Committee on the Cost of Medical Care Studies, or CMCS, who collected data from 1928-1933. They received funding from private organizations to collect information on health care delivery and payments in 17 states and Washington, D.C. As one of the first studies to collect large representative data, results from these studies demonstrated that 40 percent of medical care costs were incurred by 10 percent of the families

surveyed within these states. A major limitation of this effort was that it did not collect data from underrepresented groups, particularly black or African American families. From 1935-1936, another national survey was conducted by the Public Health Service. The focus of this survey was to measure the incidence, or new cases, of illness and use of medical services. This survey used a multi-stage area sampling design with a larger sample size. It collected data from 21 states. Both of these surveys provided data into the 1950s. Then in 1953, the Health Information Foundation in New York and the National Opinion Research Center in Chicago collaborated to develop the first survey using a nationwide probability sample of all of the states. This laid the groundwork for national surveillance systems such as the NHIS, which is in practice today.

Slide 5

Section 2.3 National Health Interview Survey. The National Health Interview Survey, or NHIS, has been collected since 1957. The sponsoring agency for the survey is the National Center for Health Statistics. The study design is cross-sectional. Participants that answer survey questions during one year are not followed up in the future to see if their health behavior changes. The data collection frequency is annual. The data collection methods include a computer-assisted personnel interviewing system used during in-person interviews. Research personnel go to participants' homes and interview them with a computer system. Some questions are asked verbally while other data are obtained by asking participants to fill out their responses directly on the computer. The purpose of the NHIS is to monitor and explore trends in the health status and health care utilization of adults and children in the US.

Slide 6

The sample size is roughly 35,000 households and 87,500 individuals surveyed each year. The response rates can range anywhere from roughly 60 percent to a little over 70 percent. For example, in 2014, the response rate was 73.8 percent and in 2018, the response rate 64.2 percent. For more information on the design, questionnaires, public use data, and reports that are available, you can go to the website provided on this slide.

Slide 7

Section 2.4 Medical Expenditure Panel Survey. The Medical Expenditure Panel Survey, or MEPS, has collected data since 1996. The sponsoring agency is the Agency for Health Care Research and Quality, which is often shortened to "ARC" or AHRQ. The study design can be cross-sectional or longitudinal as a retrospective cohort. The data collection frequency is a little bit different than the NHS because it includes five different panels from one person over a two-year period. Each person has data collected from them five times over two years. Another interesting fact about the MEPS is that the sample is a subsample of the previous year's NHIS, Data are collected using a computer assisted personal interviewing system, like the NHIS, which includes in-person interviews among individuals' households. A unique feature of the MEPS is that it also includes self-administered questionnaires. Individuals are asked to fill out paper

surveys and mail them back to AHRQ. It also includes information collected from medical providers directly by telephone. The purpose of the MEPS is to gather information on health services used by adults and children in the US, including cost, frequency and payment structures.

Slide 8

The sample size of the MEPS is smaller than many other surveys. As I mentioned before, it is collected from a subsample of households who participated in the previous years' NHIS. The sample includes roughly 15,000 individuals per year. The response rates are also slightly lower. In 2014, the response rate for the MEPS was 48.5 percent and in 2018 the response rate was 42.7 percent. More information on design, questionnaires, public-use data, and reports are available from the website provided.

Slide 9

Section 2.5 Health Information National Trends Survey. The HINTS has been collected since 2003. The sponsoring agency is the National Cancer Institute. The study design is cross-sectional. HINTS collects data almost annually. So far, there have been 5 iterations of the HINTS and 4 annual cycles. For example, the data started being collected in 2003. Then they waited 2 years and the next iteration was collected in 2005. Then weighted to collected data until 2008. Starting in 2011, they collected data from 4 one-year cycles. There was a small break as well, and then HINTS 5 data were not collected until 2017. HINTS data have also been collected using different data collection methods in comparison to the NHIS or the MEPS. HINTS has collected data using random digit dialing, mailing, and online methods. The purpose of the HINTS is to evaluate how patterns of health information technology and health communication are related to health-related knowledge, attitudes, and behaviors among the US adult population.

Slide 10

The sample size for the HINTS is much smaller than the NHIS and MEPS. Data are collected from roughly 3,500 to 6,000 individuals per iteration. Response rates are lower as well at roughly 30 percent. More information on design, questionnaires, public-use data, and reports are available from the website provided.

Slide 11

Section 6.6. Behavioral Risk Factor Surveillance System. The BRFSS has been collected since 1984. The sponsoring agency for the BRFSS is the Center for Disease Control and Prevention, or the CDC. The study design is cross-sectional. Data are collected annually. The data collection method differs from other surveys. Instead of going to individuals' homes and collecting data using a computer-assisted personal interviewing system, they use a computer-assisted *telephone* interviewing system. The focus of the data collection has always been to provide state level estimates. However, in 1993 it was expanded to produce national estimates. You can use the BRFSS for

state-specific information as well as looking at nationally specific information. It allows for a collection of optional modules. However, the optional modules are specifically selected for each state. For example, the BRFSS has collected information on health literacy, but it's not been collected by all states so only state level comparisons can be made. Also, the BRFSS allows for adding questions on race and ethnicity. So, for example, the state of Michigan collects information on Arab Americans as an ethnicity since it has one of the highest concentrations of Arab Americans in Detroit. However, other states do not. The purpose of the BRFSS is to collect data on health behaviors, physical activity, diet, hypertension and preventive safety measures such as seat-belt use among US adults. If you are interested in research questions evaluating risk factors and behaviors, this may be a good survey to use for secondary data analysis.

Slide 12

The sample size is actually much larger than the other surveys. They survey roughly 400,000 individuals each year. The overall response rate 49.4% in 2019, which when you break it down further, the landline response rate was 53.3 percent and the cell phone response rate was 43.4 percent in 2018. More information on design, questionnaires, public-use data, and reports are available from the website provided.

Slide 13

Section 2.7 National Health and Nutrition Examination Survey. The NHANES has been collected since 1999. However, its content has been collected just not specifically listed as the NHANES since 1960. The sponsoring agency is the National Center for Health Statistics, who is also the agency who collects data using the National Health Interview Survey. The study design is cross-sectional, and the data collection frequency is biennial. So instead of just collecting data every year, it collects data across a two-year period. The data collection methods are different from other surveys because they actually collect subjective as well as objective data. They collect self-reported data using in-person household interviews, as well as objective measurements by physical examination and laboratory tests in mobile examination centers. If your research question is focused on diabetes, there are data describing whether someone self-reports that they have ever been diagnosed with diabetes and those individuals may also be asked to give blood to test whether or not they actually have diabetes. The purpose is to collect data on the health and diet of individuals in the US.

Slide 14

The sample size is roughly 5,000 individuals per data collection period. The interviewed response rate is roughly 51.9% percent. The examination response rate is roughly 48.8 percent. These two findings are from the 2017-2018 iteration. More information on design, questionnaires, public-use data, and reports are available from the website provided.

Slide 15

Here is a timeline of the national health surveys that were covered in this textbook based on the year that they began collecting data. In 1957, we have the NHIS. In 1984, the BRFSS. In 1996, the MEPS. In 1999, the NHANES and in 2003, the HINTS.

Slide 16

Section 2.8 Restricted Data. Any data that may compromise the confidentiality of the survey's participants are considered restricted data. These data require special procedures and approvals in order to access them. Both the NHANES and the NHIS offer restricted data accessible by the National Center of Health Statistics and Federal Statistical Research Data Centers. These data sources can be analyzed at different data centers across the US, including a data center that I analyze data at, the Dallas Fort Worth Federal Statistical Research Data Center. For example, the NHANES has restrictions on variables related to geography. Specific census blocks, genetics, as well as dates of participants' interviews and examinations are restricted.

Slide 17

The NHIS has restrictions on geography as well, including for those born in the US their state of birth. Whether or not a participant lives in an urban or rural residency is restricted. Among foreign-born individuals, their country of birth and year the participant came to the US are also restricted. My colleagues and I have previously conducted research on Arab Americans looking at the country of birth restricted variable in the National Health Interview Survey. There is a region of birth variable that is publicly available called REGIONBR. This variable combines all the information from participants based on the country of birth that was collected during the in-person interviews. The Middle East region includes individuals born in 25 countries located within the Middle East. However, countries such as Iran are located in the Middle East but are not part of the Arab League of Nations. So, in order to make inferences about Arab American health specifically using this data, we would be including "non-Arab" individuals in that category if we use the region of birth variable.

Slide 18

So, for 4 studies, we obtained access to analyze data that were restricted using the country of birth variable and created a new variable that comprised of 15 countries from the Arab League of Nations that were also geographically located in the Middle Eastern region. The studies evaluated chronic disease prevalence, preventive cancer screenings and vaccinations among men and women and smoking status.

Slide 19

Section 2.9 Linked Data. Another feature of these national health surveys is that you can link them to other data sources. Both the NHANES and the NHIS can be linked to data from the National Death Index in order to determine differences in mortality based on characteristics of these surveys that are not normally included in the National Death

Index. You can also link the MEPS to the NHS. Because the MEPS is a subsample of the previous year's NHIS sample, you can easily link them at a restricted data center. An example here comes from a study that my colleague conducted lining NHANES III data collected from 1988-1994 with 2015 National Death Index mortality data. Another example is from study that I'm currently working on, which is linking 2001-2018 MEPS data with 2000-2017 NHIS data to examine cognitive health outcomes among Arab Americans, specifically Middle Eastern or North Africans.

Slide 20

Section 2.10 Other Surveys. The surveys that I have presented in this textbook are not comprehensive. There are several other nationally representative surveys and I have listed here that you can review on your own time. The surveys I have included are the American Community Survey, Health and Retirement Survey, National Death Index, National Health and Aging Trends Survey, National Longitudinal Study of Adolescent to Adult Health, National Vital Statistics System, National Survey of Family Growth and the Youth Risk Factor Surveillance System. These surveys have specific content that may be of interest to you in the future when creating your own research questions. The surveys include cross-sectional as well as longitudinal data collection methods.

Slide 21

Section 2.11 Summary. So, in summary, this chapter provided an overview of the national health surveys covered in this textbook. It also included some details of some expanded data analysis procedures, including restricted and linked data. I provided a few examples of some other national health surveys that may be of interest to you in the future. More specific details on each of the surveys that are used in this textbook are provided in Chapter 6 through 10.

Slides 22-23

Section 2.12 References. The next two slides include the references that were used for creation of this chapter.

Slide 24

That concludes my presentation on "Chapter 2: Overview of National Health Surveys" in the textbook Big Data for Epidemiology: Applied Data Analysis using National Health Service. Thank you for listening.