INSTITUTE FOR SOCIAL RESEARCH • SURVEY RESEARCH CENTER
PANEL STUDY OF INCOME DYNAMICS
UNIVERSITY OF MICHIGAN

Technical Series Paper #25-01

The Panel Study of Income Dynamics Linked Medicare and Medicaid Claims Data: An Update on Data Availability and Representativeness

Esther M. Friedman and Malak B. Kalasho
Institute for Social Research, University of Michigan

October 2025

This technical paper was supported by funding from the National Institute on Aging (R01-AG040213). We are grateful for research assistance from Qiyu Yang. This report is an update to Technical Series Paper #14-01

This paper can be cited as:

Friedman, EM; Kalasho, MB. (2025). The Panel Study of Income Dynamics Linked Medicare and Medicaid Claims Data: An Update on Data Availability and Representativeness. (Panel Study of Income Dynamics Technical Paper Series, #25-01).

Executive Summary

This paper provides an update on the 2014 PSID-Medicare technical report (Freedman et al, 2014) and includes a brief description of the Centers for Medicare and Medicaid (CMS) claims data in the Panel Study of Income Dynamics (PSID), including the updated process for obtaining these restricted data and new CMS data availability. PSID was one of the first national panel studies to link to Medicare claims during the 1990s. Linkages recently resumed with support from the National Institute on Aging through the newly established LINKAGE enclave. As of the writing of this report, linkages to more than 30 years of CMS claims data have been made for over 4,000 participants, these data are available under restricted data use contract. Comparisons between the PSID's Medicare eligible respondents in 2021 and beneficiaries in the Medicare Current Beneficiary Survey (MCBS) show that the two samples are very similar on a number of demographic and health characteristics, except for the youngest group (<65), where PSID respondents report less disability than MCBS respondents.

After a brief hiatus, in 2005, PSID resumed asking eligible respondents to provide their Medicare number and consent to link to their Medicare records. In 2005, approximately 60% of eligible respondents (i.e., study respondents who were 65 or older or reported receiving Medicare) gave permission to link CMS data to their survey responses; over the next eight waves of PSID, an additional 16% of 2005-eligible respondents agreed after being re-asked, so that by 2021 the cumulative percentage of the 2005 cohort providing consent reached over 76%. Altogether, permission to link to claims records was obtained from about 51% of all respondents asked in any of waves between 2005 and 2021.

The PSID Medicare eligible sample consists of 7,121 sample members who were age 65 or older in any year from 1991 through 2021; under age 65 but reported receiving Medicare in any year 1999 through 2021; or age 65 or older or had Medicare when they died and year of death was between 1991 and 2021. Approximately 60% of eligible cases altogether − including consenting respondents and decedents − were sent to the NIA Data LINKAGE Program and about 97% of those sent were successfully matched. Comparisons between the eligible sample and the sample successfully linked shows that the proportion of individuals in PSID's core ("SRC") sample and who are of white, non-Hispanic race-ethnicity are slightly over-represented among those linked. The samples line up fairly well in terms of health, though the linked sample having slightly higher reports of poor health and at least one limitation in their activities of daily living compared to the eligible sample.

1. Overview of Medicare and Medicaid Claims Linkage in the PSID

The PSID is the world's longest running national household panel study, collecting data on family composition changes, housing and food expenditures, marriage and fertility histories, employment, income, time use, health, consumption, wealth, philanthropic behavior, and more for more than five decades. The study significantly expanded the collection of health-related information in the 1990s and has continued expanding data to include additional information on the global pandemic of 2019, caregiving, and other health information. For additional information about the PSID, see McGonagle, Schoeni, Sastry, and Freedman (2012).

One important source of health information in the PSID comes from linked CMS claims data. PSID was one of the first national panel studies to link Medicare-eligible respondents to claims data starting in 1990. In that year, Medicare-eligible respondents were asked to give written consent to link to one year of claims. Through 1995 consent was renewed annually and another 1 -year linkage was made. Research based on these data has provided important insights into predictors of health insurance status, service utilization and cost (e.g., Lillard & Farmer 1997; Lillard & Rogowski 1995; Lillard et al. 1997).

After a brief hiatus, the linkage of CMS claims to eligible respondents resumed in 2005. Currently, with support from the National Institute on Aging, data were moved to the NIA LINKAGE enclave, a secure cloud-based statistical analysis environment where PSID data may be securely linked to an expanded collection of CMS data under restricted data contract. The focus of this report is on data available for Medicare-eligible PSID respondents through the 2021 wave. This report updates a prior technical report on data linkages through 2013 (Freedman et al, 2014). The most up-to-date information about obtaining the data and available linkages may be found on the PSID webpage, here:

https://simba.isr.umich.edu/restricted/ProcessReqCMSData.aspx

There are several features of the PSID claims data that make it especially valuable for aging-related research. More than thirty years of Medicare utilization/spending data are embedded in the PSID, a unique nationally representative dataset that includes more than 50 years of prospective life course measures of economic well-being. Having CMS claims linked to prospectively collected data allows the exploration of how early- and mid-life social, economic, health and environmental factors influence later life medical care utilization. In addition, distributional issues related to the financing and benefit structure of the Medicare program may be explored in- depth with the PSID data, and can take advantage of both cross-sectional and life course measures of economic well-being.

2. Benchmarking PSID-based Medicare-eligible population

We began by benchmarking the 2021 PSID against the 2021 Medicare Current Beneficiary Survey (MCBS) public use files. The MCBS provides comprehensive data on health and functional status, health care expenditures, and health insurance for Medicare beneficiaries. The MCBS design is longitudinal, and interviews each sample person three times a year, with membership in the study over four years. Both samples are limited to the non-institutionalized population ages 18 and over. PSID is further limited to reference persons and spouse/partners who are Medicare-eligible (either 65 and over or report receiving Medicare since the last wave of data collection). MCBS public-use data includes all adults enrolled in Medicare. In the PSID, Medicare-eligible respondents account for 10% of non-institutionalized reference persons and spouse/partners. Medicare beneficiaries comprise 18% of the total U.S. population.

The 2021 PSID sample of Medicare-eligible respondents has a lower percentage oldest old (ages 75+) and a greater percentage of individuals who are less than age 65 than does the Medicare population, and a greater percentage in the 65-74 age range (Table 1). The PSID sample is fairly comparable to the Medicare population. PSID is slightly more female (60% in PSID and 55% in the MCBS), has similar marital status (52 % vs. 53%), has a similar racial distribution (72% vs. 75% non-Hispanic white), and has higher educational attainment (5% in the PSID have less than a high school education compared to 12% in the MCBS).

The overall rate of the PSID sample reporting at least one limitation in their activities of daily living – including eating, dressing, getting in/out of bed, taking a bath or shower, and using the toilet – is comparable to the MCBS overall (24% vs. 26%), with differences for the youngest group, less than age 65 (36% in PSID vs. 51% in MCBS). Finally, a slightly greater percentage of sample members in PSID had at least one overnight hospital stay in the past year (11% in PSID vs. 6% in MCBS).

Table 1. Comparison between PSID 2021 Medicare-Eligible Reference Persons and Spouses/Partners and MCBS 2021

	2021 PSID Medicare Eligible Reference Persons and Spouse/Partners (Non-Institutionalized)				2021 Medicare Current Beneficiary Survey (MCBS) (Non-Institutionalized)			
	All	Age Group			All	Age Group		
		<65	65-74	75+		<65	65-74	75+
% Eligible	10.2	19.5	62.5	18.0	18.4	12.6	52.0	35.4
Of the Medicare Eligible:								
% Female	60.2	60.5	60.3	59.5	54.7	49.8	53.8	57.7
% Married	51.6	42.0	56.0	46.8	53.1	33.8	61.3	47.8
% Non-Hispanic White	72.3	52.7	76.2	80.0	74.9	60.9	75.0	79.7
% < high school education	5.3	9.9	2.7	9.2	11.7	20.4	8.9	12.8
% at least 1 ADL*	23.6	36.1	18.5	27.6	26.1	50.5	17.6	29.9
% with at least 1 hospital stay	11.3	15.0	9.4	13.9	6.3	7.0	4.9	8.1

Cross-sectional weights are used for all analyses.

*Includes difficulty bathing or showering, dressing, eating, getting in/out of bed/chair, walking, and using the toilet.

3. Gaining consent to link 2005 – 2021

Beginning in 2005, PSID resumed asking eligible respondents (i.e., who were either age 65 or older or younger adults who report receiving Medicare) to provide their Medicare number and consent to cumulatively link to all their CMS records (rather than a single year). Consent was obtained orally during the Core PSID telephone interview or via the self-administered web mode. Information on consent was preloaded from one wave to the next so that respondents who refused or provided an unusable number were re-asked permission.

As shown in Table 2, about 60% of eligible respondents gave permission to link Medicare data to their survey responses at the initial request in 2005; over the next eight waves, approximately an additional 16% of 2005-eligible respondents agreed after being re-asked, so that by 2021 the cumulative percentage of the 2005 cohort providing consent was more than 76%.

Table 2. Gaining consent from the 2005-Medicare eligible cohort (N=909) over 9 waves of data collection

	Providing consent	
Wave of data collection	Cumulative N	Cumulative %
2005	549	60.4
2007	650	71.5
2009	674	74.1
2011	683	75.1
2013	686	75.5
2015	691	76.0
2017	693	76.2
2019	695	76.5
2021	695	76.5

Note: Excludes decedents.

Table 3 shows the proportion of Medicare-eligible respondents giving consent across all nine waves and by wave of data collection. 3790 participants were ever eligible to be asked consent and 1917 provided permission to link (51%). In 2007-2021, respondents who previously refused or did not provide a valid Medicare number were re-asked and newly eligible respondents were asked for the first time. The percentage of eligible respondents providing permission to link fell with each wave from 38% in 2007 to 10% in 2021.

Table 3. Proportion of eligible respondents giving consent by wave

		Providing consent in each wave	
Wave of data collection	Total eligible in each wave	N	%
2005	909	549	60.4
2007	619	237	38.3
2009	528	168	31.8
2011	635	135	21.3
2013	735	193	26.3
2015	813	176	21.6
2017	960	178	18.5
2019	1052	165	15.7
2021	1167	116	9.9
Cumulative 2005-2021	3790	1917	50.6%

Note: Excludes decedents.

4. Linking to claims data

Sample members who met the following criteria were considered eligible for linking:

- age 65 or older in any year from 1991 through 2021;
- under age 65 but reported receiving Medicare in any year from 1999 through 2021[1];
- age 65 or older when they died or reported receiving Medicare and year of death was between 1991 and 2021.

Living sample members who were 65 or older or reported Medicare coverage were eligible for linkage only if they were the respondent (and could provide consent); therefore, not all individuals with Medicare were necessarily eligible for linkage. For deceased individuals who either were never asked for consent or never refused to provide consent for linkages, eligibility was determined by age at death or reports of Medicare coverage. Of these 7,121 Medicare eligible sample members, approximately 60% (n=4,274) were sent for linkage. The sample sent for linking consisted of 1,907 respondents (out of 3790 eligible to be asked for consent; 50%) providing permission and a usable Medicare number for the linkage in the waves between 2005 and 2021, and an additional 2,367 decedents from 1991 or later who were at least age 65 when they died or reported receiving Medicare (who had not refused permission between 1991 and 1995).

About 40% (n=2,847) of the sample was not sent for linkage. This was due to any of the following reasons, individuals: were not asked for permission (i.e., was eligible in a year when permission was not asked); were asked but did not provide permission; provided permission but did not know their Medicare number or provided an unusable number (e.g., missing digits); refused permission prior to 2005; or were missing identifying information required to perform the linkage (i.e., Medicare number for living respondents; other identifying information for deceased ones). Beginning in 2019, individuals who provided consent to linkage but had an invalid number were still sent for linkage. Individuals with missing Medicare numbers were not sent for linkage as of 2021.

Among those sent in 2021, 97% were linked to CMS data (n=4,162).[2] Nearly all respondents and approximately 96% of decedents were successfully linked. Table 4 compares characteristics of the eligible sample (n=7,121) and three groups: the sample sent for linking and linked (n=4,162), the sample sent for linking and not linked (n=112), and the sample not sent for linking (n=2,847).

Sample Type. There are four types of sample membership in the PSID. The first two are defined by the original 1968 sample which consisted of families who were part of the original low-income Survey of Economic Opportunity (SEO) oversample and families who were part of the original national probability sample drawn by the Survey Research Center (SRC) at the

---

[1] The year 1999 is used as the starting year for the under 65 eligibility criteria because the PSID began collecting data on type of health insurance coverage in that year.

[2] Respondents were matched using: First and last name, Social Security number, claim account number, beneficiary identifier, year of birth, month of birth, day of birth, gender, state, city, and zip code. For deceased individuals, the year, month, and day of death are also used. Each respondent is assigned a match confidence level: *Strong, Strong-Fair, Strong-Fair-Weak,* which is available in the crosswalk file in the enclave. There were 3,937 strong matches, 218 fair matches, and 7 weak matches. Users may choose to exclude weak matches.

University of Michigan. In 1990, a third sample was added with the inclusion of approximately 2,000 Latino households originally from Mexico, Puerto Rico, and Cuba. Due to budgetary reductions, this sample was dropped after 1995. In 1997, an immigrant refresher effort added 511 families that had recently immigrated to the U.S.; an additional 615 families who immigrated after 1997 were added in 2017.

Table 4. Sample characteristics of eligible cases
by whether they were sent for linkage and linkage outcome

	Total eligible sample	Sample sent (n=4274)		Sample not sent
	(n=7121)	Linked (n=4162)	Not Linked (n=112)	(n=2847)
Sample membership				
SRC	58.1	67.2	49.1	45.1
SEO	30.4	26.6	40.2	35.7
Immigrant	4.8	4.0	7.1	6.0
Latino	6.6	2.2	3.6	13.2
Total	100.0	100.0	100.0	100.0
Gender				
Male	42.1	44.1	43.8	39.1
Female	57.9	55.9	56.2	60.9
Total	100.0	100.0	100.0	100.0
Race and Ethnicity				
Non-Hispanic white	54.9	63.4	53.0	43.4
Non-Hispanic black	32.2	29.5	39.4	35.7
Hispanic	10.6	5.0	4.5	18.2
Other	2.3	2.0	3.0	2.6
Total	100.0	100.0	100.0	100.0
Education				
Less than high school	28.5	28.3	51.6	28.0
Health				
Poor general health	71.7	76.7	65.5	64.7
At least 1 ADL	60.2	68.5	71.8	49.3

Notes: Measures of race, education, and health were obtained from the PSID's Social, Health, and Economic Longitudinal File (PSID-SHELF). The proportion of missing values in the total eligible sample was: race, 6.4%; education, 1.6%; general health, 2.1%, and ADL, 9.8%.

As shown in Table 4, more than half of the total eligible sample falls into the SRC sample (58%), about 30% into the SEO sample, 5% in the Immigrant sample, and 7% in the Latino sample. Within the sample successfully linked, the SRC sample is overrepresented (67%) − and the SEO sample is somewhat underrepresented (27%) relative to the eligible sample. Because the Latino sample was dropped from the PSID in 1995, only about 2% of the sample sent for linking was drawn from the Latino sample.

Gender and race. The eligible and linked samples line up quite well by gender and reasonably well by race. Women make up about 58% of the eligible sample and 56% of those successfully linked. Whites make up about 55% of the eligible sample and 63% of those successfully linked.

Education and health. The eligible and linked samples also line up well by education and health. About 28% of each group has less than a high school education. The eligible sample shows comparable, but slightly lower rates of poor health (72% vs. 77%) and ADL limitations (60% vs. 69%) than the linked sample.

5. Available claims data

More than 30 years of claims data (1991-2022) have been linked to 30 years of PSID data (1991-2021) and are currently available on the LINKAGE enclave. Claims data begin with a respondent's first service after Medicare entitlement, which is at age 65 or when qualification for disability insurance is met.

The 2023 PSID data are already being processed for linkage and there are plans to continue expanding linkages every two years when new waves of PSID data are available. With each wave of PSID data, CMS data are also updated to the most recent claims information available.

As of 2021, a more extensive set of data have been made available for linking. Below we list the CMS claims data currently available under data use agreement to qualified researchers (Table 5), but these will be updated as new data become available. For the most up to date information on data availability, please see: https://simba.isr.umich.edu/restricted/ProcessReqCMSData.aspx. To learn more about the available CMS data and to download the NIA CMS Data Dictionary, please visit: https://www.nia.nih.gov/research/dbsr/nia-data-linkage-program-linkage/available-data.

Table 5. CMS data available for linkage

Medicare Enrollment Data
Denominator (DN)
Master Beneficiary Summary File (MBSF) Base
Medicare Summary Files
Master Beneficiary Summary File (MBSF): Chronic Conditions
Master Beneficiary Summary File (MBSF): Cost & Utilization Master Beneficiary Summary File (MBSF): Other Chronic or Potentially Disabling Conditions
Medicare Part A & B Claims Data
Medicare Carrier (PB) Claims
Medicare Durable Medical Equipment (DM) Claims
Medicare Home Health (HH) Claims
Medicare Hospice (HS) Claims
Medicare Inpatient (IP) Claims
Medicare Outpatient (OP) Claims
Medicare Skilled Nursing Facility (SN) Claims
MedRIC-Built Medicare Provider Analysis & Review (MedPAR)
Medicare Part C Claims Data
Medicare Carrier Encounter Claims
Medicare Durable Medical Equipment (DME) Encounter
Medicare Home Health Agency (HH) Encounter Claims
Medicare Inpatient (IP) Encounter Claims
Medicare Outpatient (OP) Encounter Claims
Medicare Skilled Nursing Facility (SNF) Encounter Claims
Medicaid Data
Medicaid Analytic eXtract (MAX) Personal Summary (PS) Enrollment Data
Medicaid Analytic eXtract (MAX) Inpatient (IP) Claims
Medicaid Analytic eXtract (MAX) Long Term Care (LT) Claims
Medicaid Analytic eXtract (MAX) Other Services (OT) Claims
Medicaid Analytic eXtract (MAX) Prescription Drug (RX) Data
TMSIS Analytic Files (TAF) Demographic and Eligibility (DE) Enrollment Data
TMSIS Analytic Files (TAF) Inpatient (IP) Claims
TMSIS Analytic Files (TAF) Long Term Care (LT) Claims
TMSIS Analytic Files (TAF) Other Services (OT) Claims

TMSIS Analytic Files (TAF) Pharmacy (RX) Data
Other Data
Medicare Part D Medication Therapy Management (MTM)
Healthcare Effectiveness Data and Information Set (HEDIS
Inpatient Rehab Facility-Patient Assessment Instrument (IRF-PAI)
Long Term Care Minimum Data Set (MDS)
Medicare Part D Prescription Drug Event (PDE)
Outcome and Assessment Information Set (OASIS)

References

Centers for Medicare & Medicaid Services. Medicare Current Beneficiary Survey, 2021 Survey File Public Use File data. Baltimore, MD: U.S. Department of Health and Human Services, 2023.

Cubanski, J., Huang, J., Damico, A., Jacobson, G. and Neuman, T. (2010). *The Medicare Chartbook,* Fourth Edition.

Freedman, VA; McGonagle, K; Andreski, P. (2014). The Panel Study of Income Dynamics Linked Medicare Claims Data. (Panel Study of Income Dynamics Technical Paper Series, #14-01).

Lillard, L., and Farmer, M.M. (1997). Linking Medicare and National Survey Data. *Annals of Internal Medicine,* 127 (8, Part 2): 691-695.

Lillard, L., and Rogowski, J. (1995). Does Supplemental Private Insurance Increase Medicare Costs? RAND, Labor and Population Program.

Lillard, L.; Rogowski, J., and Kington, R. (1997). Long-Term Determinants of Patterns of Health Insurance Coverage in the Medicare Population. *The Gerontologist*. 37(3):314323.

McGonagle, K.A., Schoeni, R.F., Sastry, N., and Freedman, V.A. (2012). The Panel Study of Income Dynamics: Overview, Recent Innovations, and Potential for Life Course Research. *Longitudinal and Life Course Studies,* 3(2): 268-284. PMCID: PMC3591471

Panel Study of Income Dynamics, public-use dataset, Produced and distributed by the Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI: 2025.

Pfeffer, Fabian T., Davis Daumler, and Esther Friedman. *PSID-SHELF, 1968–2021: The PSID's Social, Health, and Economic Longitudinal File (PSID-SHELF), Beta Release.* Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2025-02-23. DOI:10.3886/E194322.