PSID A national study of socioeconomics and health
over lifetimes and across generations

Technical Series Paper #14-01

The Panel Study of Income Dynamics
Linked Medicare Claims Data

Vicki A. Freedman, Katherine McGonagle, Patricia Andreski
Survey Research Center, Institute for Social Research
University of Michigan

February 2014

This technical paper was supported by funding from the National Institute on Aging (R01-AG040213).

Executive Summary

This paper provides a brief description of the Medicare claims data in the Panel Study of Income Dynamics (PSID). 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. As of 2013, linkages to nearly 20 years of Medicare claims have been made for over 2,000 participants. These data are available under restricted data use contract.

Comparisons between the PSID's Medicare eligible sample in 2007 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, but at the oldest ages PSID respondents report less disability than MCBS respondents.

In 2005, approximately 60% of eligible respondents gave permission to link Medicare data to their survey responses; over the next three waves of PSID, an additional 15% of 2005-eligible respondents agreed after being re-asked, so that by 2009 the cumulative percentage of the 2005 cohort providing consent reached 74% (75% by 2011). Altogether, permission to link to claims records was obtained from about 61% of all respondents asked in any of the four waves (2005-2011).

The PSID Medicare eligible sample consists of 5,817 sample members who were age 65 or older in any year from 1991 through 2009; under age 65 but reported receiving Medicare in any year 2005 through 2009; or age 65 or older when they died and year of death was between 1991 and 2009. Approximately 38% of eligible cases altogether were sent to the Medicare Research Information Center (MedRIC) for linkage to Medicare data and about 95% 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 are slightly over-represented among those linked.

1. Overview of Medicare 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 nearly five decades. The study began expanding the collection of health-related information in the 1990's, and by 2011, nearly 19 minutes of the 89-minute questionnaire was devoted to this domain. For additional information about the PSID, see McGonagle, Schoeni, Sastry, and Freedman (2012).

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).

Recognizing that Medicare claims data would be a valuable addition to the expanded content collected on health in the Core interview, the linkage of Medicare claims to eligible respondents resumed in 2005, with support from the National Institute on Aging to, continue these linkages through 2015. As of 2013, the Medicare Research Information Center (MedRIC) has made linkages to 20 years of Medicare claims for over 2000 PSID respondents. To access these data files, researchers are required to complete separate data use agreements with the Centers for Medicare and Medicaid Services, and the University of Michigan. Additional information about obtaining the data may be found here: http://simba.isr.umich.edu/restricted/RestrictedUse.aspx

There are several features of the PSID Medicare claims data that make it especially valuable for aging-related research. Nearly 20 years of Medicare utilization/spending data are embedded in a unique nationally representative dataset that includes more than 40 years of prospective life course measures of economic well-being. Having Medicare claims linked to prospectively collected data will allow 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 benchmarked the PSID 2007 Medicare eligible sample against the Medicare Chartbook (Cubanski et al. 2010). The Chartbook includes data for Medicare beneficiaries collected in the 2006 Medicare Current Beneficiary Survey (MCBS). 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.

The 2007 PSID Medicare eligible sample has fewer oldest old (ages 85+) and fewer less than age 65 than the Medicare population and about 6 percentage points more in the 65-84 age range (79% in the PSID and 73% in the MCBS; see Table 1). The PSID sample is slightly more female (60% in PSID and 55% in the MCBS). Compared to the MCBS sample, those in the PSID are less likely to be married (48% vs. 54%), have similar racial distribution (76% vs. 78% non-Hispanic white), and have higher educational attainment (21% in the PSID have less than a high school education compared to 27% in the MCBS).

The overall rate of 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 somewhat lower than in the MCBS sample across all age groups (18% vs. 29%) and differences are greatest for the 85+ (23% in PSID vs. 48% in MCBS). Finally, approximately 21% of sample members in both groups had at least one overnight hospital stay in the past year.

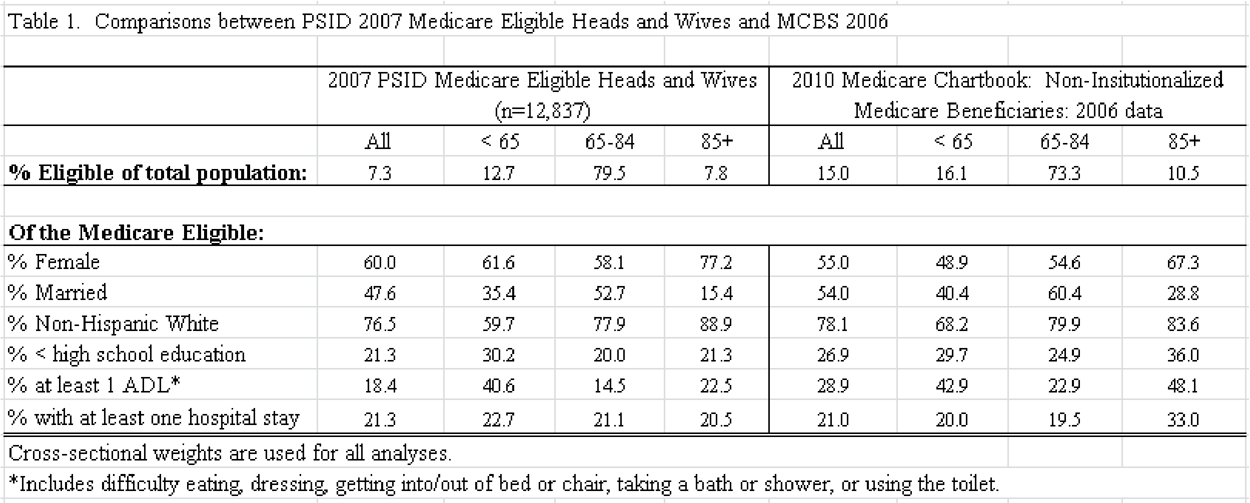

Table 1. Comparisons between PSID 2007 Medicare Eligible Heads and Wives and MCBS 2006

	2007 PSID Medicare Eligible Heads and Wives (n=12,837)				2010 Medicare Chartbook: Non-Insitutionalized Medicare Beneficiaries: 2006 data			
	All	< 65	65-84	85+	All	< 65	65-84	85+
% Eligible of total population:	7.3	12.7	79.5	7.8	15.0	16.1	73.3	10.5
Of the Medicare Eligible:								
% Female	60.0	61.6	58.1	77.2	55.0	48.9	54.6	67.3
% Married	47.6	35.4	52.7	15.4	54.0	40.4	60.4	28.8
% Non-Hispanic White	76.5	59.7	77.9	88.9	78.1	68.2	79.9	83.6
% < high school education	21.3	30.2	20.0	21.3	26.9	29.7	24.9	36.0
% at least 1 ADL*	18.4	40.6	14.5	22.5	28.9	42.9	22.9	48.1
% with at least one hospital stay	21.3	22.7	21.1	20.5	21.0	20.0	19.5	33.0

Cross-sectional weights are used for all analyses.

*Includes difficulty eating, dressing, getting into/out of bed or chair, taking a bath or shower, or using the toilet.

3. Gaining consent to link 2005 – 2011

Beginning in 2005, PSID resumed asking eligible respondents (who were either age 65 or older or receiving Medicare through the Disability Insurance (DI) program) to provide their Medicare number and consent to link to all their Medicare records (rather than a single year). Consent was obtained orally during the Core PSID telephone interview. 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, 60.4% of eligible respondents gave permission to link Medicare data to their survey responses at the initial request in 2005; over the next three waves, an additional 14.7% of 2005-eligible respondents agreed after being re-asked, so that by 2011 the cumulative percentage of the 2005 cohort providing consent reached 75.1%.

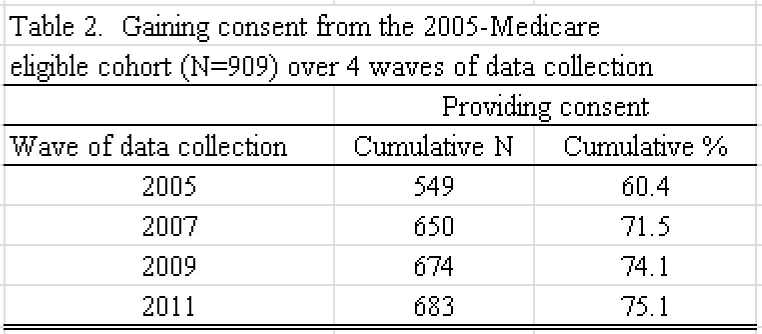

Table 2. Gaining consent from the 2005-Medicare eligible cohort (N=909) over 4 waves of data collection

	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

Table 3 shows the proportion of Medicare eligible respondents giving consent across all four waves and by wave of data collection. 1787 participants were ever eligible to be asked consent and 1089 provided permission to link (60.9%). In 2007-2011, respondents who previously refused 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 38.3% in 2007 to 21.3% in 2011.

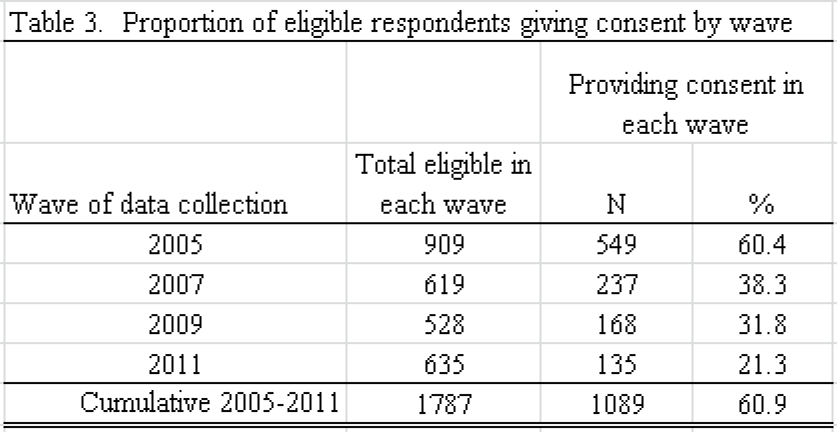

Table 3. Proportion of eligible respondents giving consent by wave

		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
Cumulative 2005-2011	1787	1089	60.9

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 2009;
- under age 65 but reported receiving Medicare in any year 2005 through 2009;
- age 65 or older when they died and year of death was between 1991 and 2009.

Of these 5,817 Medicare eligible sample members, approximately 38% (n=2,185) were sent to MedRIC for linkage. The sample sent for linking consisted of 950 respondents (out of 1512 eligible; 63%) providing permission for the linkage in PSID waves 2005, 2007, and 2009 and an additional 1,235 decedents from 1991 or later who were at least age 65 when they died (who had not refused permission between 1991 and 1995).

About 62% (n=3,632) of the sample was not sent for linkage either because they were not asked for permission or were asked but never gave permission, or they refused permission prior to 2005, or because there was a lack of adequate identifying information to perform a linkage.

Among those sent, 95% were linked (n=2,078).[1] Table 4 compares characteristics of the eligible sample (n=5,817) and three groups: the sample sent for linking and linked (n=2,078), the sample sent for linking and not linked (n=107), and the sample not sent for linking (n=3,632).

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 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.

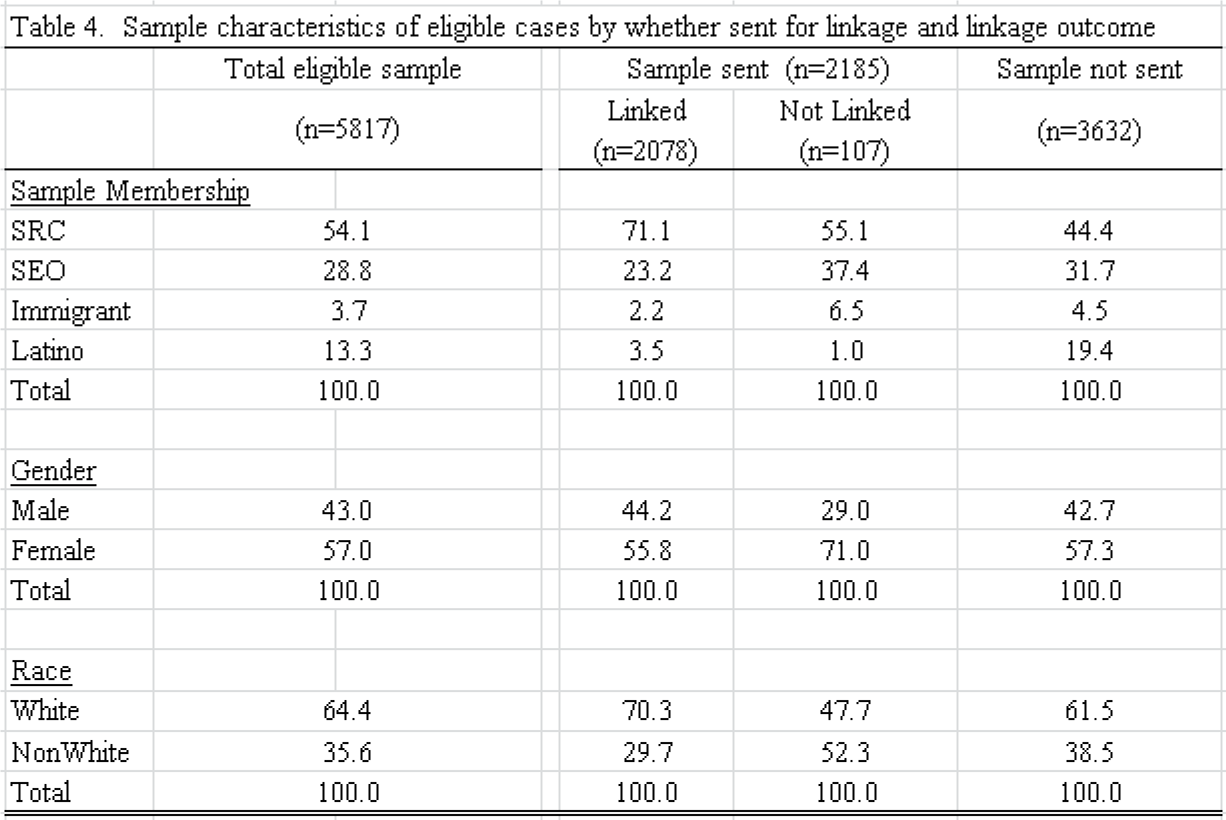

Table 4. Sample characteristics of eligible cases by whether sent for linkage and linkage outcome

	Total eligible sample	Sample sent (n=2185)		Sample not sent
	(n=5817)	Linked (n=2078)	Not Linked (n=107)	(n=3632)
Sample Membership				
SRC	54.1	71.1	55.1	44.4
SEO	28.8	23.2	37.4	31.7
Immigrant	3.7	2.2	6.5	4.5
Latino	13.3	3.5	1.0	19.4
Total	100.0	100.0	100.0	100.0
Gender				
Male	43.0	44.2	29.0	42.7
Female	57.0	55.8	71.0	57.3
Total	100.0	100.0	100.0	100.0
Race				
White	64.4	70.3	47.7	61.5
NonWhite	35.6	29.7	52.3	38.5
Total	100.0	100.0	100.0	100.0

As shown in Table 4, more than half of the total eligible sample falls into the SRC sample (54%), about 29% into the SEO sample, 4% are in the Immigrant sample, and 13% are in the Latino

---

[1] Respondents were matched using: Social Security number, claim account number, beneficiary identifier, year of birth, month of birth, day of birth, and gender. Weights were assigned by MedRIC to each identifier and then summed. There were 2,049 cases that matched with 7 points or higher and 29 cases that matched with fewer than 7 points.

sample. Within the sample successfully linked, the SRC sample is over represented (71%) – and the SEO sample is underrepresented (23%) relative to the eligible sample. Because the Latino sample was dropped from the PSID in 1995, only about 4% 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 57% of the eligible sample and 56% of those successfully linked. Whites make up about 64% of the eligible sample and 70% of those successfully linked.

4. Available claims data

Up to 19 years of claims data spanning the years 1991 through 2010 are available for the linked PSID sample. 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. Linked claims files made available by MedRIC under data use agreement to qualified researchers are listed below.

File	Type of file	Information included
Denominator file	Beneficiary enrollment information in a calendar year.	Beneficiary unique identifier, monthly entitlement indicators (A/B/Both), reasons for entitlement state buy-in indicators, and monthly managed care indicators (yes/no).
Inpatient SAF	Final action claims data submitted by inpatient hospital providers for reimbursement of facility costs.	ICD-9 diagnosis and procedures code
Skilled Nursing Facility SAF	Final action claims data submitted by SNF providers.	ICD-9 diagnosis and procedure codes, dates of service, and reimbursement amounts.
Outpatient SAF	Final action claims data submitted by institutional outpatient providers.	ICD-9 diagnosis and procedure codes, CMS Common Procedure Coding System (HCPCS), dates of service, and reimbursement amounts.
Home Health Agency SAF	Final action claims data submitted by HHA providers.	Number of visits, type of visit (skilled-nursing care, home health aides, physical therapy, speech therapy, occupational therapy, and medical social services), ICD-9 diagnosis codes, dates of visits, and reimbursement amounts.
Carrier SAF (old file name: Physician/Supplier Part B)	Final action claims data submitted by non-institutional providers.	ICD-9 diagnosis and procedure Codes, CMS Common Procedure Coding System (HCPCS), dates of service, and reimbursement amounts.
Hospice SAF	Final action claims data submitted by Hospice providers.	Level of hospice care received (e.g., routine home care, inpatient respite care), ICD-9 terminal diagnosis, dates of service, and reimbursement amounts.
Durable Medical Equipment SAF	Final action claims data submitted by Durable Medical Equipment providers.	CD-9 diagnosis codes, services provided (CMS Common Procedure Coding System (HCPCS)), dates of service, and reimbursement amounts.
MEDPAR file	Inpatient hospital and skilled nursing facility (SNF) and final action stay records.	Inpatient "stay" record summarizes all services rendered to a beneficiary from the time of admission to a facility through discharge. Each record may represent one claim or multiple claims, depending on length of beneficiary's stay and amount of inpatient services used throughout.

Also included are Beneficiary Annual Summary File (BASF) and the Beneficiary Quarterly Summary File (BQSF), which are summarized at the beneficiary level for different time periods.

Both summary files condense Medicare Part A and B claims and enrollment information into one observation per beneficiary in a specific period of time. Summary measures related to health and medical expenditures for each type of service and payer, along with utilization measures associated with these services, are provided. Potentially identifiable geographic information has been suppressed.

Linked claims data will be added to the archive as Medicare numbers continue to be collected from eligible respondents during waves 2011, 2013 and 2015 of PSID. MedRIC will continue to be provided updated finder files for linking to claims, to make the linked files available to users through restricted data use contracts, and to continue to provide user support.

References

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

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):314-323.

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