

Longitudinal Data from IPUMS

Prepared for the Summer 2022 NDiRA Pre-Conference Workshop

July 19, 2022

Resources Covered

1. Survey-NDI-linked data: the National Health Interview Survey Linked Mortality File (NHIS-LMF)
2. The Medical Expenditure Panel Survey Household Component
3. The Multigenerational Longitudinal Panel (Steve Ruggles and Megan Schouweiler)
4. The Current Population Survey Panel (Renae Rodgers)

Intro to NHIS Linked Mortality File Data

National Health Interview Survey (NHIS)

- Collected annually through in-person interviewers from the Census Bureau on behalf of the National Center for Health Statistics (NCHS) since 1957 (digital copies of the data available from 1963)
- 30-60K households per year
 - 2018 and earlier years 70-100K persons per year
 - Longest-running annual survey of health in the world
- Cross-sectional with linkage to various admin sources, including the National Death Index
- Representative of civilian, non-institutionalized US population

DATA-INTENSIVE RESEARCH CONFERENCE

Pre-conference workshop
presented by **IPUMS** and **NDiRA**

Topics covered by NHIS

- A rich array of measures covering health status and medical conditions
- Health care access and utilization
- Socioeconomic and demographic correlates of health
- Widely used for the study of health inequities

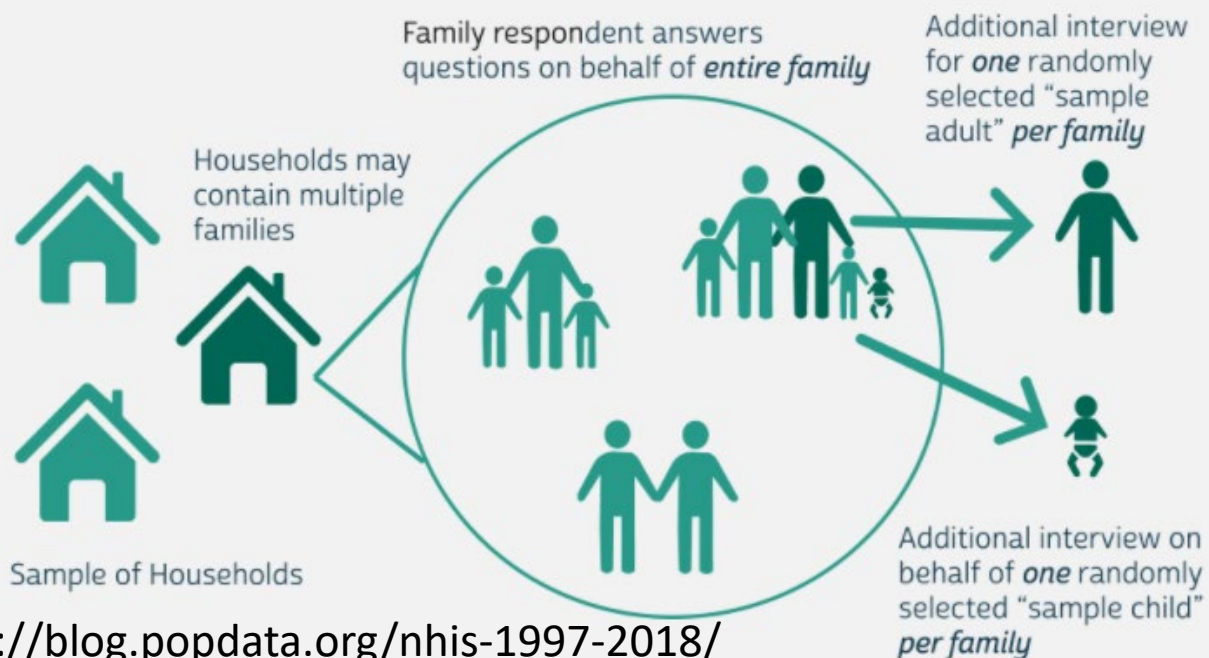
DATA-INTENSIVE RESEARCH CONFERENCE

Pre-conference workshop
presented by **IPUMS** and **NDiRA**

NHIS is a Complex Survey

- Different sets of questions are fielded to subsets of the sample, affecting sample sizes and appropriate choice of sampling weights

Figure 1. 1997-2018 NHIS Data Collection



<http://blog.popdata.org/nhis-1997-2018/>

Guidance on appropriate weighting and use of survey design variables:

nhis.ipums.org/nhis/userNotes.shtml

DATA-INTENSIVE RESEARCH CONFERENCE

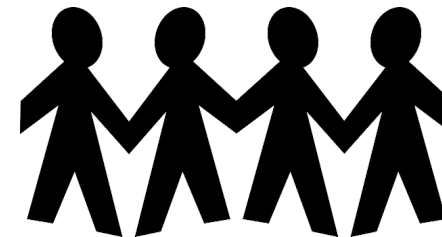
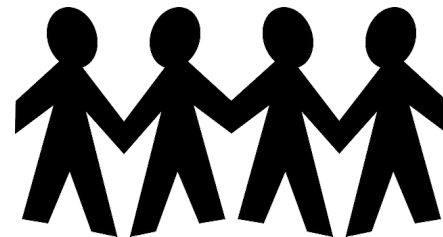
Pre-conference workshop
presented by **IPUMS** and **NDiRA**



Format for 1997-2018

Sample of households

**Households can
contain multiple
families**



DATA-INTENSIVE RESEARCH CONFERENCE

Pre-conference workshop
presented by **IPUMS** and **NDiRA**



One person – **family respondent** – interviewed about health of all family members

DATA-INTENSIVE RESEARCH CONFERENCE

Pre-conference workshop
presented by **IPUMS** and **NDiRA**

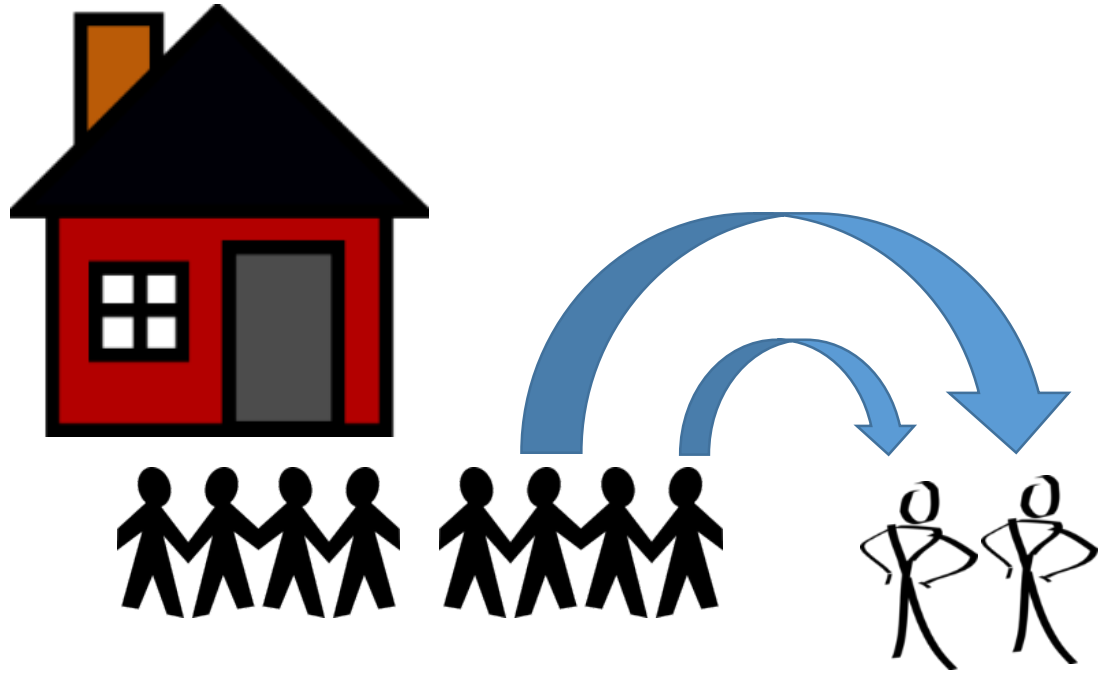


One person – **family respondent** – interviewed about health of all family members

- Demographics,
- Health status,
- Chronic conditions,
- Health insurance,
- Injuries

DATA-INTENSIVE RESEARCH CONFERENCE

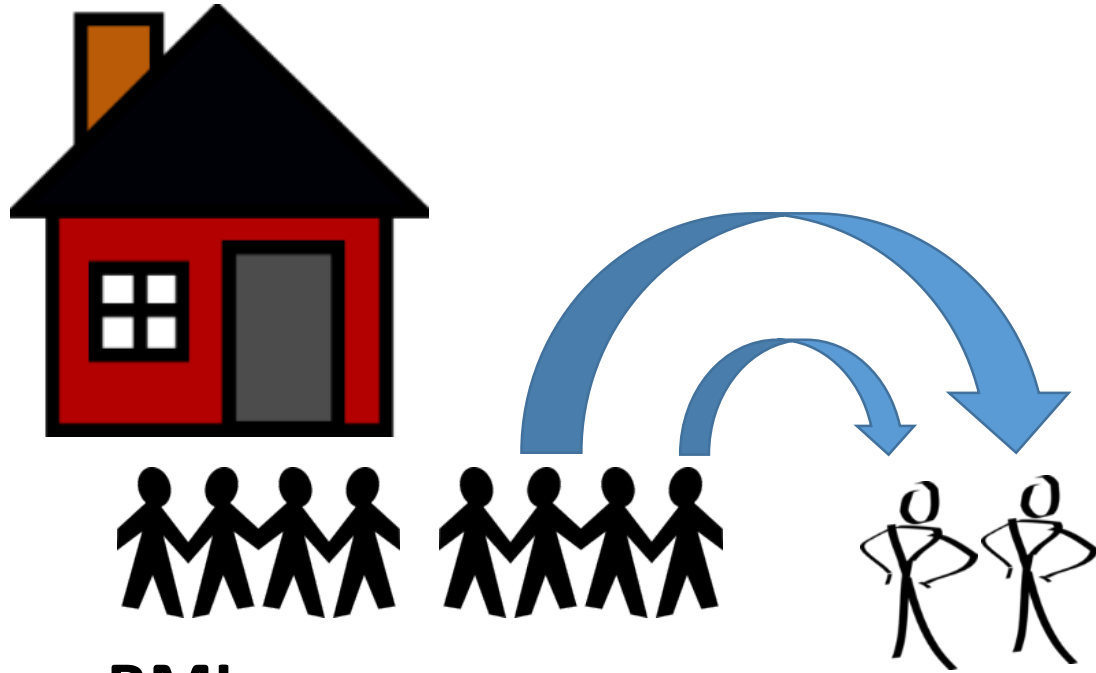
Pre-conference workshop
presented by **IPUMS** and **NDiRA**



One **sample adult**
and **sample child** per
family selected to
complete an
additional interview

DATA-INTENSIVE RESEARCH CONFERENCE

Pre-conference workshop
presented by **IPUMS** and **NDiRA**



One **sample adult**
and **sample child** per
family selected to
complete an
additional interview

- BMI,
- Mental health
- Select access to care and health behaviors measures
- Supplements
- (For adults) sexual orientation, job info

DATA-INTENSIVE RESEARCH CONFERENCE

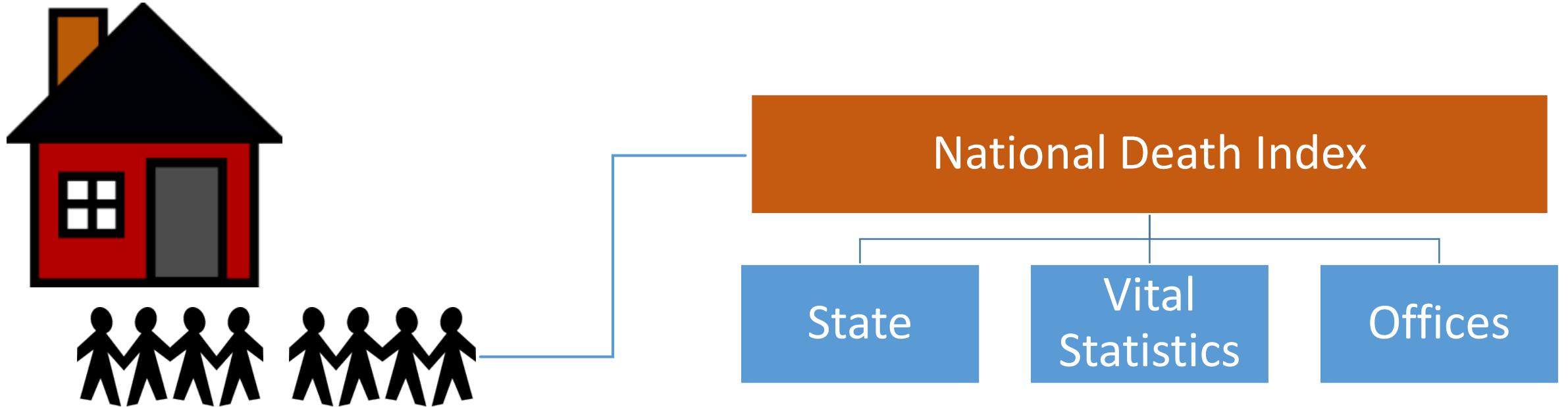
Pre-conference workshop
presented by **IPUMS** and **NDIRA**

NHIS-Linked Mortality File (NHIS-LMF)

- Produced by the Data Linkage Team at the National Center for Health Statistics
- Linkage of person records from the National Health Interview Survey to the **National Death Index (NDI)**
 - This is a centralized database containing ALL DEATHS IN THE UNITED STATES
 - Information on whether an NHIS participant has died during the mortality follow up period and, if so, when and how they died
 - More information available on restricted data available from NCHS (cdc.gov/nchs/data-linkage/mortality-restricted.htm)

DATA-INTENSIVE RESEARCH CONFERENCE

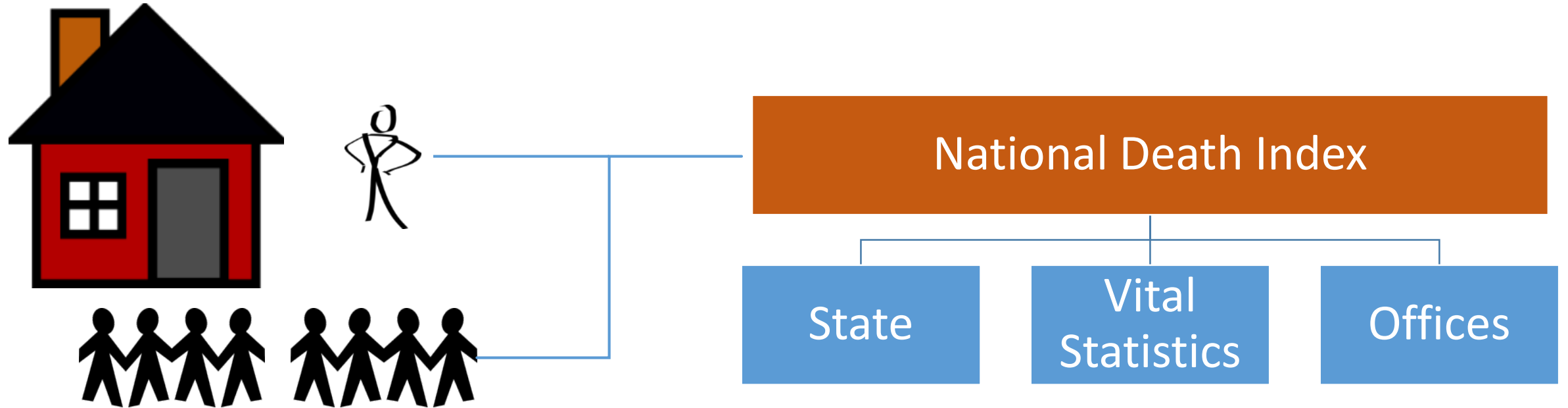
Pre-conference workshop
presented by **IPUMS** and **NDiRA**



All eligible adults 18+ who participated in the 1986-2014 NHIS

DATA-INTENSIVE RESEARCH CONFERENCE

Pre-conference workshop
presented by **IPUMS** and **NDiRA**



All eligible adults 18+ who participated in the 1986-2014 NHIS



Sample adults 18+ who participated in the 2015-2018 NHIS

Currently available mortality follow up through **December 31, 2019**

What kinds of research can you do
with the NHIS-LMF data?

At least 3 ways to use NHIS-LMF data to study mortality risks

1. Understanding disparities in mortality risks and how they may be changing over time
2. Capturing measures of health status, medical conditions, and/or health behaviors while decedents are still alive and evaluating associations with premature mortality or specific causes of death
3. Using 2018 or earlier data, can use a snapshot of family context -- including health and other characteristics of family members -- to study association with mortality risks

Second, I estimate all-cause death rates by age, sex, education, and period (m_{isep}) from weighted counts of deaths and quarter-years of exposure using the NHIS (Minnesota Population Center and State Health Access Data Assistance Center 2015). I use the NHIS because calculating education-specific death rates from death certificates and census population counts is problematic owing to dual data-source bias from education misreporting on death certificates (Hendi [forthcoming](#); Rostron et al. 2010; Sorlie and Johnson 1996). The NHIS is representative of the U.S. civilian

The Contribution of Drug Overdose to Educational Gradients in Life Expectancy in the United States, 1992–2011

Jessica Y. Ho¹

how it impacts different education groups. I use vital statistics and National Health Interview Survey data to examine the contribution of drug overdose to educational gradients in life expectancy from 1992–2011. I find that over this period, years of life lost due to drug overdose increased for all education groups and for both males and females. The contribution of drug overdose to educational gradients in life expectancy has increased over time and is greater for non-Hispanic whites than for the population as a whole. Drug overdose accounts for a sizable proportion of the increases in educational gradients in life expectancy, particularly at the prime adult ages (ages 30–60), where it accounts for 25 % to 100 % of the widening in educational gradients between 1992 and 2011. Drug overdose mortality has increased more rapidly for females than for males, leading to a gender convergence. These findings shed light on the processes driving recent changes in

Occupation, employment status, and “despair”-associated mortality risk among working-aged U.S. adults, 1997–2015

Iliya Gutin*, Robert A. Hummer

Department of Sociology and Carolina Population Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27516, United States of America

Preventive Medicine 137 (2020) 106129



ABSTRACT

The recent rise in U.S. midlife mortality has been conceptualized as a “working-class” crisis, defined by increasing mortality among blue-collar and/or unemployed workers and the decline of manual labor; yet research on the topic overwhelmingly focuses on educational attainment as the key socioeconomic determinant of midlife mortality, especially among “despair”-related deaths. The present study addresses this gap by using data on 360,146 adults ages 25–64 from restricted-use National Health Interview Survey-Linked Mortality Files (1997–2015; average follow-up 9.87 years) to estimate associations between individuals' occupation and employment status and alcoholic liver disease, suicide, or accidental poisoning mortality risk, net of confounders. Adults in service, manual labor, and transport occupations exhibited *two-to-three times* the risk of mortality from accidental poisonings compared to those in managerial/administrative positions. Notably, health professionals exhibited the highest accidental poisoning mortality risks. Relative to managerial/administrative professionals, adults not in the labor force had *double* the suicide risk and nearly *seven times* the accidental poisoning risk, net of confounders. Unemployed adults and those having never worked also had elevated risks from accidental poisoning mortality. Critically, the fact that individuals' occupations and employment status are independently associated with midlife mortality due to deaths of despair – especially accidental poisoning – highlights the need for measures of socioeconomic status beyond educational attainment and income in understanding rising midlife mortality. Moreover, policies addressing working-aged mortality must target particular workplace contexts and the consequences of unemployment, both of which affect a large and growing segment of the working-aged U.S. population.

Occupation is based on verbatim responses for “usual” occupation converted to an occupational code; IPUMS, which provides the harmonized NHIS data used in this analysis, has standardized codes to 1995 Standard Occupational Classifications (Blewett et al., 2016). Re-



Psychological Distress and Alzheimer's Disease Mortality in the United States: Results from the 1997–2014 National Health Interview Survey-National Death Index Record Linkage Study

Gopal K. Singh, PhD¹ and Hyunjung Lee, PhD^{1,2} 

Abstract

Objective: This study examines the association between psychological distress and Alzheimer's disease mortality among US adults aged ≥ 45 . **Methods:** We analyzed the Kessler 6-item psychological distress scale as a risk factor for Alzheimer's mortality using the pooled 1997–2014 National Health Interview Survey (NHIS)- National Death Index (NDI) database ($N = 265,089$). Cox regression was used to model mortality as a function of psychological distress and sociodemographic and behavioral covariates. **Results:** The Alzheimer's mortality risk was 97% higher ($HR = 1.97$; 95% confidence interval [CI] = 1.37, 2.84) in adults with serious psychological distress compared with those without psychological distress, controlling for sociodemographic covariates. The relative mortality risk remained statistically significant ($HR = 1.49$; 95% CI = 1.04, 2.13) after additional adjustment for smoking, alcohol consumption, health status, activity limitation, and body mass index. **Discussion:** US adults had significantly higher risks of Alzheimer's disease mortality at higher psychological distress levels. These findings underscore the significance of addressing psychological well-being as a strategy for reducing Alzheimer's disease mortality.

The data for this study are derived from the National Health Interview Survey (NHIS) linked to the National Death Index (NDI) (Blewett et al., 2019). As a nationally representative,

Labor force status as a buffer against mortality risks associated with alcohol consumption: A study of adult U.S. women, 2001–2015

Muntasir Masum^{a,*}, Johnelle Sparks^b

^a *Prevention Research Center, The Pennsylvania State University, 314 Biobehavioral Health Building, University Park, PA 16802, United States*

^b *Department of Demography, The University of Texas at San Antonio, One UTSA Circle, San Antonio, TX 78249, United States*

Preventive Medicine 161 (2022) 107139

A B S T R A C T

We utilized public-use data from the 2001–2013 National Health Interview Survey-Linked Mortality Files (NHIS-LMF), followed through December 31, 2015, derived from the IPUMS NHIS database (Blewett et al., 2019). The NHIS is a nationally representative cross-sectional

The association between women's labor force participation, their alcohol consumption patterns, and mortality risk is unclear. This study assessed all-cause mortality risk among women in the United States, considering their labor force status and alcohol drinking. This study used discrete-time hazard models to examine this association using 2001–2015 National Health Interview Survey-Linked Mortality Files (NHIS-LMF) data ($n = 147,714$) for women aged 25 to 65 with 5725 deaths in this sample. Complex survey-weighted adjustments and *E*-values calculations were used to limit quantitative and observational biases. Alcohol consumption and labor force status together lead to substantial mortality risks. There is a statistically significant mortality risk among unemployed women (HR 2.15, 95% CI 1.18–3.91) and women not in labor force (HR 2.38, 95% CI 1.87–3.01). In the stratified models, non-Hispanic blacks (HR 1.48, 95% CI 1.30–1.67) and Asians (HR 1.93, 95% CI 1.54–2.44) have heightened mortality risks borne out of employment. Women with higher psychological distress have a 26% higher risk of all-cause mortality when not in labor force. With the help of cross-sectional data, this study demonstrates that women not in labor force and unemployed women are more likely to be affected by their drinking habits, and their employment status is associated with lower mortality risk. Further research should be focused on cause-specific mortality, gender roles and norms, reasons for unemployment, and comorbidities using more recent data, causal modeling techniques, and an extended mortality follow-up period.

Things to note and future plans

- At this time, the IPUMS NHIS site includes the NHIS-LMF public use data released in 2019, with NHIS participants from the 1986-2014 samples with mortality follow up through December 31, 2015
- Detailed cause of death codes and information such as location of death are not available on the PUF and require accessing the restricted version of the NHIS-LMF data
- In May 2022, the NCHS Data Linkage team released a public use version of the NHIS-LMF data with NHIS participants from the 1986-2018 samples with mortality follow up through December 31, 2019
- However...

Things to note and future plans, continued

- ...they made two major changes to the NHIS-LMF data in the most recent update
 1. Additional samples (2015-2018) only link *SAMPLE ADULTS* (one randomly selected adult per family) to the NDI
 2. They substantially changed the linking algorithm used to generate probabilistic matches between person records in the NHIS and death records in the NDI (now uses the Fellegi-Sunter method)
 - Algorithm change has led to changes in mortality information
 - Preliminary investigations show that e.g., 13,497 persons (1.7% of all eligible cases) identified as deceased in the last LMF data update were identified as alive in the most recent update

Things to note and future plans, continued

- We will release a technical note with the results from our assessment of discrepancies between LMF updates
- To enhance transparency and reproducibility of research using the NHIS-LMF data, we are in the process of making available on our website sets of mortality variables, one for each update of the NHIS-LMF data
- Work currently under way

Using MEPS-HC for longitudinal analysis

MEPS Household Component in Summary

- Collected annually by Agency for Healthcare Research and Quality (AHRQ) since 1996
- Short panel covering two calendar years
- Only nationally-representative source for information on US medical expenditures
- Representative of civilian non-institutionalized US population

MEPS Household Component in Summary

- 12-13K households and 30-35K persons interviewed per year
- Re-interviews a subsample of households participating in the previous year's NHIS
 - Can be linked to NHIS using restricted linking keys, adding more information and another observation

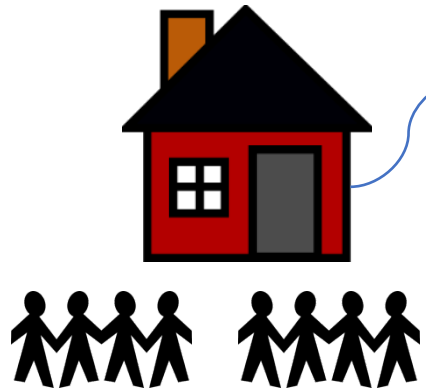
Topics covered by MEPS

- Health care utilization: Types and dates of medical care use throughout the two-year panel, reason for visit, types of services used, type of provider seen
- Health care expenditures: Amounts for care, amounts paid by specific payers (e.g., self/family, Medicare, private insurance)
- Medical conditions

DATA-INTENSIVE RESEARCH CONFERENCE

Pre-conference workshop
presented by **IPUMS** and **NDiRA**

(1) Household Component



MEPS public use data

(2) Medical Provider Component



(3) Pharmacy Component



Using MEPS for Longitudinal Analysis

1. Study the relationship between short-term employment and health trajectories, e.g.:
 - What is the effect of retirement on own mental health and that of other family members?
 - What happens to health insurance and/or health status, after job loss?
 - What are the effects of spousal/parental job loss on spousal and/or child health outcomes?
2. Study what happens following health shocks, e.g.,
 - What are the short-term effects of a new diagnosis of a serious medical condition, an injury, or a major episode of health care on employment?
 - Effects of own health shock on health and employment of co-residential family members, family income
3. Study what happens after family change, e.g.,
 - What are the effects of marriage, relationship dissolution, the birth of a new child, or retirement on own mental health and that of other family members?

Challenges to using the MEPS-HC data longitudinally

1. Information is scattered across multiple data files, most containing different units of analysis
 - Full panel file is only available for person-level data

DATA-INTENSIVE RESEARCH CONFERENCE

Pre-conference workshop
presented by **IPUMS** and **NDiRA**

Select by year and/or data file type

Year:

Data file types to include in search (check all that apply). Click information icon ⓘ for file details. Click link for full list of file types in category.

Search all data files ⓘ

Household Component Full-Year files ⓘ

Expenditure and utilization data for the calendar year from several rounds of data collection.

Full-Year Consolidated Data files

Full-Year Population Characteristics files

Full-Year Medical Organizations Survey files

Medical Conditions files

Risk Adjustment Scores files

Employment Variables file

Jobs files

Food Security file

Person Round Plan files

Longitudinal Data files

Preventive Care Self-Administered Questionnaire file (2014)

Supplemental Variables files (1996-2000)

Health Insurance Plan Abstraction file (1996)

Long Term Care file (1998)

Household Component Event files ⓘ

Data for the calendar year on unique household-reported medical events.

Prescribed Medicines files

Dental Visits files

Other Medical Expenses files

Hospital Inpatient Stays files

Emergency Room Visits files

Outpatient Visits files

Office-Based Medical Provider Visits files

Home Health files

Appendix to MEPS Event files

Including the condition-event linkage file (CLNK) and PMED-event linkage file (RXLK).

Information about demographics, mental health, health insurance

Information about medically-treated medical conditions

Detailed information about jobs and employment

Information about hospitalizations

Linking keys to combine information on hospitalizations with medical conditions

Challenges to using the MEPS-HC data longitudinally

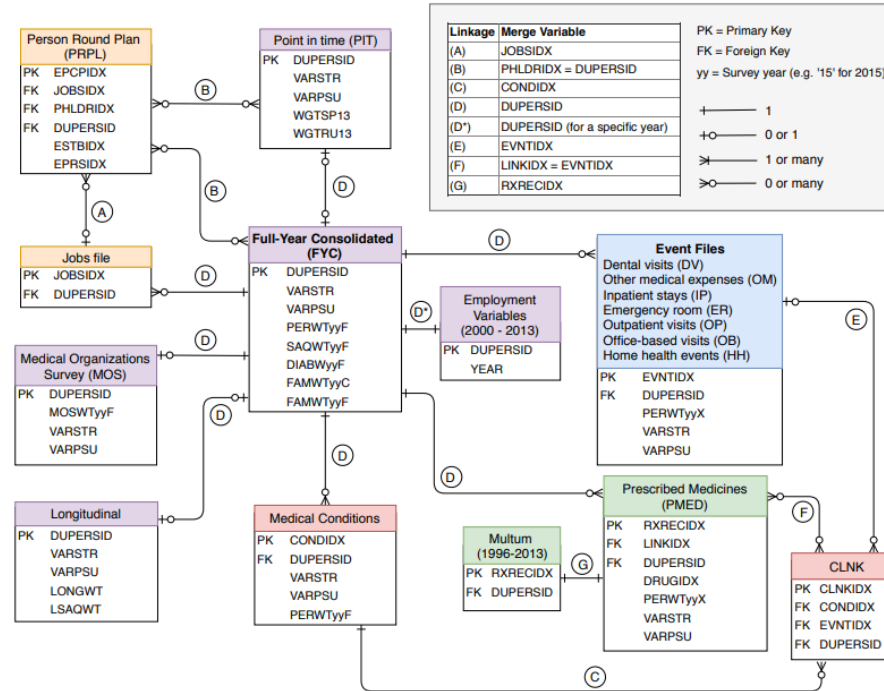
1. Information is scattered across multiple data files, most containing different units of analysis
 - Full panel file is only available for person-level data
 - Linking across units of analysis in the MEPS is especially complex

DATA-INTENSIVE RESEARCH CONFERENCE

Pre-conference workshop
presented by **IPUMS** and **NDiRA**

MEPS Public Use Files (PUFs)

Entity Relationship Diagram (ERD) with survey and linkage variables



Type	Variable	Label
ID	CLNKIDX	CLNK ID: CONDIDX + EVNTIDX
	CONDIDX	CONDITION ID
	DRUGIDX	DRUG ID (DUPERSID + COUNTER)
	DUPERSID	PERSON ID
	EPCPIDX	UNIQUE RECORD ID (EPRSDX + DUPERSID)
	EPRSDX	ESTBIDX + POLICYHOLDER ID + ROUND NUMBER
	ESTBIDX	ESTABLISHMENT ID
	EVNTIDX	EVENT ID
	JOBSIDX	JOB ID NUMBER
	LINKIDX	ID FOR LINKAGE TO COND/OTHER EVENT FILES
	PHLDRIDX	POLICYHOLDER'S DUPERSID
	RXRECIDX	UNIQUE RX/PREScribed MEDICINE IDENTIFIER

Type	Variable	Label
Survey	VARPSU	VARIANCE ESTIMATION PSU
	VARSTR	VARIANCE ESTIMATION STRATUM
Weight	DIABWyyF	FINAL DIABETES CARE SUPPLEMENT WEIGHT
	FAMWyyC	POVERTY ADJUSTED FAMILY WEIGHT (CPS FAMILY)
	FAMWyyF	FINAL FAMILY WEIGHT
	LONGWT	LONGITUDINAL WEIGHT
	LSAQWT	LONGITUDINAL SAQ WEIGHT
	MOSWyyF	MOS FILE FINAL PERSON WEIGHT
	PERWyyF	FINAL PERSON WEIGHT
	SAQWyyF	FINAL SAQ PERSON WEIGHT
	WGTRU13	FAMILY WEIGHT, ROUND 1/3
	WGTSP13	PERSON WEIGHT, ROUND 1/3

This graphic contains information for the most common linkages among the most frequently used datasets. For complete descriptions of all MEPS Public

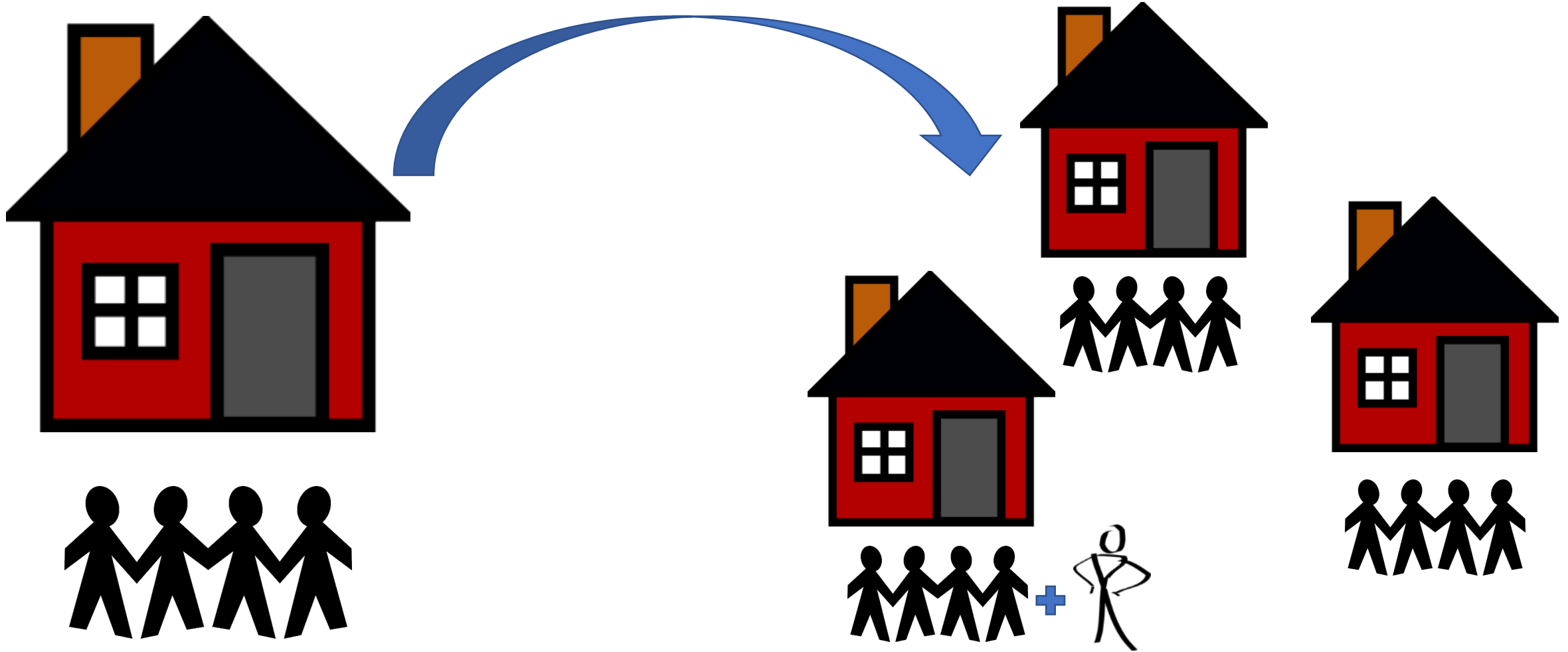


Challenges to using the MEPS-HC data longitudinally

1. Information is scattered across multiple data files, most containing different units of analysis
 - Full panel file is only available for person-level data
 - Linking across units of analysis in the MEPS is especially complex
2. Only some family members are included in the longitudinal file

DATA-INTENSIVE RESEARCH CONFERENCE

Pre-conference workshop
presented by **IPUMS** and **NDiRA**



Last Year's National Health Interview Survey

This Year's MEPS

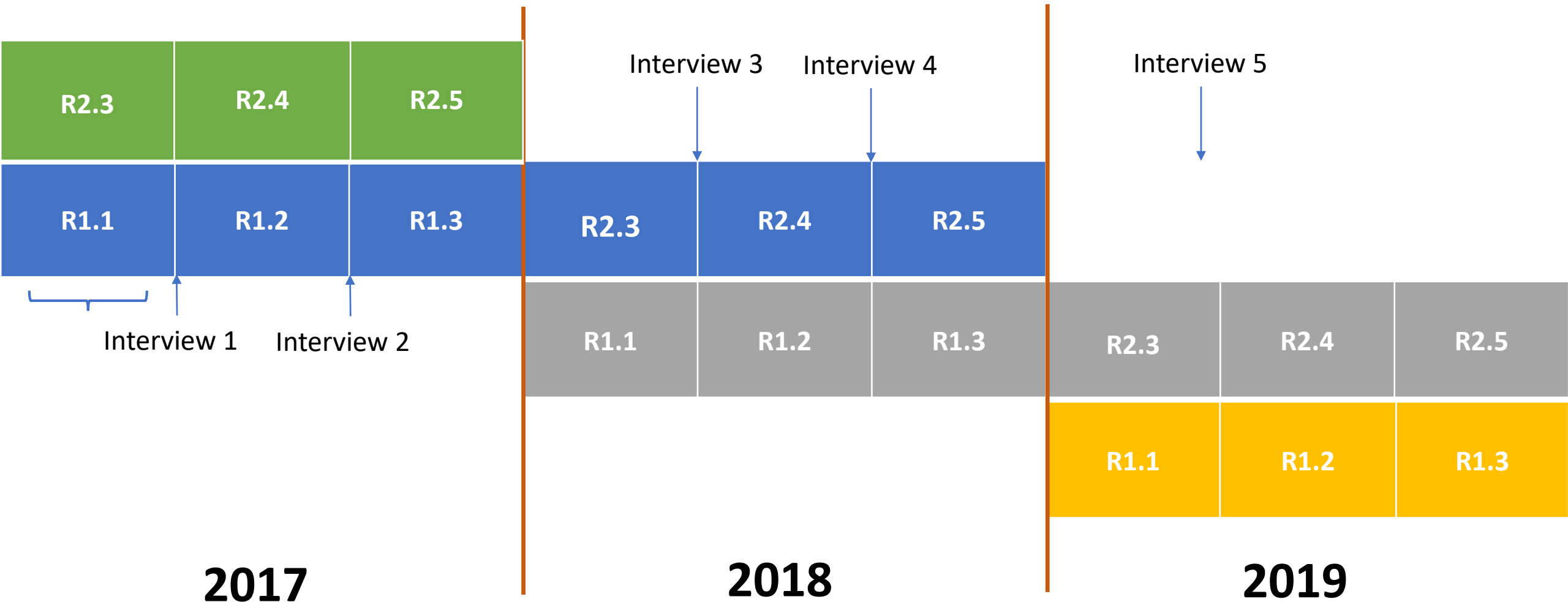
Challenges to using the MEPS-HC data longitudinally

1. Information is scattered across multiple data files, most containing different units of analysis
 - Full panel file is only available for person-level data
 - Linking across units of analysis in the MEPS is especially complex
2. Only some family members are included in the longitudinal file
3. Different time scales and file structures make it difficult to combine information across files

DATA-INTENSIVE RESEARCH CONFERENCE

Pre-conference workshop
presented by **IPUMS** and **NDiRA**

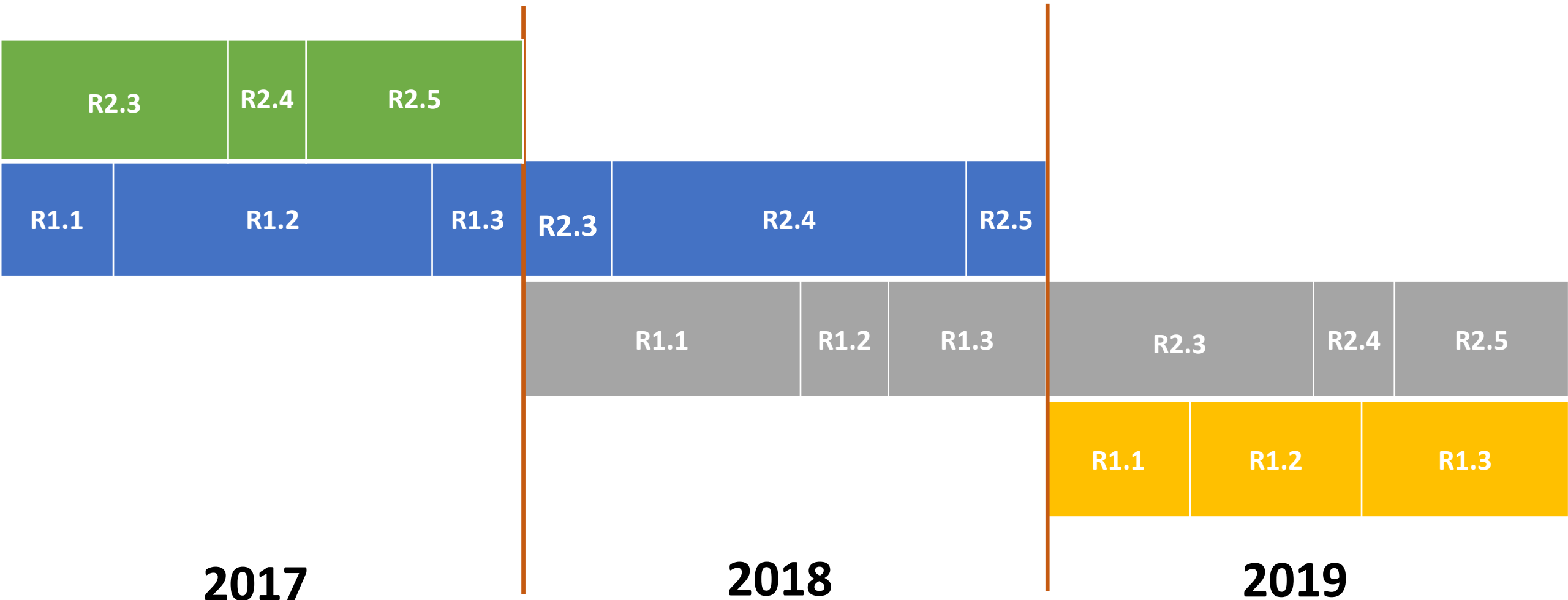
Sequential Overlapping Panel Design



DATA-INTENSIVE RESEARCH CONFERENCE

Pre-conference workshop
presented by **IPUMS** and **NDiRA**

Sequential Overlapping Panel Design



DATA-INTENSIVE RESEARCH CONFERENCE

Pre-conference workshop
presented by **IPUMS** and **NDiRA**

Illustration of Medical Encounter File Structure

mepsid	evntid	# days in month received care	Month	Year	Type of home health care provider	Care provider was a home health aide	Self/Agency
2291700101	003001	8	1	2018	Other (paid)	No	Agency
2291700101	003201	1	1	2018	Other (paid)	No	Agency
2291700101	003501	20	1	2018	Relative	n/a	n/a
2291700101	003101	8	2	2018	Other (paid)	No	Agency
2291700101	003301	1	2	2018	Other (paid)	No	Agency
2291700101	003601	20	2	2018	Relative	n/a	n/a
2291700101	003701	20	3	2018	Relative	n/a	n/a
2291700101	003801	28	4	2018	Relative	n/a	n/a
2291700101	103901	28	5	2018	Relative	n/a	n/a
2291700101	104701	2	5	2018	Other (paid)	n/a	Self

Our Approach to Handling Challenges

1. Make it easier to combine data across different record structures and over time
2. Make it easier to match up differing time scales
3. Make it easier to construct user-defined longitudinal resources from the cross-sectional data

DATA-INTENSIVE RESEARCH CONFERENCE

Pre-conference workshop
presented by **IPUMS** and **NDiRA**

Maximize Flexibility in Data Structures Offered

The image shows a screenshot of the IPUMS Health Surveys web application. At the top left is the IPUMS logo and 'HEALTH SURVEYS' text. To the right, there are tabs for 'NHIS' and 'MEPS', and the text 'MEDICAL EXPENDITURE PANEL SURVEY'. Below this is a navigation bar with 'HOME | SELECT DATA | MY DATA | SUPPORT'. A 'DATA CART' widget shows 'YOUR DATA EXTRACT' with '0 VARIABLES' and '0 SAMPLES'. Below the navigation bar is a 'SELECT VARIABLES' section with dropdown menus for 'ANNUAL', 'ROUND', 'CONDITION', 'EVENT', and 'A-Z', and a 'SEARCH' input. A 'SELECT SAMPLES' button is on the left. A 'CHANGE DATA STRUCTURE' button is circled in red. To the right of it is a 'DISPLAY OPTIONS' button and a 'HELP' link. Below the 'SELECT VARIABLES' section is the text 'Select samples and variables to build a data extract.' A modal window titled 'EXTRACT DATA STRUCTURE' is open, showing four radio button options: 'Rectangular', 'person (default)', 'Hierarchical', and 'Wide'. The 'Wide' option is selected. A yellow arrow points to the 'Wide' option, and another yellow arrow points to the 'EXTRACT DATA STRUCTURE' title. At the bottom of the modal are 'CANCEL' and 'APPLY SELECTIONS' buttons. At the bottom of the page, there is a footer with 'SUPPORTED BY: THE EUNICE KENNEDY SHRIVER NATIONAL INSTITUTE OF CHILD HEALTH AND HUMAN DEVELOPMENT, STATE MANAGER, STATE HEALTH ACCESS DATA'.

Coming Features: MEPS variable builder

- Needing to work directly with data on the child records can act as a barrier to using the MEPS data in analysis
- Planned activity to develop a VARIABLE BUILDER that would allow users to
 - Define summary variables based on child record types through an easy-to-use interface
 - Include any user-defined summary variables in your IPUMS extract as a person-level variable
- Examples:
 - a count of medical encounters for diabetes by type of provider
 - the sum of medical expenditures paid for by self/family for cancer across all medical encounters

DATA-INTENSIVE RESEARCH CONFERENCE

Pre-conference workshop
presented by **IPUMS** and **NDiRA**

Thank you!