Comparison of National Data Sources to Assess Preventive Care in the US Population

Glen B. Taksler, PhD[1,2,3] (iD), Elizabeth R. Pfoh, PhD, MPH[1], Kathryn A. Martinez, PhD, MPH[1], Megan M. Sheehan, BS[4], Niyati M. Gupta, MD[5], and Michael B. Rothberg, MD, MPH[1]

[1]Center for Value-Based Care Research, Cleveland Clinic Community Care, Cleveland Clinic, Cleveland, OH, USA; [2]Department of Quantitative Health Sciences, Cleveland Clinic, Cleveland, OH, USA; [3]Center for Health Care Research and Policy, Population Health Research Institute, Case Western Reserve University at The MetroHealth System, Cleveland, OH, USA; [4]Cleveland Clinic Lerner College of Medicine, Cleveland, OH, USA; [5]Department of Medicine, Maimonides Medical Center, Brooklyn, NY, USA.

Each year, nearly half of US deaths are attributable to preventable risk factors.[1–3] Recent work suggests that optimal use of preventive care would add >2 million healthy life-years to the US population.[4, 5] Although numerous public health resources exist to facilitate assessment of the health of the US population, including preventive care, each differs in its design, strengths, and limitations. The Centers for Disease Control and Prevention (CDC) conducts 6 nationally representative surveys,[6–9] many of which seem similar at first glance. For example, the purpose of the National Health Interview Survey (NHIS), "[T]o monitor the health of the United States population through the collection and analysis of data on a broad range of health topics,"[7] is nearly identical to that of the National Health and Nutrition Examination Survey (NHANES), "[T]o assess the health and nutritional status of adults and children in the United States."[6] Additionally, the Agency for Healthcare Research and Quality and higher educational institutions conduct at least 3 nationally representative surveys.[10–12] Researchers interested in these databases need to devote substantial time to identify the most appropriate resource.

In this study, we compared nationally representative, publicly available surveys with preventive care data. Our purpose was to aid investigators to quickly identify the most appropriate survey for specific research questions.

METHODS

We searched 9 national, open-access surveys for information on preventive services rated A or B by the US Preventive Services Task Force (USPSTF) or recommended by the Advisory Committee on Immunization Practices (ACIP) for adults aged ≥18 years.[13, 14] Although each study collects data on a broad range of medicine and public health topics, our focus was preventive care. Specifically, we considered the National Health Interview Survey (NHIS), intended to assess overall health of the US population;[7] National Health and Nutrition Examination Survey (NHANES), broadly similar in purpose to NHIS but with an in-person examination component;[6] Medical Expenditure Panel Survey (MEPS), a longitudinal subset of households participating in the NHIS;[10] Behavioral Risk Factor Surveillance System (BRFSS) and BRFSS-Selected Metropolitan/Micropolitan Area Risk Trends (BRFSS-SMART), which collect state- and local-level data, respectively;[8, 15] National Ambulatory Medical Care Survey (NAMCS) and National Hospital Ambulatory Medical Care Survey (NHAMCS), which collect data on provision and use of ambulatory and hospital services, respectively;[9] and 2 social sciences surveys: the Health and Retirement Study (HRS), intended to capture the retirement, savings, and health of Americans aged >50 years;[11] and the Panel Study of Income Dynamics (PSID), intended to provide data on income, wealth, labor force characteristics.[12] Although the NHAMCS is a hospital-based survey, its intention is to document use of ambulatory care services in emergency and outpatient departments; therefore, preventive care topics were included. We did not consider the Longitudinal Study of Aging (LSOA), a cohort study of adults aged >70 years, which last collected data in 2000.[16]

We reviewed study design, sampling techniques and size, data collection method (e.g., telephone, in-person), file organization (e.g., one large file vs. many small files), and restricted access components. Then, we reviewed documentation for 2016–2018 to find relevant questions on preventive care. We also included the NHIS Cancer Control Supplement from 2015 because of extensive questions on cancer screening. (Although survey questions differ over time, each has a stable "core" component of questions asked for many years.

Appendix Table 1 provides links to websites with more information on historical data availability.) We reviewed the following areas: diagnosis of major cardiovascular conditions (hypertension, hyperlipidemia, diabetes), medical examination (physical exam, laboratories, and imaging), medications, screenings (e.g., cancer, osteoporosis), lifestyle (weight, tobacco, alcohol, illicit drug use, nutrition, physical activity, sexual behavior, sexually transmitted infections and sleep), mental health, vaccinations, family history, and demographics.

To aid investigators in determining which survey(s) best met their needs, we summarized similarities and differences between surveys, counted the number of relevant preventive care topics within each survey (and supplement), and identified if a survey addressed a particular topic. To count topics, 2 members of the study team independently identified the number of topics and then met to reconcile their results. The number of topics was determined using 2 rules. First, topics had to be distinct. Questions that asked essentially the same information in different ways were grouped. For example, "What year did you stop smoking" and "How old were you when you stopped smoking?" were counted as 1 topic. Second, to avoid overcounting, for topics with an initial question and follow-up questions conditional on the response (e.g., follow-up if a participant responded "Yes"), we only counted follow-up questions. For example, the initial question "During the past 7 days, did you walk to get some place that took you at least 10 minutes?" was not counted but the follow-up question "In the past 7 days, how many times did you do that?" was counted.

RESULTS

Tables 1 and 2 summarize the surveys, and Table 3 provides detail on the number of preventive care topics within each survey. (The online Supplement lists specific questions, Appendix Tables 2–25.) Surveys substantially differed in design and data elements.

Design

Six surveys are cross-sectional and 3 are longitudinal. MEPS follows subjects every 6 months for 2.5 years, HRS follows subjects biennially until death, and PSID (an intergenerational survey) follows subjects and their offspring annually until death.

Sampling Unit. Seven surveys randomly sample US households and individuals within households. NAMCS and NHAMCS sample providers at community practices or hospitals, respectively, for data on specific encounters. For PSID, random sampling occurred at survey inception in 1967.

Oversampling. HRS, NHANES, NHIS, MEPS, and PSID oversample select minorities, typically Black race and

Hispanic ethnicity. PSID fixed the oversampling proportion in 1968 (not updated annually). BRFSS/BRFSS-SMART oversample sub-state (e.g., rural) regions.

Size. Sample size ranges from about 6000 to 10,000 (HRS, NHANES) to nearly 500,000 individuals (BRFSS) per survey period. NAMCS, NHAMCS, MEPS, and PSID each have a sample size of about 20,000–30,000 individuals/period, and NHIS about 100,000.

Shared Questions. All BRFSS-SMART questions are included in BRFSS (results combined below as "BRFSS/BRFSS-SMART"). NAMCS and NHAMCS have many shared questions.

Representativeness. All surveys are intended to provide national estimates; BRFSS also provides state-level data and BRFSS-SMART provides estimates for selected urban areas.

Major Cardiovascular Conditions

All surveys ask about hypertension, hyperlipidemia, diabetes, and other chronic conditions. NHANES provides the most detail, asking up to 10 topics on hypertension and hyperlipidemia, and 20 topics on diabetes. NHANES distinguishes itself through extensive medical examination (below) and recording names and doses of medications. For researchers not requiring this level of detail, NHIS still provides substantial information; for example, on hypertension, known diagnosis, time since diagnosis, use of prescription medication (yes/no), and resulting limitations on activity (yes/no). Although less detailed overall, HRS was the only survey to ask about medication adherence and reasons for nonadherence, such as cost, insurance coverage, and side effects. For researchers requiring state/local-level data, a 2017 BRFSS/BRFSS-SMART supplement provides greater detail ("modules" asked in selected states).

All surveys ask about other major medical conditions, with NHANES and NHIS providing the most detail (>30 major medical conditions such as asthma, arthritis, bronchitis, heart disease, and liver disease).

Medical Examination

NHANES and HRS perform a physical examination including vital signs and body mass index and collect laboratories relevant to preventive care (e.g., lipids, HbA1c). In NHANES, >95% of subjects undergo the exam which includes 4 blood pressure measurements, pulse, weight, height, and other body measurements (e.g., waist circumference), and depending on the year, a dental exam, hearing exam, vision test, lung function, muscle strength, and DEXA scan. Laboratories are extensive, including complete blood count, vitamins and minerals, hepatitis, HIV, inflammatory markers, pregnancy testing, sexually transmitted infections, and sugars. Depending on

Table 1 Overview of Nationally Representative Surveys of the US Population

Survey	Sponsor	Sample size (year)	Years	Frequency	Goal
BRFSS, BRFSS-SMART	CDC	BRFSS: 486,303 (2016) BRFSS-SMART: 249,011 (2016)	BRFSS: 1984–1992 (15 states), 1993–present (nationwide) BRFSS-SMART: 1992–present	Annual	BRFSS: Collect state-level data BRFSS-SMART: Collect Metropolitan and Micropolitan Statistical Area (MMSA) data
HRS	University of Michigan, NIA, SSA	6624 (2014)	1992–present	Biennial	Collect data on retirement, disability, and the economic and health impact of aging
NAMCS	NCHS of CDC	1737 physicians and 28,332 patient visits (2015)	1973–1981, 1985, 1989 –present	Annual since 1989	Collect data on ambulatory medical care services
NHAMCS	NCHS of CDC	267 hospitals and 21,061 emergency department visits (2015)	1992–present	Annual	Collect data on use of medical care services in hospital emergency and outpatient departments and ambulatory surgery centers
NHANES	NCHS of CDC	9971 interviewed, 9544 medically examined (2015–16)	1960–present with 8 major redesigns; 1999–present for current design ("continuous NHANES")	Biennial, 1999 –present; frequency varied prior to 1999	Collect data on health and nutritional status of adults and children
NHIS	NCHS of CDC	40,220 household, 97,169 persons (2016)	1957–present (redesigned every 10–15 years); most recent design began in 2019 (data not yet released)	Annual	Collect data on health, burden of illness, and disability
MEPS	AHRQ	13,800 families, 33,893 individuals (2015)	1996–present	Annual	Collect data on health care utilization, cost, and health insurance
PSID	University of Michigan, NSF, NIA, NICHHD	9048 families, 24,637 individuals (2015)	1968–present	Annual	Collect intergenerational data on economic, social, and health factors over-the-life course

AHRQ, Agency for Healthcare Research and Quality; CDC, Centers for Disease Control and Prevention; MMSA, Metropolitan and Micropolitan Statistical Area; NCHS, National Center for Health Statistics; NIA, National Institute on Aging; NICHHD, National Institute of Child Health and Human Development; NSF, National Science Foundation; SSA, Social Security Administration

the survey year, one-quarter to one-third of subjects obtain a fasting lipids panel. In HRS, physical examination data are considered sensitive, requiring special permission.[17] Biomarkers were collected beginning in 2006 and generally included lipids, A1c, C-reactive protein, and cystatin C. In 2016 (not 2018), HRS laboratories were more extensive, adding complete blood count, comprehensive metabolic panel, b-type natriuretic peptide, and vitamins and minerals (more limited than for NHANES). These labs were requested at the end of the interview in exchange for a $50 incentive payment (approximately two-thirds of subjects consented) and were performed after the interview (typically within 4 weeks).

Medications

NHANES, NAMCS, NHAMCS, and MEPS verify prescription medications, although only NHANES and MEPS report dosage. (In 2007, HRS conducted a related Prescription Drug Study of participants.) NHANES also reports nonprescription antacids and dietary supplements. NHANES asks subjects to bring their prescription and over-the-counter medication bottles with them (for review by the interviewer) and MEPS contacts subjects' pharmacies. NAMCS and NHAMCS ask providers to report up to 30 medications (prescription or over-the-counter, including dietary supplements) ordered or administered during an encounter, and to specify whether each medication was new or continued.

All surveys except NAMCS, NHAMCS, and HRS ask about low-dose aspirin. BRFSS, NHAMCS, NAMCS, NHIS, and MEPS ask about contraindications to aspirin.

Screenings

Cancer. All surveys ask about the history of cancer (helping to establish eligibility for screenings), and all except PSID ask about cancer screenings. General cancer history questions include whether a respondent currently has cancer or has ever been told they have cancer. BRFSS, HRS, NHIS, and PSID ask about cancer treatment. All surveys ask about breast, cervical, colorectal, and prostate cancer screenings except NHANES which only asks about cervical cancer. NHIS supplemental cancer control surveys (every 5 years; most recent in 2015) ask the most detail, ranging from 4 topics on prostate cancer screening to 22 topics on colorectal cancer screening, including colonoscopy, stool-based testing, flexible sigmoidoscopy, CT colonography, and outcome (e.g., polyps). On an annual basis, BRFSS asks the most detail, with 1-8 topics on each type of cancer screening. However, BRFSS only asks residents of selected states.

Osteoporosis. All surveys except NHIS and BRFSS ask about osteoporosis screening and diagnosis. NHANES included osteoporosis in earlier years (1999–2000 through 2009–2010, 2013–2014).

Table 2 Methods and File Organization of Nationally Representative Surveys of the US Population

Survey	Design	Oversampling	Method	File organization*
BRFSS, BRFSS-SMART	Cross-sectional sample of non-institutionalized US population	48 states sampled disproportionately from sub-state regions (to increase sample size for smaller geographically-defined populations of interest)	Telephone (including cellular since 2011), Internet. Data collected monthly in each state and territory.	BRFSS: 1 file BRFSS-SMART: County data, MMSA data
HRS	Longitudinal	Black race, Hispanic ethnicity, residents of Florida	Self or proxy reported. Initial interview in-person with telephone follow-up	HRS Core, HRS Exit (data on decedents), HRS Post-Exit (additional financial information on decedents), Supplemental studies
NAMCS	Cross-sectional sample of private practices and nonfederal community health centers	Selected states (2012−2015 only); otherwise, area-based sampling (national, 4 census regions, 9 census divisions)	Paper, electronic (since 2012)	1 file
NHAMCS	Cross-sectional sample of ambulatory care utilization in nonfederal hospital outpatient departments, emergency departments, and ambulatory surgery centers	Selected states (2012−2015 only); otherwise, area-based sampling (national, 4 Census regions, 9 Census Divisions)	Paper, electronic (since 2012)	1 file
NHANES	Cross-sectional sample of civilian, non-institutionalized US population	Hispanic ethnicity (specifically Mexican-American for 1999−2006), Black race, income <130% of federal poverty level; ages 12−19 y (1999−2006), ≥70 y (1997−2006), ≥80 y (2007−present)	In-person, computer-assisted, audio computer-assisted	Demographics, Dietary, Examination, Laboratory, Questionnaire, Limited Access
NHIS	Cross-sectional, household sample of civilian, non-institutionalized households, one adult, and one child from each household sampled for more information	None at household level. Adults aged ≥65 y and Black race, Asian race, and Hispanic ethnicity at adult within household level.	In-person, computer-assisted	Household, Family, Sample Adult/Child, Injury Episode, Adult/Child Alternative Medicine, Adult Cancer, Child Immunization
MEPS	Longitudinal subsample of households participating in NHIS.	None at household level. Adults aged ≥65 y and Black race, Asian race, and Hispanic ethnicity at adult within household level.	Computer-assisted	Household component, insurance component
PSID	Longitudinal cohort, genealogic design (offspring of survey participants are invited to participate)	Low-income families and African American populations. Proportion of oversampling fixed as of 1968, rather than updated each year.	Internet, paper as backup format	Main interview, child development supplement (CDS), transition into adulthood supplement (TAS), disability, and use of time supplement (DUST)

Currently, all surveys offer sample SAS, SPSS, and Stata code. Historical data prior to the 1990s/early 2000s typically offered sample SAS code or only a PDF documentation manual

Lifestyle

All surveys ask about lifestyle including weight, healthy diet, physical activity, smoking, and alcohol use. Five surveys ask about sleep and 6 about sexual health.

Weight. NHANES provides the most detail about weight (18 topics) including self-reported weight history, attempts to lose weight, and methods (e.g., diet, exercise, medication, surgery). The next most detailed is NHIS, which asks about current weight, duration of any weight problem, and whether weight interferes with daily activities. BRFSS also asks subjects if a health care provider has ever counseled them to lose weight.

Diet. NHANES contains detailed nutrition information (14 topics), including a 2-day dietary interview and nutritional analysis. However, every 5th year, the NHIS asks a 26-item dietary screener about food/beverage consumption over the past month; these questions are almost identical to NHANES.[18] MEPS asks if subjects have received advice on healthy eating from a doctor.

Physical Activity. NHANES includes 16 topics, focusing on respondents' knowledge of guidelines (e.g., "[H]ow many minutes [do you] think you should exercise?" and exertion (light/moderate/vigorous)). HRS addresses 4 topics, including mild, moderate, and vigorous activity.

Tobacco. All surveys ask about current smoking; NHANES and NHIS also ask about years of smoking, quit attempts/former smoking, and use of other tobacco products (including e-cigarettes). NHIS, MEPS, and NAMCS/NHAMCS identify whether the respondent (or patient) received tobacco cessation counseling.

Alcohol. BRFSS, HRS, NHANES, NHIS, and PSID ask about the quantity of alcohol consumption. HRS inquires about

Table 3 Number of Preventive Care Topics by Survey in 2017−2018

Reporting	BRFSS and BRFSS-SMART*	HRS	NAMCS	NHAMCS	NHANES	NHIS	MEPS	PSID
	Self	Self	Provider	Provider	Self	Self	Self/insurance	Self
Design	Cross-sectional	Longitudinal	Cross-sectional	Cross-sectional	Cross-sectional	Cross-sectional	Longitudinal	Longitudinal
Number of preventive care topics in core questionnaire (parentheses indicate topics in supplements)								
Major cardiovascular conditions								
Hypertension	0	3	3	3	10	6	4	4
Hyperlipidemia	0	3	5	5	10	4	3	0
Diabetes	8 (9)	6	5	5	20	8 (9)	3	3
Screenings								
Breast cancer	4	1	1	1	0	6	2	0
Cervical cancer	7	1	7	7	4	7 (12)	4	0
Colorectal cancer	4	2	2	2	1	13 (22)	4	0
Lung cancer	1	0	0	0	0	0 (5)	0	0
Prostate cancer	8	1	2	2	0	6 (4)	3	0
Personal history of cancer†	12	7	1	1	3	3 (25)	2	4
Osteoporosis	0	2	2	2	13	0	2	0
Lifestyle								
Exercise	1	4	1	1	16	3 (4)	1	4
Nutrition	5 (2)	0	1	1	14	4 (26)	0	5
Weight	2	4	2	2	18	5	3	4
Sleep	5	13	0	0	8	5	0	0
Mental health	5 (4)	12	3	3	11‡	18	7	6
Sexual history/health	2	0	3	3	47	3	0	0
Tobacco	5	6	3	3	34	12 (29)	7	9
Alcohol	5	8	3	3	12	7	3	5
Drugs	3	0	2	2	23	2	0	2
Vaccinations								
Influenza	3	1	0	0	0	5	4	0
Pneumonia	1	1	0	0	0	2	1	0
Tetanus	2 (1)	0	0	0	0	2	0	0
Zoster	1 (1)	2	0	0	0	6	1	0
Family history								
Family history	0	1	0	0	5¶	1 (19)	0	0

We counted the number of topics related to preventive care in 9 national surveys. Topics assessed during medical examination (physical, labs, and medications) were categorized by disease and not separately reported. Parentheses indicate topics asked in the NIHS Cancer Control Supplement (2015) or the BRFSS diabetes module (2018, only asked of participants in selected states)
*BRFSS and BRFSS-SMART questions were identical
†Included to establish screening eligibility
‡The 1 topic is the validated Patient Health Questionnaire-9 for depression
¶NHANES also asks about family history of asthma and angina, excluded from the topic count because of a focus on preventive care

negative feelings associated with alcohol, and BRFSS includes questions on counseling. NHAMCS/NAMCS ask whether the visit is related to alcohol use.

Sleep. HRS addresses most topics including trouble sleeping, feeling rested, and medical conditions associated with sleep. BRFSS/BRFSS-SMART and NHIS ask about hours of sleep, schedule (NHANES), snoring (BRFSS/BRFSS-SMART, NHANES), and sleep medications (NHIS).

Sexual Health. NHANES includes 47 topics and asks about partners of both genders. BRFSS/BRFSS-SMART and NHIS also ask about sexual behavior (only 2–3 topics each). NAMCS and NHAMCS ask whether the patient ever has been diagnosed with HIV or hepatitis B or C.

Mental Health

All surveys ask about mental health. NHANES and BRFSS/BRFSS-SMART include either the validated Patient Health Questionnaire-9 for depression or its component questions.[19, 20] HRS and MEPS ask about depression symptoms during the prior 12 months or 4 weeks, respectively. NAMCS and NHAMCS document attention-deficit disorder and depression diagnosis. NHIS and PSID ask about developmental disorders and PSID about memory loss and mentally demanding activities.

Vaccinations

BRFSS, HRS, and NHIS ask about influenza, pneumonia, tetanus, and shingles vaccines, with NHIS covering the most topics. MEPS only asks about influenza and shingles vaccines.

Family History

The NHIS Cancer Control Supplement asks about family cancer history of cancer (19 topics) including the type of cancer, specific relatives, and age <50 years at diagnosis. NHANES asks about family history of diabetes, asthma, and angina.

Specific USPSTF and ACIP Recommendations

Biennially beginning in 2018, MEPS includes questions about receipt of 15 USPSTF- or ACIP-recommended services, following exact guideline recommendations (e.g., had blood pressure checked during past 24 months, without regard to hypertension control).[21]

Demographics

All surveys include demographics, such as age, sex, race, ethnicity, income, and education. HRS and PSID provide the most economic detail (e.g., income, employment, savings, debts).

Health Care Access and Utilization

MEPS and HRS provide the most detail on access to care, such as subjects' usual source of care, specialist needs, language barriers (MEPS), and reasons for delaying medical care (HRS). All surveys ask about health insurance; HRS and PSID include detailed information about premiums. Owing to their encounter-based design, NAMCS and NHAMCS have the most detail on utilization, although NHANES, MEPS, NHIS, HRS, and PSID offer restricted linkages to administrative claims (described below).

Linkages to Other Data Sources

Surveys offer linkages to external data sources, but most require an application for restricted/sensitive data access. Typically, researchers must visit a facility of the National Center for Health Statistics (NCHS) (Washington, DC metropolitan area, or Atlanta, GA) or a Federal Statistical Restricted Data Center (nationwide).[22] (As of May 18, 2020, facilities are closed due to COVID-19.)[23] For NCHS-sponsored surveys, the minimum cost is $3000 plus travel expenses.[24] Costs for other surveys are available online and may depend on the amount of data required.[25–30] Table 4 summarizes restricted data linkages.

File Organization

For each survey cycle, NHANES and NHIS are each organized in 6 files. For NHANES, files are demographics (including sample weights), dietary, examination, laboratory, a questionnaire, and documentation of limited access (restricted) data. For NHIS, files are household, family, injury, sample adult, sample child, and supplemental surveys. By contrast, NAMCS, NHAMCS, and BRFSS/BRFSS-SMART are each condensed in 1 file per year. Complexity for HRS, PSID, and

MEPS is in-between. HRS data comprise 1 "core" file, 1 "exit" file for decedents, and 1 "post-exit" file for additional financial information on decedents. PSID is organized into 4 files but 1 file ("cross-year individual") is most relevant to preventive care. It contains 1 record per individual present in an interviewed family in any year. MEPS data are available in 1 household file which includes most variables, plus 9 "event" files detailing healthcare utilization. Three are relevant to preventive care: prescribed medicines, office-based medical provider visits, and outpatient visits.

All surveys include documentation, often with statistical software code. Typically, data are downloadable in SAS Transport or comma-delimited formats. Stata (College Station, TX) and R (Vienna, Austria) have commands to read SAS Transport data; R's *nhanesA* package can directly import NHANES data. Surveys also provide documentation on how to weight observations to obtain nationally representative (or for BRFSS/BRFSS-SMART, state-/locality specific) results.[31–36]

Statistical Briefs

For some researchers, previously published overviews of disease epidemiology and known risk factors may suffice. These reports are readily available online for NCHS- and CDC-sponsored surveys.[37–40] Additionally, HRS and PSID provide a list of peer-reviewed studies utilizing their datasets.[41, 42]

DISCUSSION

In this study, we reviewed 9 nationally representative surveys of relevance to preventive care researchers. Our work may assist researchers with identifying the most appropriate resource(s) to consider in future research, an otherwise difficult task given their extensive documentation.

Generally, NHANES offers the most detail on preventive care, with a comprehensive medical exam and questionnaire including detail on management of chronic conditions relevant to disease prevention. These strengths are somewhat offset by biennial frequency, small sample size (about 10,000/wave), limited historical comparisons owing to major revisions in 1999, and complex file organization. However, we anticipate that for many preventive care studies, NHANES will serve as the most valuable nationally representative data source.

Two exceptions are NHANES' limited data on cancer screenings (only asking about cervical cancer) and vaccines (only asking only about human papilloma virus). For these topics, NHIS offers the most detail, the Cancer Control Supplement which asks dozens of cancer screening questions. Historically, the supplement has been asked every 5th year but transitioned to annual beginning 2019, with rotating questions addressing each cancer type every 2–5 years. Yet, even away from these topics, researchers may benefit from consulting the NHIS questionnaire before choosing NHANES. Should NHIS's more limited questionnaire suffice, advantages

Table 4 Restricted Data and Linkages to External Data Sources Available from Nationally Representative Surveys of the US Population

	Exact Dates	Mortality	Health care utilization	Geography	Socioeconomic position
BRFSS, BRFSS-SMART	-	-	-	-	Industry and occupation*
HRS	Birthdate, interview date, death date	National Death Index, next-of-kin exit interviews	Veterans' Affairs (2360 participants from 1999 to 2013)	Zip code	Social security, food access, crime, air quality, Dartmouth Atlas of Health Care
NAMCS	Patient birth date, encounter date	-	Data collected in surveyed encounter (public access), number of patient visits to provider's practice in past 12 months (restricted access, 2007−present)	County	Provider and practice characteristics†
NHAMCS	Patient birth date, encounter date	-	Limited to data collected in surveyed encounter (public access)	County	Provider and hospital characteristics†
NHANES	Interview date, examination date	National Death Index‡	Medicare, Medicaid claims	Latitude/longitude; census block, block group, tract; county; state	Housing assistance
NHIS	Birth date, interview date	National Death Index‡	Medicare, Medicaid claims	Latitude/longitude; census block, block group, tract; county; state	Housing assistance
MEPS¶	Dates of visits/home health events, diagnosis and procedure codes	-	Health insurance, charges, utilization	Latitude/longitude; census block, block group, tract; county; state	-
PSID	Dates from Medicare claims data; death date	National Death Index, manual research (e.g., next-of-kin contact, obituary search)	Medicare claims	Latitude/longitude; census block, block group, tract; county; state	Housing assistance

No linkages are listed on the National Center for Health Statistics website; however, the National Institute for Occupational Safety and Health states that restricted data on industry and occupation are available since 2013

†For example, provider year of birth, sex, number of providers in the practice, and availability of on-site tests and procedures (e.g., mammography, colonoscopy)

‡NHANES and NHIS offer both public and restricted mortality linkages; the former contains broad cause-of-death (e.g., cardiovascular disease) while the latter contains exact dates and ICD-9/10 level cause-of-death

¶Additionally, MEPS offers a restricted linkage to the original NHIS questions which were asked of the same participants in the year prior to their initial MEPS participation

include annual frequency, breadth of historical data (1957–present), larger sample size (approx. 100,000/year), and simpler file organization.

Overall, both surveys offer more preventive care information than BRFSS and BRFSS-SMART, which are less thorough but offer state- and metropolitan-specific data, respectively. BRFSS offers an exceptional sample size (nearly 500,000/year), allowing improved power for subgroup analyses (e.g., minorities).

NAMCS and NHAMCS (annual surveys) are the only studies sampling ambulatory encounters, with physician subjects. They provide encounter-specific prescriptions, laboratories, examination, and health education/counseling, but more limited lifestyle and vaccination data than other surveys. Researchers interested in provider- and practice-level heterogeneity may benefit from covariates (e.g., practice size) available in restricted data.

One limitation of the above surveys is their cross-sectional design, which may be attenuated by combining survey waves. Longitudinal surveys offer substantially less preventive care data, but for researchers needing this design, HRS generally offers the most depth, except for cancer screenings. However,

MEPS and PSID offer 4–5 times the sample size of HRS (approx. 25,000–35,000 vs. 6600) and are conducted annually, versus biennially for HRS. Additionally, in 2018, MEPS began asking about the receipt of 15 preventive services, with wording that matches guideline recommendations. PSID offers the greatest historical data (1968–present), while HRS and PSID offer extensive detail on socioeconomic position.

As an example of the types of research questions most suited to particular surveys, consider diabetes. Several high-profile studies utilizing NHANES have assessed historical trends in diabetes prevalence and risk factors associated with diabetes-specific mortality (e.g., physical activity, healthy diet).[43–46] Although NHIS does not provide as much depth as NHANES, it allows for longer historical analysis; one study considered trends in diabetes prevalence and incidence since 1980.[47] BRFSS/BRFSS-SMART allow for state- and local-level analyses; studies have found the greater burden of diabetes and associated risk factors in the south and lower burden in the west.[48, 49] Studies with MEPS have been linked to costs, such as assessing the financial impact of missed workdays in patients with diabetes or measuring healthcare expenditures in multimorbid patients.[50, 51] Analyses with NAMCS have

considered provider factors; for example, finding that primary care providers provide a majority of outpatient care for patients with diabetes,[52] but that lifestyle management is addressed in just one-quarter of office-based encounters.[53] Studies with HRS, which in our opinion offers more in-depth analysis of diabetes than PSID, have considered the impact of sociodemographic factors on diabetes incidence and mortality, such as stress, marital quality, and food insecurity.[54–56] We note that for all surveys, it is possible to assess disparities in these topics by sex and race/ethnicity.

Limitations

Survey content changes over time, limiting generalizability beyond our 2016–2018 timeframe. However, questions often remain in place for years, facilitating comparison or pooling across survey cycles. Second, although two researchers reviewed the number of preventive care topics and resolved differences, different methods, such as the counting of follow-up questions, would yield different results. Our methods were intended to describe available resources rather than provide a meta-analysis or identify the "best" survey. Researchers may find it helpful to review specific questions presented in the Appendix in the Supplementary Information. Third, our focus was adult preventive services. Researchers interested in children may wish to consult documentation for the NHIS, NHANES, MEPS, and Youth BRFSS (YBRFSS). Finally, our study was specific to preventive care. The strengths and weaknesses of national surveys may differ for other aspects of primary care.

CONCLUSIONS

Understanding national surveys can help in choosing the most appropriate data source for preventive care research.

Corresponding Author: Glen B. Taksler, PhD; Center for Value-Based Care Research, Cleveland Clinic Community Care, Cleveland Clinic, Cleveland, OH, USA (e-mail: taksleg@ccf.org).

Declarations:

This study did not meet the definition of research including human subjects at the Cleveland Clinic Institutional Review Board.

REFERENCES

1. **Rhodes HG**, Committee on Population (Division of Behavioral and Social Sciences and Education), Board on Health Care Services (Institute of Medicine). *Measuring the Risks and Causes of Premature Death: Summary of Workshops.* Washington, DC: The National Academies Press; 2015.
2. **Mokdad AH, Marks JS, Stroup DF, Gerberding JL**. Actual causes of death in the United States, 2000. *JAMA.* 2004;291(10):1238-1245.
3. **McGinnis JM, Foege WH.** Actual causes of death in the United States. *JAMA.* 1993;270(18):2207-2212.
4. **Maciosek MV, LaFrance AB, Dehmer SP, et al.** Updated Priorities Among Effective Clinical Preventive Services. *Ann Fam Med.* 2017;15(1):14-22.
5. **Isham G, Sanchez E, Jones WA, Teutsch S, Woolf S, Haddix A**. Prevention Priorities: Guidance for Value-Driven Health Improvement. *Ann Fam Med.* 2017;15(1):6-8.
6. National Health and Nutrition Examination Survey. https://www.cdc.gov/nchs/nhanes/index.htm. Accessed February 25, 2019.
7. National Health Interview Survey. https://www.cdc.gov/nchs/nhis/index.htm. Accessed June 30, 2020.
8. Behavioral Risk Factor Surveillance System. https://www.cdc.gov/brfss/index.html. Accessed June 30, 2020.
9. Ambulatory Health Care Data. https://www.cdc.gov/nchs/ahcd/index.htm. Accessed February 25, 2019.
10. Medical Expenditure Panel Survey (MEPS). https://www.ahrq.gov/data/meps.html. Accessed February 25, 2019.
11. Health and Retirement Study. https://hrs.isr.umich.edu/data-products. Accessed February 25, 2019.
12. Panel Study of Income Dynamics. https://psidonline.isr.umich.edu/. Accessed February 25, 2019.
13. USPSTF A and B Recommendations. http://www.uspreventiveservicestaskforce.org/Page/Name/uspstf-a-and-b-recommendations/. Accessed May 15, 2020.
14. **Kim DK, Riley LE, Hunter P**. Advisory Committee on Immunization Practices Recommended Immunization Schedule for Adults Aged 19 Years or Older - United States, 2018. *MMWR Morb Mortal Wkly Rep.* 2018;67(5):158-160.
15. SMART: BRFSS City and County Data and Documentation. https://www.cdc.gov/brfss/smart/Smart_data.htm. Accessed February 25, 2019.
16. Longitudinal Studies of Aging. https://www.cdc.gov/nchs/lsoa/index.htm. Accessed February 25, 2019.
17. Sensitive Health Data. http://hrsonline.isr.umich.edu/index.php?p=healthdat. Accessed June 26, 2020.
18. Dietary Screener Questionnaire in the National Health Interview Survey Cancer Control Supplement 2015. https://epi.grants.cancer.gov/nhis/2015-screener/. Accessed June 26, 2020.
19. **Kroenke K, Spitzer RL, Williams JB**. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med.* 2001;16(9):606-613.
20. **Spitzer RL, Kroenke K, Williams JB**. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. *JAMA.* 1999;282(18):1737-1744.
21. **Borsky A, Mitchell E, Machlin S, et al.** *Use of Clinical Preventive Services in the United States: Estimates from the Medical Expenditure Panel Survey (MEPS), 2015. Research Findings #41.* Agency for Healthcare Research and Quality, Rockville, MD; May 2019.
22. Federal Statistical RDC. https://www.cdc.gov/rdc/b2accessmod/acs220.htm. Accessed June 26, 2020.
23. Research Data Center (RDC). https://www.cdc.gov/rdc/. Accessed June 26, 2020.
24. Location of Access. https://www.cdc.gov/rdc/b2accessmod/ACs210.htm. Accessed June 26, 2020.

25. HRS Restricted Data FAQs. https://hrs.isr.umich.edu/data-products/restricted-data/faqs. Accessed January 19, 2021.

26. Restricted NCHS Variables. https://www.cdc.gov/rdc/b1datatype/Dt122.htm. Accessed January 19, 2021.

27. Fees and Invoicing. https://www.cdc.gov/rdc/b5aprovproj/ap540.htm. Accessed January 19, 2021.

28. BRFSS Industry and Occupation (I&O) Data: How external researchers can request access. https://www.cdc.gov/niosh/topics/surveillance/brfss/access.html. Accessed January 19, 2021.

29. Restricted Data Files Available at the Data Centers. https://meps.ahrq.gov/data_stats/onsite_datacenter.jsp. Accessed January 19, 2021.

30. Restricted Use Data. https://psidonline.isr.umich.edu/guide/Brochures/sens-data.pdf. Accessed January 19, 2021.

31. NAMCS Estimation Procedures. https://www.cdc.gov/nchs/ahcd/ahcd_estimation_procedures.htm. Accessed January 19, 2021.

32. NHANES Survey Methods and Analytic Guidelines. https://wwwn.cdc.gov/nchs/nhanes/analyticguidelines.aspx#estimation-and-weighting-procedures. Accessed January 19, 2021.

33. HRS Survey Design and Methodology. https://hrs.isr.umich.edu/documentation/survey-design#22. Accessed January 19, 2021.

34. **Machlin S, Yu W, Zodet M**. Computing Standard Errors for MEPS Estimates. https://meps.ahrq.gov/survey_comp/standard_errors.jsp. .

35. Complex Sampling Weight and Preparing 2018 BRFSS Module Data for Analysis. https://www.cdc.gov/brfss/annual_data/2018/pdf/Complex-Smple-Weights-Prep-Module-Data-Analysis-2018-508.pdf. Accessed January 19, 2021.

36. Sample weights. https://psidonline.isr.umich.edu/Guide/FAQ.aspx?Type=10. Accessed January 19, 2021.

37. BRFSS Statistical Briefs. https://www.cdc.gov/brfss/data_documentation/statistic_brief.htm. Accessed January 19, 2021.

38. Survey Results and Products. Accessed January 19, 2021.

39. NCHS Data Briefs. https://www.cdc.gov/nchs/products/databriefs.htm. Accessed January 19, 2021.

40. Publication type: Statistical Brief. https://meps.ahrq.gov/mepsweb/data_stats/Pub_ProdLookup_Results.jsp?ProductType=Statistical%20Brief. Accessed January 19, 2021.

41. Publications. https://hrs.isr.umich.edu/publications. Accessed January 19, 2021.

42. Pubs & Media. https://psidonline.isr.umich.edu/Publications/. Accessed January 19, 2021.

43. **Saint-Maurice PF, Troiano RP, Bassett DR, Jr., et al.** Association of Daily Step Count and Step Intensity With Mortality Among US Adults. *JAMA.* 2020;323(12):1151-1160.

44. **Cheng YJ, Kanaya AM, Araneta MRG, et al.** Prevalence of Diabetes by Race and Ethnicity in the United States, 2011-2016. *JAMA.* 2019;322(24):2389-2398.

45. **Micha R, Penalvo JL, Cudhea F, Imamura F, Rehm CD, Mozaffarian D**. Association Between Dietary Factors and Mortality From Heart Disease, Stroke, and Type 2 Diabetes in the United States. *JAMA.* 2017;317(9):912-924.

46. **Murphy D, McCulloch CE, Lin F, et al.** Trends in Prevalence of Chronic Kidney Disease in the United States. *Ann Intern Med.* 2016;165(7):473-481.

47. **Geiss LS, Wang J, Cheng YJ, et al.** Prevalence and incidence trends for diagnosed diabetes among adults aged 20 to 79 years, United States, 1980-2012. *JAMA.* 2014;312(12):1218-1226.

48. **Dwyer-Lindgren L, Mackenbach JP, van Lenthe FJ, Flaxman AD, Mokdad AH**. Diagnosed and Undiagnosed Diabetes Prevalence by County in the U.S., 1999-2012. *Diabetes Care.* 2016;39(9):1556-1562.

49. **Patel SA, Winkel M, Ali MK, Narayan KM, Mehta NK**. Cardiovascular mortality associated with 5 leading risk factors: national and state preventable fractions estimated from survey data. *Ann Intern Med.* 2015;163(4):245-253.

50. **Egede LE, Walker RJ, Bishu K, Dismuke CE**. Trends in Costs of Depression in Adults with Diabetes in the United States: Medical Expenditure Panel Survey, 2004-2011. *J Gen Intern Med.* 2016;31(6):615-622.

51. **Bishu KG, Gebregziabher M, Dismuke CE, Egede LE**. Quantifying the Incremental and Aggregate Cost of Missed Workdays in Adults with Diabetes. *J Gen Intern Med.* 2015;30(12):1773-1779.

52. **Pilla SJ, Segal JB, Maruthur NM**. Primary Care Provides the Majority of Outpatient Care for Patients with Diabetes in the US: NAMCS 2009-2015. *J Gen Intern Med.* 2019;34(7):1089-1091.

53. **Wu J, Ward E, Lu ZK**. Addressing Lifestyle Management During Visits Involving Patients with Prediabetes: NAMCS 2013-2015. *J Gen Intern Med.* 2019;34(8):1412-1418.

54. **Bergmans RS, Zivin K, Mezuk B**. Depression, food insecurity and diabetic morbidity: Evidence from the Health and Retirement Study. *J Psychosom Res.* 2019;117:22-29.

55. **Smith BE, Miles TP, Elkins J, Barkin JL, Ebell MH, Ezeamama AE**. The Impact of Stressful Life Events on the Incidence of Type 2 Diabetes in U.S. Adults From the Health and Retirement Study. *J Gerontol B Psychol Sci Soc Sci.* 2020;75(3):640-649.

56. **Whisman MA, Li A, Sbarra DA, Raison CL**. Marital quality and diabetes: results from the Health and Retirement Study. *Health Psychol.* 2014;33(8):832-840.