

Supplementary Information for

The risk of losing health insurance in the US is large, and remained so after the  
Affordable Care Act

Liran Einav and Amy Finkelstein<sup>1</sup>

<sup>1</sup> To whom correspondence may be addressed. Email: [afink@mit.edu](mailto:afink@mit.edu).

**This PDF file includes:**

Supplementary text

Tables S1 to S3

## Supplementary Information

### The risk of losing health insurance in the US is large, and remained so after the Affordable Care Act

Einav and Finkelstein

## Data

### MEPS

The Medical Expenditure Panel Survey (MEPS) is a two-year panel survey overseen by the Agency for Healthcare Research and Quality of the Department of Health and Human Services. The MEPS is designed to produce nationally representative estimates of the use and cost of health services and insurance among the US civilian non-institutionalized population. Each year, a new panel wave is formed as a sub-sample of the households from the previous year’s National Health Interview Survey (NHIS), with racial and ethnic minority populations over-sampled until 2016 (Medical Expenditure Panel Survey 2021a). Households are followed for a period of 24 months, such that in any given year there are two active cohorts.

Specifically, the MEPS samples occupied housing units (“dwelling units”, DUs) from the previous year’s NHIS. Within each dwelling unit, reporting units (RUs) are identified as “a person or group of persons in the sampled DU who are related by blood, marriage, adoption, or other family association [including self-identification as such]” (Medical Expenditure Panel Survey 2021a). One individual in the RU (the “respondent”) answers all questions on behalf of other members.

Households (i.e. RUs) are interviewed in five waves over the two-year sample period. Each interview collects information about the “reference period” between the previous interview (or panel start date) and the current interview. Waves are approximately evenly spaced over the 24 months; however, as interviews are conducted throughout the wave, exact reference periods vary from household to household. After the conclusion of the interview, and with respondent permission, the MEPS contacts a sample of the medical providers identified by respondents in the household component to verify and, if necessary, revise the data on respondent healthcare utilization.

Survey weights are required to produce nationally representative estimates with the MEPS. As with many surveys, some individuals participate in the MEPS for only part of the time that they are “in scope” (i.e. civilian non-institutionalized). The MEPS longitudinal weights, assigned to individuals in the panel data release, are designed to account for attrition from such individuals and also include adjustments for the oversampling and clustering built into the MEPS design (Medical Expenditure Panel Survey 2021b). We use these weights throughout our analysis.

### SIPP

The US Census Bureau’s Survey of Income and Program Participation (SIPP) is a multi-year panel survey designed to provide nationally representative estimates of income, employment, and government program participation. In the current structure, respondents are interviewed

once annually over the course of four years, with new cohorts beginning each year. Similar to the MEPS, the SIPP is a sample of the civilian non-institutionalized population. The SIPP samples in two stages: first selecting primary sampling units (PSUs) of one or more counties, and then selecting addresses within that PSU. The SIPP stratifies by income within these PSUs, and oversamples low-income households. In the current structure, launched in 2014, interviews are conducted once annually, and survey participants are followed for a period of 48 months. During the first wave of interviews, the SIPP attempts to interview all members of the sampled households (members are defined as an individual who usually sleeps at that address). Individuals over 15 answer for themselves (or by proxy if necessary) and for any individuals under 15 in the household. After the first wave (i.e. the first year), however, the SIPP follows all *individuals* interviewed in Wave 1, whether they continue to live in the Wave 1 household or not (Irving and Smith 2021). In each wave of annual interviews, individuals are asked questions about the previous calendar year, with month-level data generated as necessary (i.e. if individuals report insurance at any point during the reference period—see detail below).

The SIPP generates survey weights that account for the oversampling of certain populations and differing response rates in order to produce representative population estimates. These survey weights are generated for each year of interview, and for multi-year periods which cover multiple interviews (“longitudinal weights”). The SIPP assigns longitudinal weights to individuals with non-missing survey response in each month of the multi-year file. These weights include adjustments for non-response, attrition, and other post-stratification adjustments. Note that in 2019 and 2020, the SIPP experienced substantial data collection difficulty due to a lapse in funding and COVID-19. While longitudinal weights are in principle designed to account for attrition, the magnitude of this non-response is sufficient that estimates generated from these two years of data should be interpreted with this attrition in mind (US Census Bureau 2021a, 2021b).

## Sample Construction

We make several restrictions to the MEPS data to arrive at our baseline sample. Table S1 shows the impact of the restrictions on our sample sizes; Table S2 shows the impact on the sample characteristics. For each cohort, we restrict the sample to individuals who are under 65 at the end of the two-year so that they are not eligible for Medicare coverage for the elderly. We also require that they have recorded responses to all five survey waves, which also means that they must survive until the end of the survey and be at least two years old by that time. Finally, we require that they have non-missing information on insurance in every survey wave. The resultant post-ACA sample consists of 59,784 unique individuals who are 2-64 years old by the end of the two-year survey.

We make analogous restrictions in the 2017-2019 SIPP. Specifically, we require that individuals are under 65 at the end of the three year survey. We also require that they have recorded responses each year, which also means that they must survive until the end of the survey period and be at least three years old by that time. Finally, we require that they have non-missing information on insurance in every survey. These restrictions are shown in greater detail in Table S3.

## Measuring Insurance Coverage

### MEPS

Although most data are collected at the round level (i.e. since the last interview, or as of the current interview), insurance status is elicited during each interview for each month of the reference period corresponding to that round for that household (January 1 of the first survey year or the previous interview date). Respondents are first asked if they or the household members for whom they are answering have had any insurance coverage of a given type since the beginning of the reference period. If affirmative, respondents are then asked if they (or the household member) have been continuously covered since the last interview date (approximately 4-5 months prior). If not, the current month’s coverage is first assessed. Then, each month between the last interview month and the current interview month are displayed and respondents are asked about whether for that month “For each of the following months, {were/was} {you/{PERSON}} covered for the whole month, part of the month, or not at all during the month?”

We define an individual as having any insurance coverage in a given month if she reports any private coverage and/or any public coverage. We then consider whether the individual has each of the following three (non-exhaustive) sub-types of insurance: employer-provided coverage (which may be provided by an employer or union), Medicaid, or private coverage through the private exchanges (post-ACA; note that private exchange coverage becomes an option only in 2014). These variables are measured directly from existing MEPS variables (*MCD*, *PEG*, *PRX*, respectively). Note that, following the approach of the US Census Bureau (e.g. Keisler-Starkey and Bunch 2021), we allow individuals to have multiple insurance types in a given month throughout our analysis. 1.3% of our analysis sample reports having more than one of the three types of insurance we study in a given month; when we conduct analyses by insurance type they are double counted.

### SIPP

For each of several types of possible health insurance, respondents are asked whether they currently receive insurance from that source. If not, they are asked whether they did at any point during the reference period. If respondents either currently or have had insurance of a given type, they are asked for the start and end dates of that spell. For example, someone who has not previously reported Medicaid would be asked: “When we talked to you in [the previous interview month], you were not receiving Medicaid. When did you start? When did you stop?” If the end date is before the current month, respondents are asked if they had another spell of that insurance type, and the process repeats. The SIPP then constructs monthly indicators from these start and end dates.

We code an individual as having insurance in a given month if she reports any public or private insurance in that month. For summary statistics in Table S3, we construct an indicator for private coverage through the exchanges as the logical intersection of any private coverage and coverage purchased through Healthcare.gov or a state-based exchange (the latter is a SIPP-provided variable). Indicators for Medicaid and employer-provided are directly available as SIPP variables. For summary statistics, we also bin together all individuals who report insurance, but not of one of the aforementioned types.

## Other Variable Definitions

The following section provides detail on the construction of the other demographic variables used either for subgroup analysis or summary statistics in the MEPS and the SIPP.

### Age

In the MEPS, we define age as of December 31 of the second survey year. This is directly reported in the MEPS data, as *AGEY2X*. These are the cuts that we use for sample construction and analysis by age. We construct an analogous measure in the SIPP: the highest observed age over the sample period (i.e. as of the final survey month, for individuals in scope for all 36 months).

Note that education is a slight exception in both of these cases. In the MEPS, we report education for individuals age 18 or older as of the end of the first survey year—this is because education is assessed at the first interview. Similarly, when we report education in the SIPP summary statistics, we restrict the sample to individuals age 18 or older as of the 12th survey month (the first time that education is assessed).

### Race

In the MEPS instrument, individuals are categorized as Hispanic or not (variable *HISPANX*) and as belonging to one of 5-6 “race” categories (*RACEX* or *RACEV1X*, depending on survey year). In our analyses, we allow Hispanic to supersede other classifications, and code individuals as Hispanic or non-Hispanic White, Black, American Indian/Alaska Native, Asian/Pacific Islander, or Multiracial.

Variables for Hispanic/Non-Hispanic and Race (White/Black/Asian/Other) are available directly from the SIPP (*EORIGIN* and *ERACE*). We use the intersection of these two variables to generate Non-Hispanic White/Black/Asian and Hispanic indicators in the data.

### Education

During the 2012-13, 2013-14, and 2014-15 MEPS surveys, the standard education variable (*HIDEG*) was replaced with several alternatives (including a crosswalk, *EDRECODE*), and then reinstated. This results in two coding schemes. Prior to 2012-13 and after 2014-15, we determine education from *HIDEG*, “highest degree reached.” We code four education bins as follows: Below High School is no degree; High School is either a GED or high school diploma; College is a Bachelor’s degree; Above College is a master’s degree or doctoral degree (we combine these latter two categories in our analysis). “Other degree” cannot be classified in this ordinal scheme—therefore, along with juveniles, non-response, and inapplicable participants, such individuals are left with missing education status. In 2012-13, 2013-14, and 2014-15, we use *EDRECODE*, the MEPS crosswalk to the *HIDEG* variable used in other years. To create an equivalent scale, we define Below High School as (i) less than 8th grade and (ii) 9-12th grade, no degree; High School as (i) a GED or high school degree and (ii) beyond high school with no BA; College as a bachelor’s degree; and Above College as higher than a bachelor’s degree. Once again, missing and inapplicable responses and juveniles are left with missing education status. Note that in the MEPS, education level is assessed at

the first interview, and we define education only for individuals age 18 or older (when age is first observed, at the end of the first survey year).

Our SIPP education estimates mirror the MEPS. We classify individuals as having less than a high school degree if they are at least 18 years of age and report between “less than first grade” and “12th grade, no diploma” as the highest level of school completed when education is assessed in the first survey year. We similarly assign “high school” to educational attainment ranging from “high school graduate” to “associates degree,” and bachelor’s, master’s, professional school, and doctoral degrees to “college or above.”

### **Prior diagnoses**

We code individuals in the MEPS as having or not having had a prior diagnosis of coronary heart disease, diabetes, high blood pressure, heart attack, or stroke. This initial health status is determined by MEPS health assessments from the first survey year. We consider an individual to have a prior diagnosis if she (1) reports having ever received a diagnosis of the given condition *and* (2) reports having received it when she was at least one year younger than her age in the first survey interview.

We do not consider prior health conditions in the SIPP.

Table S1: MEPS cohort sample size

Cohort	Under 65 (1)	& Alive and Observed all Rounds (2)	Analysis Sample (3)
2007-08	10,992	9,941	9,938
2008-09	16,455	14,740	14,730
2009-10	14,334	13,006	12,993
2010-11	12,831	11,664	11,656
2011-12	16,276	14,835	14,833
2012-13	15,974	14,639	14,638
Total pre-ACA	86,862	78,825	78,788
2013-14		-----excluded cohort-----	
2014-15	13,951	12,736	12,725
2015-16	14,731	13,444	13,434
2016-17	13,402	12,264	12,259
2017-18	11,995	11,001	10,995
2018-19	11,344	10,372	10,371
Total post-ACA	65,423	59,817	59,784

Notes: Table presents sample size for each MEPS cohort after sequentially applying three sets of sample restrictions: that individuals are under the age of 65 by the end of the survey period (column (1)), alive and observed for all interview rounds (column (2)), and have observed insurance status each month (column (3)). Cohorts beginning in 2007 through 2018 are considered. The 2013-2014 cohort is excluded from our analyses, as the individuals included are “partially treated” by the ACA implementation in 2014.

Table S2: Post-ACA MEPS summary statistics, by sample restriction

	Under 65 (1)	& Alive and Observed all Rounds (2)	Analysis Sample (3)
No. of Obs.	65,423	59,817	59,784
Share of Unrestricted Sample	1.000	0.914	0.914
Age			
<18	0.288	0.242	0.242
18-30	0.205	0.212	0.212
31-50	0.295	0.317	0.317
51-64	0.213	0.229	0.230
Race			
non-Hispanic White	0.570	0.575	0.575
non-Hispanic Black	0.129	0.127	0.127
non-Hispanic Asian/Pacific Islander	0.062	0.063	0.063
Hispanic	0.199	0.197	0.197
Education <sup>a</sup>			
Below High School	0.125	0.120	0.120
High School	0.531	0.530	0.530
College or Above	0.344	0.350	0.350
Female	0.503	0.506	0.506
Prior Adverse Health Events	0.242	0.236	0.236
Insurance in first survey month <sup>b</sup>			
Medicaid	0.195	0.190	0.189
Employer provided	0.548	0.560	0.561
Private exchange	0.024	0.024	0.024
Other insurance coverage	0.091	0.084	0.084
Uninsured	0.151	0.151	0.151

Notes: Table presents sample size and summary statistics for the pooled post-ACA MEPS cohort after sequentially applying three sets of sample restrictions: that individuals are under the age of 65 by the end of the survey period (column (1)), alive and observed for all interview rounds (column (2)), and have observed insurance status each month (column (3)). The pooled cohort combines individuals from cohorts 2014-15 through 2018-19. Age is reported as of December 31 of the second survey year. Prior adverse health events are diagnoses of coronary heart disease, diabetes, high blood pressure, heart attack, or stroke. This initial health status is determined by MEPS health assessments from the first survey year. An individual is considered to have a prior diagnosis if she (1) has ever received a diagnosis of the given condition *and* (2) received it when they were at least one year younger than the first observed age in the MEPS (i.e. at the first survey interview). Insurance status is reported as of January of the first survey year. Those with “other” insurance are individuals who report insurance but do not report coverage through Medicaid, employer/union coverage, or private exchanges. MEPS survey weights are applied for all demographic calculations.

<sup>a</sup> Education is defined for individuals 18 and older.

<sup>b</sup> Insurance types are not mutually exclusive. About 1% of the sample (N=573) is coded as having multiple types of coverage.



Table S3: SIPP 2017-2019 summary statistics, by sample restriction

	Under 65 (1)	& Alive and Observed all Rounds (2)	Analysis Sample (3)
No. of Obs.	15,068	14,893	14,893
Share of Unrestricted Sample	1.000	0.988	0.988
Age			
<18	0.264	0.246	0.246
18-30	0.207	0.212	0.212
31-50	0.311	0.318	0.318
51-64	0.234	0.239	0.239
Race			
non-Hispanic White	0.565	0.565	0.565
non-Hispanic Black	0.130	0.132	0.132
non-Hispanic Asian/Pacific Islander	0.065	0.066	0.066
Hispanic	0.206	0.204	0.204
Education <sup>a</sup>			
Below High School	0.095	0.095	0.095
High School	0.550	0.550	0.550
College or Above	0.355	0.355	0.355
Female	0.502	0.505	0.505
Insurance in first survey month <sup>b</sup>			
Medicaid	0.234	0.229	0.229
Employer provided	0.619	0.621	0.621
Private exchange	0.045	0.045	0.045
Other insurance coverage	0.045	0.046	0.046
Uninsured	0.102	0.104	0.104

Notes: Table presents sample size for the 2017-2019 SIPP cohort after sequentially applying three sets of sample restrictions: that individuals are under the age of 65 by the end of the survey period (column (1)), alive and observed for all 36 months of 2017-2019 (column (2)), and have observed insurance status each month (column (3); this last restriction is inconsequential for the SIPP sample, but we keep it to retain a parallel structure to the MEPS analysis). An individual's age is reported here as the highest age observed. "Private exchange" coverage represents any private coverage purchased through Healthcare.gov or a state-based exchange. Note that the full, unrestricted sample contains at least partial data for 68,241 individuals. The SIPP experienced substantial attrition during data collection in 2019 and 2020 due to a lapse in federal funding and COVID-19, such that the sample with non-missing longitudinal weights by the 2020 data release is only 20,967 individuals. In principle, the SIPP longitudinal weights are designed, to the extent possible, to compensate for attrition. For further discussion of non-response in these two data collection waves, see US Census Bureau [2021a](#) and US Census Bureau [2021b](#).

<sup>a</sup> Education is defined for individuals 18 and older.

<sup>b</sup> Insurance types are not mutually exclusive. 2.5% (N=381) of the sample is coded as having multiple types of coverage.