HHS Public Access

Identifying Cohabiting Couples in Administrative Data: Evidence from Medicare Address Data

Sasmira Matta, MHS[a,b], Joanne W. Hsu, PhD[c], Theodore J. Iwashyna, MD, PhD[d,e], Micah Y. Baum, BA[f], Kenneth M. Langa, MD, PhD[e,g], Lauren Hersch Nicholas, PhD[f,h]

[a]The Wharton School, University of Pennsylvania, Department of Health Care Management and Economics, 3641 Locust Walk, Philadelphia, PA 19106

[b]Leonard Davis Institute of Health Economics, 3641 Locust Walk, Philadelphia, PA 19106

[c]Board of Governors of the Federal Reserve System, Constitution Ave NW & 20th Street Northwest, Washington, DC 20551

[d]University of Michigan Pulmonary and Critical Care Medicine, 1500 E Medical Center Dr SPC 5360, Ann Arbor, MI 48109

[e]University of Michigan Institute for Social Research, 426 Thompson Street, Ann Arbor, MI 48104

[f]Johns Hopkins Bloomberg School of Public Health, Department of Health Policy and Management, 624 N. Broadway, Baltimore, MD 21205

[g]University of Michigan Department of Internal Medicine, 1500 E Medical Center Dr SPC 5368, Ann Arbor, MI 48109

[h]The Johns Hopkins University School of Medicine, 733 N. Broadway, Baltimore, MD 21205

Abstract

Marital status is recognized as an important social determinant of health, income, and social support, but is rarely available in administrative data. We assessed the feasibility of using exact address data and zip code history to identify cohabiting couples using the 2018 Medicare Vital Status file and ZIP codes in the 2011–2014 Master Beneficiary Summary Files. Medicare beneficiaries meeting our algorithm displayed characteristics consistent with assortative mating and resembled known married couples in the Health and Retirement Study linked to Medicare claims. Address information represents a promising strategy for identifying cohabiting couples in administrative data including healthcare claims and other data types.

Marital status is widely recognized as an important social determinant of health (National Academies of Sciences 2017, Wood et al. 2007). Compared to the unmarried or widowed, members of married couples live longer, exhibit better physical and mental health, and receive less expensive and more efficient health care (Calvillo–King et al. 2013, National Academies of Sciences 2017, Waite and Gallagher 2000, Wood et al. 2007, Wu et al. 2008, Young et al. 1998). As the Centers for Medicare and Medicaid Services (CMS) increasingly move towards paying for outcomes, the availability of a partner (potential caregiver) has emerged as an important variable for risk adjustment (National Academies of Sciences 2017). Despite the numerous health and economic benefits of marriage, administrative data rarely includes marital status.

Health services researchers have had a long-standing interest in identifying married couples using administrative data (Allison and Christakis 2006, Christakis and Allison 2006, Christakis and Allison 2009, Christakis and Iwashyna 1998, Christakis et al. 2002, Elwert and Christakis 2006, Elwert and Christakis 2008a, Elwert and Christakis 2008b, Gilden et al. 2017, Iwashyna and Christakis 2003, Iwashyna et al. 1998, Jin and Chrisatakis 2009, Smith and Christakis 2009, Subramanian et al. 2008, Zivin and Christakis 2007, Iwashyna et al. 2002). Health Insurance Claim numbers (HIC), which are frequently derived from Social Security Numbers, were previously used to identify couples in Medicare claims data (Iwashyna et al. 1998). Unfortunately, this method is no longer as useful because fewer couples claim benefits using a single partner's earnings history and CMS no longer release HICs.

Ascertaining marital status without appropriate legal documentation or self-reported data is not possible. Current trends in family demography indicate that there has been decreased entry in to marriage, delayed entry to marriage, and increased cohabitation (Smock and Schwartz 2020). Regardless of legal marital status, cohabiting partners still act as caregivers for each other (Monden 2007) and 95% of cohabiting Medicare beneficiaries in the nationally representative Health and Retirement Study are legally married.

Coresidence via the use of exact address data represents a potential new approach to identify couples in administrative data. Such information can be utilized to improve financing and enhance planning and evaluation of health care services. This information also extends beyond the realm of medical claims and can be of use to researchers interested in couples' dynamics within administrative data. In this paper, we assessed the feasibility and validity of using geographic identifiers as an alternative way of identifying co-residing couples in Medicare claims data.

METHODS

Based on trends in marriage among older adults in the Health and Retirement Study (HRS), we hypothesized that many elderly married couples could be identified if they shared a

current address as well as four years' worth of ZIP code history. We tested this address-matching algorithm in a sample of Medicare beneficiaries age 70 or older in 2014 using ZIP codes from 2011 – 2014 Master Beneficiary Summary Files and exact addresses from a June 2018 Medicare Vital Status file. The Vital Status File contains the current address for living beneficiaries and the last known address for the deceased. Addresses were cleaned and harmonized to facilitate exact text matches using geocoding software (ESRI 2019, Wilson et al. 2008) (see Appendix A).

To test whether our algorithm was correctly identifying cohabiting couples, we compared the characteristics of the address-based couples (Medicare Couples) to characteristics of those that were cohabiting based on the current address but did not have a shared ZIP code history (Medicare Non-Couples) and randomly assigned pairs of beneficiaries who lived in the same census block (Random Pairs).

The Medicare Vital Status file contained 14,568,245 Medicare beneficiaries residing in the 50 United States and DC who were at least 70 in 2014 and living in a subset of US zip codes.

We also tested whether our identified couples were similar in prevalence and demographics to known couples in other data. Because we used Medicare data from elderly beneficiaries, our method would necessarily exclude those whose spouses were not eligible for Medicare due to age or other factors. Given that our sample of Medicare beneficiaries was non-random, a standard comparison to a large household survey such as the Current Population Survey or the American Community Survey would not provide an accurate benchmark. We instead identified two samples of respondents in the HRS (Juster and Suzman 1995, Sonnega et al. 2014).

Each sample contained all beneficiaries with Alzheimer's disease and related dementias (ADRD) who were at least 70 years old and present in 2014 claims data, and a 3:1 sample of beneficiaries without ADRD following the parameters of our Medicare sample. For the first sample (HRS Couples 1), who chose non-ADRD beneficiaries from counties with the lowest Medicare Advantage (MA) enrollment to match the Medicare addresses selection process. For the second sample (HRS Couples 2), we chose non-ADRD beneficiaries with a unique ZIP code/year of birth combination in HRS to improve the likelihood of a unique respondent-spouse match. To more accurately assess our couples algorithm, we restricted HRS couples to those where both members were 70 or older and the couple had been married/partnered for at least five years and was not living with other individuals age 65 or older. The HRS samples contained 2,134 and 2,002 Medicare beneficiaries that mirrored the demographic and disease composition of the beneficiaries in the Vital Status file.

RESULTS

Sample

Table 1 presents characteristics of Medicare beneficiaries disaggregated by the number of times an exact address appeared in the Vital Status file. 8,554,309 beneficiaries (59%) resided in one-beneficiary households, 3,997,492 (27%) in two-beneficiary households, and

2,015,802 (14%) in three or more beneficiary households. The composition of the two-beneficiary households was approximately half female and half male and was predominantly white, with a mean age of 79 years. Single and 3+ beneficiary households were older, and more female compared to two-beneficiary households. These characteristics are consistent with the single person households including widows and widowers who have outlived their spouses and the 3+ including those living in nursing homes and other supportive quarters (Barford et al. 2006, Luy and Minagawa 2014, Case and Paxson 2005).

Couples Prevalence

Of the 3,997,608 beneficiaries classified to live in two-beneficiary households, 3,654,408 (91.4% of beneficiaries in two-beneficiary households, 25.1% of our sample) were identified as cohabiting couples because they also shared four years' worth of ZIP code history. There were notable differences between those who did and did not meet the ZIP code restriction as well as those who were randomly paired (Appendix B). 96% of Medicare Couples were different sex and 82% were within 5 years of each other compared to 63% and 50% in Medicare Non-Couples, and 50% and 49% in the Random Pairs. Race differences across the three Medicare samples were less pronounced than differences in age and sex, likely reflecting residential segregation patterns.

Benchmarking to the HRS

In the HRS benchmark samples, 39.5% and 37.1% of observations were in partnerships. Of those coupled, only 57.4% and 71.4% had both partners represented in claims data. This implies that we would expect to observe 21–28% of our samples to be in partnerships if we only matched couples where both members met our Medicare eligibility, age and co-residence criteria. This is consistent with our 25.1% detection rate in the Vital Status file.

Characteristics of Cohabiting Couples

Table 2 presents the demographics of the younger and older spouses in the identified Medicare couples in Column I and the two HRS samples in Columns II-III.

In Medicare Couples, 77.5% of younger spouses were female and the average younger spouse was 76.8 years old. 75.7% of the older spouses were males and the average older spouse was 80.2 years old. Although the sex composition differences between the older spouses of Medicare and HRS couples were statistically significant, they were not qualitatively meaningful. There were small differences in the share of non-Whites in HRS versus Medicare Couples, which partially reflects oversampling of minorities in the HRS (Ofstedal and Weir 2011), as well as use of self-reported versus administrative data, as not all eligible HRS spouses also consented to the Medicare linkage. A similar proportion of coupled beneficiaries and HRS respondents had an ADRD diagnosis.

Assortative Mating

Figure 1 compares and contrasts the assortative mating characteristics of the HRS Couples, Medicare Couples, Medicare Non-Couples, and Random Pairs. The characteristics of interest are percent different-sex, same-race, and aged within five years of each other. Medicare Couples most closely resembled the HRS couples because they were closer in age

and more likely to be different-sex and same-race compared to Medicare Non-Couples and Random Pairs. Additional assortative mating tables stratified by the race of the younger spouse can be found in Appendix B.

CONCLUSION

In this study we assessed the feasibility and validity of using exact address information to identify cohabiting couples in Medicare claims data. We defined two people to be a cohabiting couple if they shared an exact address as well as ZIP codes from four years (2011–2014). Among our identified cohabiting couples, we found that the majority of younger partners were females and that the majority of older partners were males, mirroring the sex composition of nationally representative HRS samples, most of whom represent married couples.

25.1% of Medicare beneficiaries age 70 and up were identified as being in a cohabiting couple using our algorithm. HRS data sampled to match the characteristics of our Medicare claims data were consistent with this population prevalence of marriage of appropriate duration—showing marriage rates of 37–39%, with 21–28% married for 5 or more years to another Medicare beneficiary 70+ living in a two-person household. Therefore, address data represent a promising way to identify cohabiting couples in administrative data.

This study extends the previous research on identifying couples in Medicare claims data by using exact address information (Iwashyna et al. 1998). Unlike previous work, this method is not limited to Medicare claims data and can be applied to any longitudinal administrative data that facilitates extraction of addresses and historical ZIP code information. Cohabitation as a proxy for marital status of Medicare beneficiaries can help health services researchers and demographers further examine the relationship between spousal presence and healthcare outcomes and utilization over time. Similarly, it can facilitate the understanding of couples' dynamics for researchers across disciplines.

Our study has several limitations. We excluded beneficiaries who were less than 70 years of age in 2014 in order to have at least four years' worth of ZIP codes. In our HRS samples we found that 6–7% of couples had a partner who is less than 65 years of age, and 76–77% of couples were within five years of each other. Our method assigns beneficiaries with younger partners to a one-beneficiary household. We therefore trade additional confirmation of relationship history for the ability to correctly classify younger couples which leads to undercounting. Next, the use of exact address information is also a limitation in and of itself. Self-reported addresses can be mis-entered which would affect the exact address matching and under-identify the number of couples in the dataset. Additionally, using exact addresses does not reveal any information on separated couples. One-quarter of our Medicare beneficiary sample was diagnosed with ADRD. The sample likely contains some institutionalized individuals (3% in HRS) with partners living in the community. If they have different registered addresses, we would again under-identify the number of couples by falsely attributing more beneficiaries to one-person households.

Our couples algorithm can more readily be adopted for research, risk-adjustment, and healthcare delivery applications through the production of an annual Vital Statistics file or similar address database for Medicare and Medicaid beneficiaries. Alternatively, another government agency such as the Internal Revenue Service could produce such a data resource.

Acknowledgments

Appendix

A. Address Cleaning

We geocoded the addresses in the Vital Status file using ArcGIS in order to obtain cleaned versions of the imputed addresses as well as each beneficiary's census block. We concatenated apartment numbers and PO box numbers to each beneficiary's cleaned address if available. The spelling sensitivity default in ArcGIS controls the amount of variation the geocoder will allow when identifying addresses in the reference data (Wilson et al. 2008). In other words, it standardizes directional terms such as "St." and "Street" and names such as "Universe' and "University," which helped us avoid undercounting addresses that contained spelling errors or abbreviated words (Wilson et al. 2008, ESRI).

B. Assortative Mating

Among the identified couples 97% are different sex, 96% are of the same race, and 82% are within five years of each other. Among those who did not meet the couple definition (Not-Identified), only 78% were different sex, 53% were of the same race, and 51% were within five years of each other. Randomly assigning beneficiaries to partners in their census block suggests that our couple identification strategy is stronger than chance alone given that 83% of these randomly identified couples are of the same race, only 50% are opposite sex and 49% are within 5 years of each other, deviating severely from what we hypothesize a couple to resemble.

Table 3:

Assortative Mating Characteristics-All Couples

Couple	HRS Couples	Medicare Couples	Medicare Non-Couples	Medicare Random Pairs
Diff Sex (%)	96.81	96.51	78.86	50.37
Same Race (%)	93.97	96.33	52.62	84.99
<= 5 yrs (%)	74.77	82.16	63.90	48.86
n	282	1,620,110	255,643	4,878,825

Table 4:

Assortative Mating Characteristics-White Younger Spouse

Couple	HRS Couples	Medicare Couples	Medicare Non-Couples	Medicare Random Pairs
Diff Sex (%)	96.34	96.87	79.01	50.69
Same Race (%)	96.75	98.27	54.91	92.46
<= 5 yrs (%)	76.42	832.83	64.31	49.18
n	246	1,486,772	226,685	4,199,892

Table 5:

Assortative Mating Characteristics-Black Younger Spouse

Couple	HRS Couples	Medicare Couples	Medicare Non-Couples	Medicare Random Pairs
Diff Sex (%)	100.00	90.90	79.22	47.47
Same Race (%)	92.59	93.87	37.69	51.56
<= 5 yrs (%)	59.26	74.66	61.53	46.16
n	27	63,305	16,577	392,219

Table 6:

Assortative Mating Characteristics-Other/Unknown Younger Spouse

Couple	HRS Couples	Medicare Couples	Medicare Non-Couples	Medicare Random Pairs
Diff Sex (%)	100.00	94.02	91.32	49.58
Same Race (%)	22.22	56.79	51.23	21.28
<= 5 yrs (%)	66.67	74.76	71.74	47.97
n	9	69,033	167,885	286,714

REFERENCES

Allison PD & Christakis NA (2006). Fixed-effects methods for the analysis of nonrepeated events. Sociol Methodol, 36, 155–172.

Barford A, Dorling D, Smith GD & Shaw M (2006). Life expectancy: women now on top everywhere. BMJ, 332, 808. [PubMed: 16601021]

Calvillo–King L, Arnold D, Eubank KJ, Lo M, Yunyongying P, Stieglitz H & Halm EA (2013). Impact of social factors on risk of readmission or mortality in pneumonia and heart failure: systematic review. J Gen Intern Med, 28, 269–282. [PubMed: 23054925]

Case A & Paxson C (2005). Sex differences in morbidity and mortality. Demography, 42, 189–214. [PubMed: 15986983]

Christakis NA & Allison PD (2006). Mortality after the hospitalization of a spouse. N Engl J Med, 354, 719–730. [PubMed: 16481639]

Christakis NA & Allison PD (2009) Inter-spousal mortality effects: Caregiver burden across the spectrum of disabling disease. Health at older ages: The causes and consequences of declining disability among the elderly. (pp. 455–477): University of Chicago Press.

Christakis NA & Iwashyna TJ (1998). Spousal illness burden is associated with delayed use of hospice care in terminally ill patients. J Palliat Med, 1, 3–10. [PubMed: 15859867]

Christakis NA, Iwashyna TJ & Zhang JX (2002). Care after the onset of serious illness: a novel claims-based dataset exploiting substantial cross-set linkages to study end-of-life care. J Palliat Med, 5, 515–529. [PubMed: 12243676]

Elwert F & Christakis NA (2006). Widowhood and race. Am Sociol Rev, 71, 16–41.

Elwert F & Christakis NA (2008a). The effect of widowhood on mortality by the causes of death of both spouses. Am J Pub Health, 98, 2092–2098. [PubMed: 18511733]

Elwert F & Christakis NA (2008b). Wives and ex-wives: a new test for homogamy bias in the widowhood effect. Demogr, 45, 851–873.

ESRI (2019) Geocoding options properties. In I. Environmental Systems Research Institute (Ed.), ArcGIS for Desktop.

Gilden DM, Kubisiak JM, Kahle-Wrobleski K, Ball DE & Bowman L (2017). A claims-based examination of health care costs among spouses of patients with Alzheimer's Disease. J Gerontol A Biol Sci Med Sci, 72, 811–817. [PubMed: 28329147]

Iwashyna TJ, Brennan G, Zhang JX & Christakis NA (2002). Finding Married Couples in Medicare Claims Data. Health Services and Outcomes Research Methodology, 3, 75–86.

Iwashyna TJ & Christakis NA (2003). Marriage, widowhood, and health-care use. Soc Sci Med, 57, 2137–2147. [PubMed: 14512244]

Iwashyna TJ, Zhang JX, Lauderdale DS & Christakis NA (1998). A methodology for identifying married couples in Medicare data: mortality, morbidity, and health care use among the married elderly. Demogr, 35, 413–419.

Jin L & Chrisatakis NA (2009). Investigating the mechanism of marital mortality reduction: the transition to widowhood and quality of health care. Demogr, 46, 605–625.

Juster FT & Suzman R (1995). An overview of the Health and Retirement Study. J. Hum. Resour, 30, S7–S56.

Luy M & Minagawa Y (2014). Gender gaps - Life expectancy and proportion of life in poor health. Health Reports, 25, 12–19.

Monden C (2007). Partners in health? Exploring resemblance in health between partners in married and cohabiting couples. Sociology of Health & Illness, 29, 391–411. [PubMed: 17470218]

National Academies of Sciences, Engineering, Medicine (2017). Accounting for social risk factors in Medicare payment: National Academies Press.

Ofstedal MB & Weir DR (2011). Recruitment and retention of minority participants in the Health and Retirement Study. J. Gerontol, 51 Suppl 1, S8–S20.

Smith KP & Christakis NA (2009). Association between widowhood and risk of diagnosis with a sexually transmitted infection in older adults. Am J Pub Health, 99, 2055–2062. [PubMed: 19762656]

Smock PJ & Schwartz CR (2020). The Demography of Families: A Review of Patterns and Change. Journal of Marriage and Family, 82, 9–34. [PubMed: 32612304]

Sonnega A, Faul JD, Ofstedal MB, Langa KM, Phillips JWR & Weir DR (2014). Cohort profile: the Health and Retirement Study (HRS). Int. J. Epidemiol, 43, 576–585. [PubMed: 24671021]

Subramanian SV, Elwert F & Christakis N (2008). Widowhood and mortality among the elderly: The modifying role of neighborhood concentration of widowed individuals. Soc Sci Med, 66, 873–884. [PubMed: 18178300]

Waite LJ & Gallagher M (2000) The case for marriage: Why married people are happier, healthier, and better off financially. (pp. Chapters 1 & 5), New York: Doubleday.

Wilson J, Swift J & Goldberg DW (2008). Geocoding best practices: Review of eight commonly used geocoding systems, Los Angeles, CA: University of Southern California GIS Research Laboratory

Wood RG, Goesling B & Avellar S (2007). The Effects of Marriage on Health: A Synthesis of Recent Research Evidence, Princeton, NJ: Mathematica Policy Research.

Wu J-R, Moser DK, Lennie TA & Burkhart PV (2008). Medication adherence in patients who have heart failure: a review of the literature. Nurs Clin North Am, 43, 133–153. [PubMed: 18249229]

Young NL, Cheah D, Waddell JP & Wright JG (1998). Patient characteristics that affect the outcome of total hip arthroplasty: a review. Can J Surg, 41, 188. [PubMed: 9627543]

Zivin K & Christakis NA (2007). The emotional toll of spousal morbidity and mortality. Am J Geriatr Psychiatry, 15, 772–779. [PubMed: 17804830]
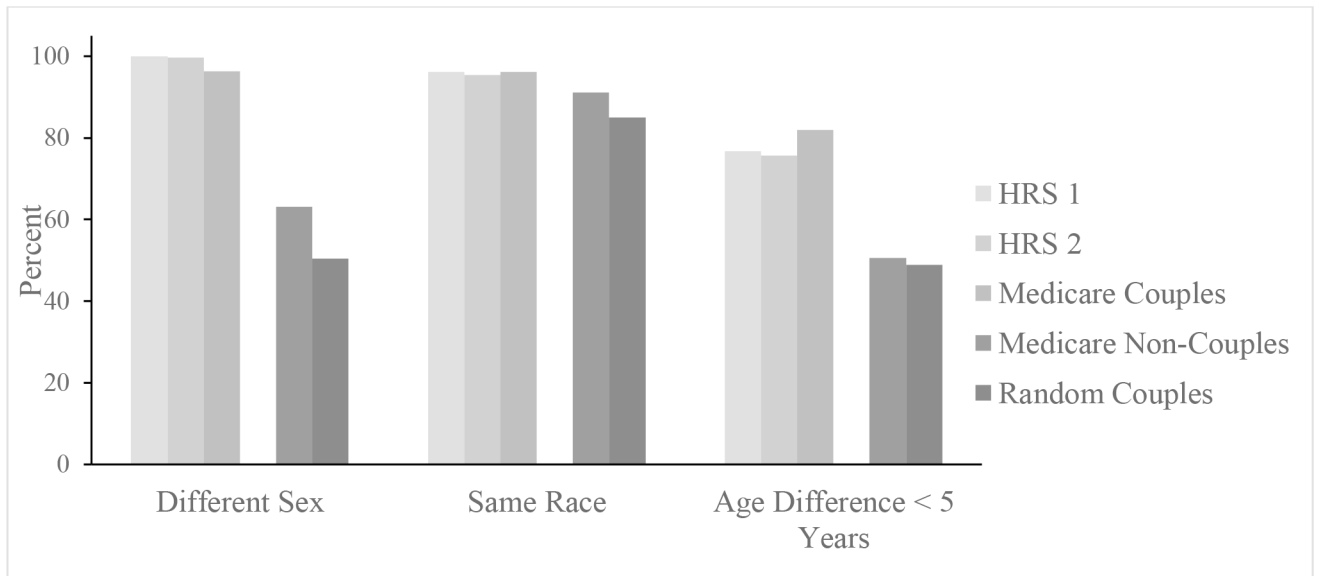
FIGURE 1.
CHARACTERISTICS OF IDENTIFIED COUPLES.

Note: Medicare Couples, Medicare Non-Couples, and Random Pairs are described here using data found in the Medicare claims file. HRS couples have known marital status, linked Medicare claims, and are sampled in two approximations to the criteria used to select the Medicare address sample. Non-couples share the same address but not ZIP code history, random couples are pairs residing in the same census block.

Table 1.

Characteristics of Medicare Beneficiaries by Number of People at the Same Address in the Medicare Vital Status 2018 File

Beneficiaries per Address	1	2	3+
Sex, %			
Male	39.42	48.37	38.55
Female	60.58	51.63	61.45
Race, %			
White	83.96	90.83	86.35
Black	9.63	4.71	8.21
Other/Missing	6.41	4.46	5.44
Age, years	81.00	78.74	82.11
Months in HMO	2.99	3.38	2.76
N	8,556,693	3,997,608	2,013,298

Note: We counted the number of times an address appeared in the Medicare Vital Status 2018 File in order to determine the number of beneficiaries at each address.

Table 2.

Characteristics of Medicare Beneficiaries Believed to be Cohabiting Couples Compared to Known HRS Couples

Couples Sample	Medicare N = 1,827,204		HRS 1[a,b] N = 2,134		HRS 2[a,b] N = 2,002	
Partner	**Younger**	**Older**	**Younger**	**Older**	**Younger**	**Older**
Share Coupled, %	25.09		39.51 **		37.07 **	
Share Coupled, Both Members 70+ Medicare Beneficiaries in 2-person household, %	25.09		28.21		21.28	
Sex, %						
Male	22.48	75.65	20.02 *	79.26 **	18.60 **	80.80 **
Female	77.52	24.35	79.98 *	20.74 **	81.40 **	19.20 **
Race (%)						
White	91.30	91.01	91.77	91.38	90.78	90.36
Black	4.52	4.54	5.49 *	5.66 *	5.95 **	6.09 **
Other/Unknown	4.18	4.46	2.73 **	2.97 **	3.27 *	3.55 *
Mean age, years	76.78	80.22	76.93	80.38	77.44 **	81.21 **
ADRD, %	12.24		13.04		13.58 *	

Note: HRS samples were compared to identified couples (I) using a t-test for slope in a weighted regression.

[a] HRS columns (II-III) contain all respondents age 70+ in the HRS-Medicare claims linkage in 2014 with ADRD, a 3:1 sample of non-ADRD respondents, and all partners. Summary statistics are weighted.

[b] Non-ADRD respondents in HRS Couples 1 were chosen by sorting respondents by lowest county-level MA penetration.

[c] Non-ADRD respondents in HRS Couples 2 were chosen from individuals with a unique ZIP code-year of birth combination, again sorting by MA penetration.

*
$p < 0.05$.

**
$p < 0.01$.