


# Validation of Self-Reported Rheumatoid Arthritis Using Medicare Claims: A Nationally Representative Longitudinal Study of Older Adults

Michael J. Booth,<sup>1</sup>  Daniel Clauw,<sup>1</sup> Mary R. Janevic,<sup>1</sup> Lindsay C. Kobayashi,<sup>1</sup> and John D. Piette<sup>2</sup>

**Objective.** To determine the validity of self-reported physician diagnosis of rheumatoid arthritis (RA) using multiple gold-standard measures based on Medicare claims in a nationally representative sample of older adults and to verify whether additional questions about taking medication and having seen a physician in the past two years for arthritis can improve the positive predictive value (PPV) and other measures of the validity of self-reported RA.

**Methods.** A total of 3768 Medicare-eligible respondents with and without incident self-reported RA were identified from the 2004, 2008, and 2012 waves of the United States Health and Retirement Study. Self-reported RA was validated using the following three claims-based algorithms: 1) a single International Classification of Diseases, ninth edition, Clinical Modification claim for RA, 2) two or more claims no greater than 2 years apart, and 3) two or more claims with at least one diagnosis by a rheumatologist. Additional self-report questions of medication use and having seen a doctor for arthritis in the past two years were validated against the same criteria.

**Results.** A total of 345 respondents self-reported a physician diagnosis of RA. Across all three RA algorithms, the PPV of self-report ranged from 0.05 to 0.16., the sensitivity ranged from 0.23 to 0.55., and the  $\kappa$  statistic ranged from 0.07 to 0.15. Additional self-report data regarding arthritis care improved the PPV and other validity measures of self-report; however, the values remained low.

**Conclusion.** Most older adults who self-report RA do not have a Medicare claims history consistent with that diagnosis. Revisions to current self-reported RA questions may yield more valid identification of RA in national health surveys.

## INTRODUCTION

Self-reports of health conditions are commonly used in national health surveys to estimate disease prevalence or identify risk factors. However, self-reported health conditions show varying agreement with medical records or administrative data (1). In general, self-reports of well-defined chronic conditions, such as diabetes mellitus and hypertension, show modest agreement with medical records, whereas those of less clearly defined conditions, such as chronic low back pain, do not (1,2). On the other hand, when people report not having a chronic disease, such as emphysema, cancer, lupus, or manic-depressive disorder, that information is often accurate and useful for identifying a disease-free population (1,3).

To validate self-reported medical conditions, researchers measure how self-report agrees with a gold standard, using standard indicators of validity such as sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and Cohen's  $\kappa$  (Figure 1). For the autoimmune disease rheumatoid arthritis (RA), previous studies conducted in the United States and elsewhere have shown that most people self-reporting physician-diagnosed RA do not have the disease according to medical records (Table 1) (4–16). The PPV of self-reported RA in prior studies ranges from 14.7% to 41%, meaning that on the high end, only 41% of people saying they had RA had confirmation by the respective gold standard. For studies reporting the sensitivity of self-report, the range was 54% to 100% (4,5,8,10,13,17), and for studies reporting  $\kappa$  statistics, the values ranged from 0.06 to

Supported by the National Institute of Aging (award T32AG000221), the Marshall Weinberg Endowment Fund (grant G002832), and the Population Studies Center Small Grants funded by the Marshall Weinberg Endowment Fund (grant G002832) at the University of Michigan Institute for Social Research.

<sup>1</sup>Michael J. Booth, MS, Daniel Clauw, MD, Mary R. Janevic, PhD, Lindsay C. Kobayashi, PhD: University of Michigan, Ann Arbor, Michigan; <sup>2</sup>John D. Piette, PhD: University of Michigan and Veterans Affairs Center for Clinical Management Research, Ann Arbor, Michigan.

Michael Booth is a Doctoral Candidate at the University of Michigan School of Public Health and a National Institute of Aging Trainee. John Piette is a Veterans Affairs Research Career Scientist. No other disclosures relevant to this article were reported.

Address correspondence to Michael J. Booth, MS, 1415 Washington Heights, Ann Arbor, MI, 48109. Email: mjbooth@umich.edu.

Submitted for publication December 15, 2020; accepted in revised form January 8, 2021.

**SIGNIFICANCE & INNOVATIONS**

- We verify self-reported rheumatoid arthritis (RA), a common measure used in National Health Surveys, against three records-based Medicare RA algorithms.
- The study innovates on existing research by including additional self-reports about care received for arthritis to verify whether the positive predictive value of self-reported RA is improved.
- The study also innovates on existing research by using incident rather than prevalent self-reported RA in a nationally representative survey of older adults.

0.46 (4,8,10,13). Prevalence affects the PPV of self-report, which is most clearly seen when the sample prevalence approaches 0% or 100%. Therefore, the most generalizable PPV studies have a sample prevalence representative of the actual population prevalence.

The consistent findings of low PPV and  $\kappa$  values of self-reported RA pose problems for researchers aiming to study RA in studies that use self-reports. To remedy this problem, three of the previously cited studies (4,7,10) looked at whether additional survey questions related to medication use could improve the PPV and other validity measures for self-reports. All three studies found

that self-reported use of disease-modifying antirheumatic drugs (DMARDs) increased the PPV of self-reported RA significantly (47-50.5% higher values) (4,7,10).  $\kappa$  statistics likewise increased to 0.47 and 0.65 (4,10). For example, Nguyen et al found that self-reported DMARD use increased the PPV of self-reported RA from 41% to 91.5% and the  $\kappa$  statistic from 0.22 to 0.87 (4). All three studies concluded that self-reported RA combined with self-reported DMARD use was satisfactory for identifying cases of RA. Although additional self-report questions related to DMARD use may increase the accuracy of self-report, such a definition simultaneously excludes those with RA not taking DMARDs.

Many publicly available national health surveys, such as the Health and Retirement Study (HRS) and the National Health and Nutrition Examination Survey, include self-reported RA questions. National surveys often include rich information about RA covariates regarding the respondent's lifestyle, diet, behaviors, socio-economic status, mental health, and more and can be a valuable resource for studying RA. The HRS is unique in that in survey years 2004 to 2012, respondents reported whether they took medication for arthritis (arthritis generally, not RA specifically) in addition to having RA. Moreover, the HRS includes linkable Medicare claims for consenting respondents, thereby allowing an evaluation of the validity of the self-report questions.

In this study, we validated self-reported RA diagnoses against diagnoses from Medicare claims in a large, nationally

Gold Standard	Self-Reported RA		Total
	Positive	Negative	
Positive	A	B	A+B
Negative	C	D	C+D
Total	A+C	B+D	N

A: True Positives, or Methods Agree

B: False Negatives, or Methods do not Agree

C: False Positives, or Methods do not Agree

D: True Negatives, or Methods Agree

Sensitivity:  $A/(A+B)$

Specificity:  $D/(C+D)$

PPV:  $A/A+C$

NPV:  $D/B+D$

Cohen's Kappa:  $(A+D/N) - \text{expected agreement} / (1-\text{expected agreement})$

Figure 1. Contingency table calculations.

**Table 1.** Validation studies of self-reported RA

Author, Year	Data Source	Study Type	Population	Gold Standard	Results
O'Rourke et al, 2019	Collected at time of survey	Online self-report questionnaire (Massachusetts)	Adult patients of a primary care clinic aged 18-70 yr	Electronic medical record review	PPV = 32%; sensitivity = 54%
Videm et al, 2017	Nord-Trøndelag Health Study	Population-based cohort study	Inhabitants of Nord-Trøndelag aged 20 yr and older	Hospital case files using EULAR 2010 criteria reviewed by immunologist	PPV = 19%
Formica et al, 2010	Black Women's Health Study	Prospective cohort study (United States)	Black women aged 21-69 yr	Medical Record and physician checklist review by blinded Rheumatologist.	No medication PPV = 29%; NSAID PPV = 61%; DMARD PPV = 76%
Oksanen et al, 2010	Finnish Public Sector Study	Prospective survey (Finland)	Employees of participating organizations	Clinical diagnosis of treating physician derived from three national registries.	PPV = 33%; sensitivity = 83%; $k = 0.47$
Cooper et al, 2008	Canadian Network for Improved Outcomes in SLE	Clinic-based questionnaire (Canada)	Probands of SLE recruited from 11 rheumatology clinics	Medical record review	PPV = 25%
Walitt et al, 2008	Women's Health Initiative	Prospective cohort study (United States)	Postmenopausal women aged 50-79 yr	Medical records reviewed by three blinded rheumatologists.	PPV = 14.7%; sensitivity = 100%; $k = 0.06$ ; DMARD PPV = 62.2%; DMARD sensitivity = 54.8%; DMARD $k = 0.53$
Karlson et al, 2004	Longitudinal Nurses' Health Study	Prospective cohort study (United States)	Female nurses aged 30-55 yr	Medical record review by two rheumatologists using ACR criteria	PPV = 29%
Mikulic, et al, 2002	Iowa Women's Health Study	Prospective cohort study (United States)	Women aged 55-69 yr in the state of Iowa	Medical record review by a combination of two trained rheumatologists, rheumatology advanced practice nurses, or rheumatology physician assistants using ACR criteria	PPV = 13%
Ling et al, 2000	Women's Health and Aging Study	Prospective cohort study (United States)	Community-dwelling disabled women aged 65 yr and older	Algorithm based on ACR criteria	PPV = 34%; sensitivity = 77%; $k = 0.46$
Karlson et al, 1999	Women's Health Cohort Study	Prospective cohort study (United States)	Female health professionals aged 45 yr and older	Medical records reviewed by two board certified rheumatologists	PPV = 35.8%
Kvien et al, 1996	Collected at time of survey	Postal survey (Norway)	Randomly selected individuals of Oslo, Norway, aged 20-79 yr	Clinical examination and Oslo RA registry	PPV = 31%
Star et al, 1996	Study of Osteoporotic Fractures	Prospective cohort study (United States)	Caucasian women aged 65 yr and older from Maryland.	Physician submitted survey of ACR criteria (1987), medical records, and radiographic hand features of RA, reviewed by investigators	PPV = 21%
Rasooly et al, 1995	Collected at time of survey	Mailed survey (England)	Rheumatology outpatients from Manchester Royal Infirmary	Recorded ICD-9 clinical diagnostic codes from clinic registry	Sensitivity = 90%

ACR, American College of Rheumatology; DMARD, disease-modifying antirheumatic drug; EULAR, European League Against Rheumatism; NSAID, nonsteroidal anti-inflammatory drug; PPV, positive predictive value; RA, rheumatoid arthritis; SLE, systemic lupus erythematosus.

representative sample of United States (US) adults aged 65 years and older enrolled in the HRS. We further examined whether including responses to the additional question about medication use improved self-reported RA accuracy. The American College of Rheumatology Guidelines for the Management of RA recommends ongoing follow-up care to assess treatment response in those with active disease (18,19). Therefore, we also sought to determine whether an additional question asked in the HRS about having seen a physician in the past two years for arthritis could improve the correspondence of self-reported RA with Medicare claims. To fully understand the accuracy of self-reported RA relative to prior studies, we used three different claims-based RA indicators as the gold standard.

## METHODS

**Data source.** We identified the study sample from the 2004, 2008, and 2012 waves of the US HRS. The HRS is a nationally representative longitudinal panel study of US residents over the age of 50 years (20). Sponsored by the National Institute on Aging (U01AG009740) and the Social Security Administration, the HRS surveys approximately 20 000 US adults each year on a variety of health and economic-related topics. Using a multistage area probability sample design, the HRS is a nationally representative sample of US adults in the target age group (21). Subjects and their spouses are followed from entry into the study until death, with cohorts entering the survey every 6 years and then surveyed every 2 years (22).

**Sample eligibility and exclusion criteria.** We chose the 2004, 2008, and 2012 waves for the current study because they are the most recent survey waves that included items measuring self-reported RA and questions about receiving care for arthritis after the RA self-report that all HRS respondents answered. We used incident rather than prevalent self-reported RA because, in the HRS, respondents self-report a lifetime history of RA. Medicare claims are available in the HRS from 1992 to 2016; however, using prevalence would give respondents an unequal lookback period. For instance, someone who is 65 years old in 2004 would not have any Medicare claims to confirm self-report, whereas someone who is 85 years old would have claims for the entirety of the available data years. By using incident self-report, all respondents have an identical 4-year lookback period at each wave.

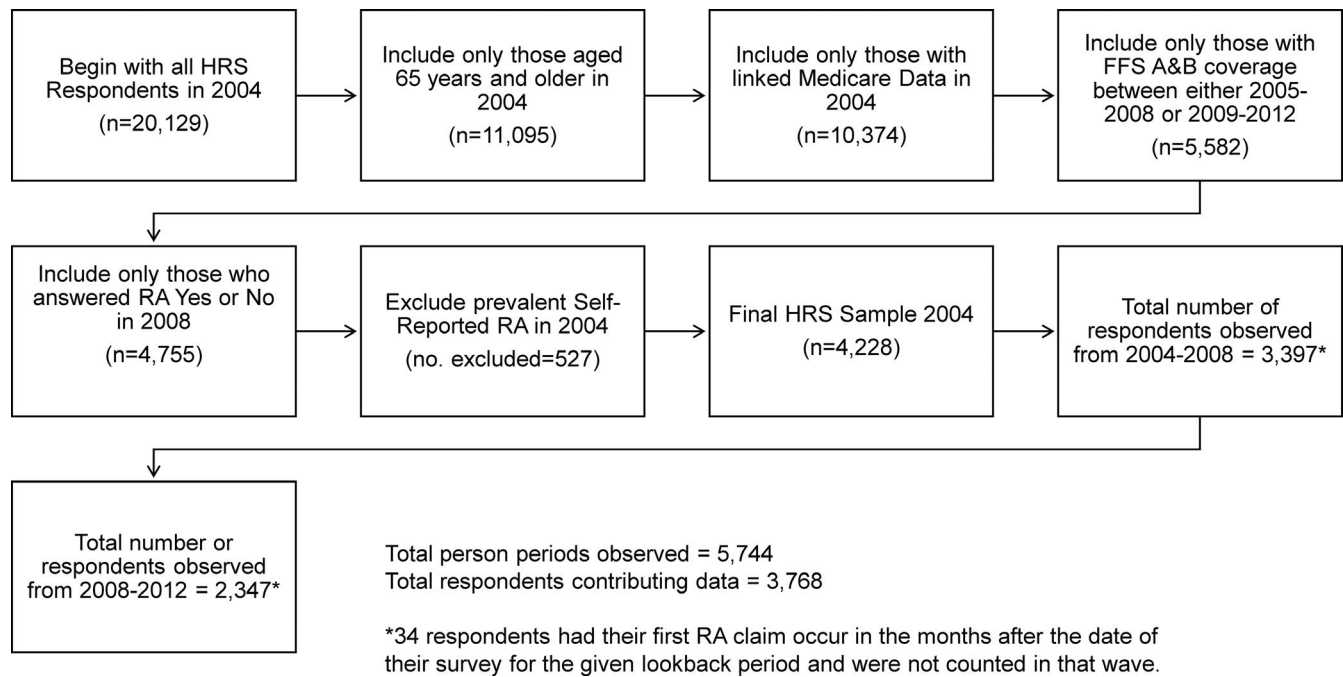
To identify incident self-reported RA, we excluded prevalent RA self-reports in 2004 and then tabulated how responses changed between waves. For example, if a respondent reported no RA in 2004 and RA in 2008 or 2012, we classified these as incident cases. Respondents to the 2006 and 2010 waves were also asked the same RA questions detailed above; however, responses at these waves were only for respondents who did not report at the previous wave or respondents who wished to alter their previous answers. Therefore, new responses in the

intervening 2006 and 2010 waves were excluded from our analysis to maintain the same 4-year lookback period, though those who wanted to change their last wave's responses were included and captured in the next wave.

Medicare claims document billing for various encounter types, including inpatient stays, outpatient primary care, specialty visits, and professional providers' claims. Claims include diagnoses listed in the encounter for billing purposes, recorded using International Classification of Diseases, ninth edition, Clinical Modification (ICD-9-CM) codes, Health Care Common Procedure Coding System, and Current Procedural Terminology codes. We employed 4-year lookback periods between the 2004 to 2008 survey wave and the 2008 to 2012 wave to verify incident self-reported RA against linked Medicare claims from Part A inpatient and skilled nursing facility files and Part B carrier files from primary care providers, outpatient care, home health services, durable medical equipment, and preventative services. Centers for Medicare & Medicaid Services (CMS) pays for care on a fee-for-service (FFS) basis (Part A and Part B) and a capitated basis via Part C, which is also called Medicare Advantage or Medicare + Choice. Part C claims are not yet available for linkage to the HRS. Not all HRS respondents consent to linking their Medicare records (20). We addressed these differences in respondents' enrollment patterns through our exclusionary criteria, detailed below.

Several exclusionary criteria limited the original sample of potentially eligible subjects (Figure 2). From the 20129 HRS respondents in 2004, we included those who 1) were 65 years or older in 2004 ( $n = 11\,095$ ), 2) had linked Medicare data ( $n = 10\,374$ ), 3) had a FFS Parts A and B Medicare coverage period defined as 11 months or more per year for all 4 years either from 2005 through 2008 or 2009 through 2012 ( $n = 5582$ ), and 4) answered "yes" or "no" in 2008 to the question about self-reported RA because nonresponses or "don't know" responses would not allow us to identify new cases of self-reported RA ( $n = 4755$ ) in 2008 or 2012. We then excluded all prevalent cases of self-reported RA in 2004 ( $n = 527$  excluded), leaving a final eligible sample of 4228. Of these 4228 respondents, we did not count 34 who had their first RA claim occur in the months after their interview date at their respective wave. These excluded respondents' late RA claims are due to overlap in the lookback period's final year and the month of the year that participants took their survey. Study wave nonresponse, participant death, or voluntary study withdrawal occurring in patterns between waves that did not allow a calculation of incident self-report further reduced the sample by 426 respondents, leaving 3768 contributing self-report data.

**Identifying RA cases in Medicare claims data.** We identified RA from Medicare claims associated with encounters occurring in the 4 years before either the 2008 or 2012 survey wave, using three different definitions based on commonly used public health surveillance algorithms and a literature review of prior epidemiologic studies. These three algorithms (described in more



**Figure 2.** Study sample flow chart.

detail below) are 1) the Center for Disease Control and Prevention (CDC) ICD-9-CM case definition of arthritis and other rheumatic conditions, 2) the Chronic Conditions Warehouse (CCW) algorithm for establishing whether a beneficiary is likely to have osteoarthritis (OA) or RA requiring medical care (restricted to RA-specific codes), and 3) a strict algorithm derived from a systematic review of validated methods for identifying patients with RA using administrative data that assessed the PPV of ICD code-based algorithms (22).

The National Arthritis Data Workgroup developed the CDC case definition of arthritis and related rheumatological conditions using ICD-9-CM codes (23,24). The definition identifies RA cases based on the presence of at least one arthritis-related ICD-9-CM code in administrative, mortality, or health care use databases. We examined the working group’s list of codes and restricted our definition to capture RA only, using a minimum of one code for 714 or 714.0 (“systemic rheumatoid arthritis”). We identified codes listed either as the principal diagnosis or in 1 of the 25 primary/secondary diagnostic fields from Medicare Part A or the 12 fields from Part B carrier files. We excluded claims from nonlicensed health care providers, such as ambulance services and durable medical equipment providers.

The CMS CCW algorithms aim to reduce the risk of false positives that may occur when trying to identify the presence of chronic conditions on the basis of billing claims (25). The CCW algorithm for identifying OA/RA cases specifies a 2-year reference period in which beneficiaries must have two or more diagnoses listed in inpatient, skilled nursing facility, home health agency, hospital outpatient, or routine outpatient visits from carrier claims. The 2-year reference period requires that beneficiaries have 11 to 12 months per year of non-health maintenance organization (HMO) Parts

A and B coverage and two claims within that period. During the 4-year lookback period between waves, we required a minimum of two RA claims at least 1 day and no greater than 730 days apart, restricted to ICD-9-CM codes 714 and 714.0. As we did in the CDC algorithm, we excluded nonlicensed health care providers’ claims and counted any RA claim occurring in the available fields.

Chung et al conducted a systematic review that assessed the PPV of ICD-9 code-based algorithms for identifying RA (22). Algorithms that had, in addition to ICD-9 codes, information related to DMARD use or the requirement that some or all RA codes come from a rheumatologist performed much better than ICD codes alone (22). For instance, Kim et al compared Medicare claims against an RA diagnosis made by a rheumatologist and found that the PPV of two or more RA claim codes was 55.7%, of three or more RA claims codes was 65.5%, and of at least two rheumatology claims codes was 66.7%. When including at least one DMARD prescription in addition to these algorithms, the PPV increased to 86.2% to 88.9% (26). The HRS has linkable Medicare pharmacy claims; however, data were not available for all study years in our analysis. Therefore, we opted to create a “strict” algorithm that was identical to the CCW criteria above but with at least one of the two minimum claims coming from a rheumatologist. We identified rheumatology clinic-based RA claims using CMS provider specialty code 66 from Part B carrier files listed in at least 1 of the 13 specialty billing fields.

**RA self-report variables.** We identified self-reported RA in two steps. First, self-report of any form of arthritis comes from HRS responses to the question “Have you ever had, or has a doctor ever told you that you have arthritis or rheumatism?” The

HRS definition of medical doctor includes “specialists such as dermatologists, psychiatrists, ophthalmologists, osteopaths, cardiologists, as well as family doctors, internists, and physician’s assistants” and also includes “diagnoses made by nurses and nurse practitioners.” Second, people who responded affirmatively further identified which type of arthritis they have as “osteoarthritis, rheumatoid arthritis (sometimes called autoimmune arthritis), gout or lupus, or arthritis related to a previous injury.” Our first definition of self-reported RA included respondents who answered “yes” to the first question and indicated “rheumatoid arthritis” in their response to the second question.

As a potentially more specific self-report, we also included information about self-reported doctor visits and arthritis-related medication use. Respondents answered whether, in the past two years, they had seen a doctor specifically for arthritis. Respondents also answered whether they were “currently taking medication or other treatments for arthritis or rheumatism.” These questions were not specific to RA but instead were asked of anyone responding yes to having a form of arthritis.

**Statistical methods.** We examined sociodemographic differences in incident RA, assessed four different ways, as follow: self-reported RA (yes versus no); the CDC ICD9-CM case definition of a single code indicating RA; the CCW algorithm; and our strict algorithm requiring, in addition to the CCW criteria, at least one diagnosis from a rheumatologist.

To calculate the sensitivity, specificity, PPV, NPV, and  $\kappa$  of self-reported RA compared with the three Medicare-based claims algorithms, we computed incident RA from both the 2008 and 2012 survey waves and created contingency tables representing the validity of self-report measures, including 1) self-reported diagnosis only, 2) self-reported diagnosis + self-reported arthritis medications, 3) self-reported diagnosis + self-reported doctor visits, and 4) self-reported diagnosis + doctor visits + medications for each algorithm.

The use of proxy respondents is common in large health surveys. Prior research conducted in an HRS sister study, the Survey on Assets and Health Dynamics Among the Oldest Old, found that the correspondence of self-report with Medicare use claims varies by whether a proxy represents the respondent (27). Therefore, we also examined whether the PPV of self-reported RA differed by proxy status. Specifically, we created a binary proxy variable for the 2008 and 2012 waves (yes or no), grouping both spouse and nonspouse proxies into a single category, and compared the PPV of incident self-reported RA by proxy status against all three RA algorithms using two-sample tests of proportions.

Finally, to verify whether our exclusionary criteria based on Medicare enrollment patterns introduced bias into our sample, we used bivariate statistics to examine sociodemographic and RA self-report differences between 1) Medicare linkage (yes versus no) and 2) full Parts A and B coverage of at least 11 months per year for all 4 years between survey waves (yes versus no).

## RESULTS

**Sample characteristics.** The study sample included 3768 HRS respondents 65 years of age and older with linked Medicare claims and complete sociodemographic data who contributed incident self-report data during either the 2008 or 2012 HRS survey wave. The mean age was 73.7 years (SD = 6.7) in 2004, 58.8% were women, 88.5% were White/Caucasian, 8.9% were Black/African American, 19.7% had less than high-school education, and 5.8% of respondents required a proxy.

A total of 345 respondents (9.2%) self-reported incident RA in either 2008 or 2012 (Table 2). Compared with those not reporting RA, those with incident self-reported RA were more likely to be Black or African American (15.6% versus 8.3%;  $P < 0.001$ ), more often had less than high-school education (29.6% versus 18.7%;  $P < 0.001$ ), and were more often from the lowest quartile of wealth (38.6% versus 23.8%;  $P < 0.001$ ).

We identified 245 respondents (6.5% of all respondents) who had a single RA claim. Compared with those without claims data indicating RA, those with a single RA claim were more often female (70.6% versus 58.0%;  $P < 0.001$ ) and more often had lower educational attainment and lower total household wealth. We identified 103 respondents meeting the CCW RA algorithm, and these respondents were proportionally more often female compared with participants without RA by the CCW criterion (72.8% versus 58.4%;  $P = 0.003$ ). Finally, we identified 29 respondents meeting our strict RA algorithm, who were, on average, 2.6 years younger than those without this claim pattern ( $P = 0.038$ ).

**Sensitivity, specificity, PPV, NPV, and  $\kappa$  of RA self-reports.** Of the 345 respondents with incident self-reported RA, 57 had a single RA claim (PPV = 16%), 41 had two or more claims according to the CCW criteria (PPV = 12%), and 16 met the strict RA criterion (PPV = 5%).  $\kappa$  statistics were 0.13 for a single diagnosis, 0.15 for the CCW criteria, and 0.07 for the strict definition.

In addition to self-reported RA, 161 respondents reported also taking medication for arthritis. Of these respondents, 43 were confirmed by a single RA claim (PPV = 27%), 31 were confirmed by the CCW criteria (PPV = 19%), and 13 were confirmed by the strict definition (PPV = 8%).  $\kappa$  statistics for those additionally reporting medication use were 0.17 for a single RA claim, 0.20 for the CCW criteria, and 0.12 for the strict definition. A total of 152 respondents reported, in addition to having RA, seeing a doctor in the past 2 years for arthritis. Of these 152 respondents, 44 were confirmed by a single RA claim (PPV = 29%), 32 were confirmed by the CCW criteria (PPV = 21%), and 16 were confirmed by the strict definition (PPV = 11%). Finally, 111 respondents reported having RA and both taking medication and seeing a doctor in the past 2 years for arthritis. Of respondents reporting both types of care received, 37 were confirmed by a single RA claim (PPV = 33%), 27 were confirmed by the CCW criteria (PPV = 24%), and 13 were confirmed by the strict definition (PPV = 12%).

**Table 2.** Population characteristics of incident self-reported RA and Medicare claims history from the HRS, 2004-2012

Population Characteristic (n = 3768)	Incident Self-Reported RA		One or More Medicare RA Claims		Two or More Medicare RA Claims		Two or More Medicare RA Claims, One by a Rheumatologist		P Value			
	Negative (n = 3423)	Positive (n = 345)	P Value	Negative (n = 3523)	Positive (n = 245)	P Value	Negative (n = 3655)	Positive (n = 103)		P Value		
Age in 2004, mean (SD), yr	73.7 (6.7)	73.5 (6.5)	0.59	73.7 (6.7)	73.5 (6.4)	0.69	73.7 (6.7)	72.8 (5.7)	0.09	73.7 (6.7)	71.1 (5.0)	0.038*
Sex, n (%)												
Male	1412 (41.3)	140 (40.6)	0.81	1480 (42.0)	72 (29.4)	<0.001*	1524 (41.6)	28 (27.2)	0.003*	1544 (41.3)	8 (27.6)	0.14
Female	2011 (58.7)	205 (59.4)		2043 (58.0)	173 (70.6)		2141 (58.4)	75 (72.8)		2195 (58.7)	21 (72.4)	
Race and ethnicity												
Non-Hispanic White	3056 (89.3)	280 (81.2)	<0.001*	3124 (88.7)	212 (86.5)	0.34	3,243 (88.5)	93 (90.3)	0.59	3,309 (88.5)	Omitted	0.56
Black/African American	283 (8.3)	54 (15.6)		309 (8.8)	28 (11.4)		328 (8.9)	Omitted		336 (9.0)	Omitted	
Other	84 (2.4)	11 (3.2)		90 (2.6)	5 (2.0)		94 (2.6)	Omitted		94 (2.5)	Omitted	
Education												
Less than high school	641 (18.7%)	102 (29.6%)	<0.001*	676 (19.2)	67 (27.4)	0.005*	721 (19.7)	22 (21.4)	0.64	737 (19.7)	6 (20.7)	0.71
High school or equivalent	1899 (55.5)	192 (55.7)		1957 (55.6)	134 (54.7)		2030 (55.4)	61 (59.2)		2073 (55.4)	18 (62.1)	
2-yr college	109 (3.2)	4 (1.2)		107 (3.0)	6 (2.5)		111 (3.0)	Omitted		113 (3.0)	Omitted	
4-yr college or more	774 (22.6)	47 (13.6)		783 (22.2)	38 (15.5)		803 (21.9)	Omitted		816 (21.8)	Omitted	
Household wealth, n (%)												
<25th percentile	815 (23.8)	133 (38.6)	<0.001*	870 (24.7)	78 (31.8)	0.04*	913 (24.9)	35 (34.0)	0.07	940 (25.1)	8 (27.6)	0.14
25th-50th percentile	849 (24.8)	87 (25.2)		872 (24.8)	64 (26.1%)		910 (24.8%)	26 (25.2%)		927 (24.8%)	Omitted	
50th-75th percentile	873 (25.5)	70 (20.3)		893 (25.4)	50 (20.4)		927 (25.3)	16 (15.5)		941 (25.2)	Omitted	
75th-99th percentile	886 (25.9)	55 (15.9)		888 (25.2)	53 (21.6)		915 (25.0)	26 (25.2)		931 (24.9)	10 (34.5)	

HRS, Health and Retirement Study; RA, rheumatoid arthritis.

\* Denotes statistically significant P value at 0.05α.

Omitted rows are not reported because of low cell counts and HRS policy that protects the confidentiality of respondents.

Continuous measures were tested with a t test of equal or unequal variance as appropriate. Categorical measures were tested using Pearson's χ<sup>2</sup>.

Across all three RA algorithms and self-report combinations, NPVs remained high, between 94% and 99% and increasing with the RA algorithm's strictness. Specificity performed similarly, with high values ranging from 92% to 98% across all combinations. Sensitivity was, however, low, ranging from 15% to 55%. The highest values for sensitivity occurred in the strict RA algorithm compared with self-reported RA and in the strict algorithm compared with self-reported RA plus having seen a doctor (55% for both).

The PPV of self-reported RA did not differ by proxy status for a single RA claim or the CCW algorithm at either the 2008 or 2012 waves ( $P$  values ranged from 0.73 to 0.96). A proxy represented no respondents who were identified by our strict RA algorithm. Therefore, we could not determine whether proxy self-report differed for that method.

**Medicare Linkage and Enrollment Patterns.** A total of 93.5% of eligible HRS respondents in 2004 had Medicare linkage ( $n = 10374$ ) (Table 3). Compared with those who did not consent to Medicare linkage, those with linkage were more often older (mean 75 years versus 74 years;  $P = 0.002$ ), were proportionally more often female (94.2% female versus 92.6% male;  $P = 0.001$ ), were non-Hispanic White more often than Black or African American (93.9% non-Hispanic White versus 91.2% Black or African American;  $P < 0.001$ ), and were proportionally more

often from the lower quartile of wealth. Important for our analysis, we found no differences in self-reported RA patterns between groups.

Of the 10374 respondents in 2004 with linked Medicare data, 5582 (54%) had four full years of non-HMO Parts A and B coverage during either lookback period. Compared with those without full coverage, those with full coverage were somewhat younger (73.8 years versus 76.3;  $P < 0.001$ ), proportionally more often female (55.1% versus 44.9%;  $P = 0.002$ ), and proportionally more often non-Hispanic White. Between groups, we found a statistically significant difference in self-reported RA ( $P = 0.005$ ); respondents with full coverage more often self-reported having RA (54.7%) compared with those without coverage (45.3%), and they more often stated they did not know if they had RA (52.2%) compared with those without coverage (47.8%).

## DISCUSSION

Our analysis expands on findings from prior studies by testing the validity of self-reported RA in a US-based nationally representative sample. We found that self-reported RA has low validity for identifying survey respondents with RA on the basis of Medicare claims data, as indicated by three different claims-based RA diagnostic algorithms. Overall, the highest PPV for self-reported RA was 16% and came from a single claim at any point during the

**Table 3.** Sociodemographic Characteristics of HRS Respondents Age 65+ with and without Medicare linkage and with and without full FFS Parts A and B coverage

Population Characteristic	Medicare Linkage			Full Parts A and B FFS Coverage for Either Reference Period		
	No ( $n = 721$ )	Yes ( $n = 10\,374$ )	$P$ Value	No ( $n = 4791$ )	Yes ( $n = 5583$ )	$P$ Value
Self-reported RA in 2004, $n$ (%)						
No	554 (6.8)	7634 (93.2)	0.16	3457 (45.3)	4177 (54.7)	0.005*
Yes	77 (5.9)	1236 (94.1)		616 (49.8)	620 (50.2)	
Do not know	90 (5.7)	1503 (94.3)		718 (47.8)	785 (52.2)	
Age in 2004, mean (SD), yr	74 (8.3)	75 (7.5)	0.002*	76.3 (8.2)	73.8 (6.7)	<0.001*
Sex ( $n$ , %)						
Male	349 (7.4)	4372 (92.6)	0.001*	2097 (48.0)	2275 (52.0)	0.002*
Female	372 (5.8)	6002 (94.2)		2694 (44.9)	3308 (55.1)	
Race and ethnicity, $n$ (%)						
Non-Hispanic White	563 (6.1)	8709 (93.9)	<0.001*	3900 (44.8)	4809 (55.2)	<0.001*
Black/African American	129 (8.8)	1342 (91.2)		726 (54.1)	616 (45.9)	
Other	29 (8.3)	322 (91.7)		164 (50.9)	158 (49.1)	
Education, $n$ (%)						
Less than high school	228 (7.2)	2960 (92.8)	0.31	1637 (55.3)	1323 (44.7)	<0.001*
High school or equivalent	354 (6.2)	5,405 (93.8)		2356 (43.6)	3049 (56.4%)	
2-yr college	22 (7.0)	292 (93.0)		140 (48.0)	152 (52.0)	
4-yr college or more	117 (6.4)	1717 (93.6)		658 (38.3)	1059 (61.7)	
Total household wealth, $n$ (%)						
<25th percentile	210 (7.6)	2567 (92.4)	0.03*	1481 (57.1)	1,113 (42.9)	<0.001*
25-50th percentile	184 (6.6)	2595 (93.4)		1261 (48.6)	1,334 (51.4)	
50-75th percentile	158 (5.7)	2608 (94.3)		1109 (42.8)	1,484 (57.2)	
75th-99th percentile	169 (6.1)	2604 (93.9)		940 (36.3)	1,652 (63.7)	

FFS, fee for service; HRS, Health and Retirement Study; RA, rheumatoid arthritis.

\* Denotes statistically significant  $P$  value at 0.05  $\alpha$ .



four-year lookback periods, meaning that this method confirmed only 16% of respondents who self-reported a new case of RA. Our results agree with prior studies showing low PPVs of RA self-report. The highest PPV we obtained (16%) for a single RA claim is within the range reported previously. However, using our strict definition of RA, the PPV (5%) is the lowest of all studies to date. A limit of the strict algorithm is that some people with RA may receive the entirety of their care from general practitioners or other physicians. The PPV of the CCW algorithm, 12%, is also the lowest of all those cited.

Our lower PPVs may reflect differences across studies in the gold standard used to validate the self-report. In most prior studies, the gold standard for validation was medical review, whereas we used Medicare claims, which are the only available means of validating self-report in the HRS. The use of Medicare claims is inferior to the use of medical records, as research shows that Medicare claims can misclassify respondents (28). Accordingly, some people self-reporting RA in the HRS who did not have a Medicare record of RA may actually have the disease, which would increase the PPV of self-report. Furthermore, our algorithms showed decreases in the PPVs of self-report with the increasing strictness of the algorithm. This could mean that the stricter algorithms exclude people with RA who do not receive frequent care for it, explaining the lower PPV. Prior studies were also not restricted to older adults, which could explain our lower PPVs if older adults are more likely to confuse general “rheumatism” with RA.

The highest sensitivity we obtained, 55% for self-reported RA and the strict algorithm, is also within the range reported by others. To better understand the low sensitivity, we examined the claim patterns of false negatives (those meeting the strict RA definition but not self-reporting RA) and found they had anywhere from 1 to 38 RA claims made by a rheumatologist, meaning that some people who say they do not have RA are receiving significant ongoing care for it.

Additional self-report information about doctor visits and medication use increased the PPV and  $\kappa$  of self-report; however, these values remained low. Our study findings differ from those of other studies reporting that self-reported DMARD use significantly improved the PPVs and  $\kappa$  values of self-report to acceptable values, likely because the HRS question is about arthritis medication use generally and not specifically about DMARD use. Furthermore, the tradeoffs of including additional questions to increase the PPV of self-report require consideration. In many cases, the sensitivity of self-report may suffer by misclassifying respondents who self-report RA but do not use medication or regularly see a doctor for care as false positives when they may be true positives. The tradeoff between PPV and sensitivity is visible in Table 4, in which, as the PPV increases across additional self-report questions, the sensitivity simultaneously decreases. Additional self-report questions also restrict the generalizability of who has RA by limiting true positives to those who regularly see a doctor for care, take medication, or both, and not anyone with RA. Researchers

**Table 4.** Self-reported RA validation with Medicare claims, 2004-2012

	Self-Reported RA (n = 345)	Self-Reported RA + RX (n = 161)	Self-Reported RA + Doctor (n = 152)	Self-Reported RA + RX + Doctor (n = 111)
CDC method: one or more RA claims				
Sensitivity	0.23	0.18	0.18	0.15
Specificity	0.92	0.97	0.97	0.98
PPV	0.16	0.27	0.29	0.33
NPV	0.95	0.94	0.94	0.94
$\kappa$	0.13	0.17	0.18	0.17
CCW method: two or more RA claims				
Sensitivity	0.40	0.43	0.31	0.26
Specificity	0.92	0.96	0.97	0.98
PPV	0.12	0.19	0.21	0.24
NPV	0.98	0.98	0.98	0.98
$\kappa$	0.15	0.20	0.23	0.23
Strict method: two or more RA claims, one from a rheumatologist				
Sensitivity	0.55	0.45	0.55	0.45
Specificity	0.91	0.96	0.96	0.97
PPV	0.05	0.08	0.11	0.12
NPV	0.99	0.99	0.99	0.99
$\kappa$	0.07	0.12	0.23	0.18

CCW, Chronic Conditions Warehouse; CDC, Centers for Disease Control and Prevention; NPV, negative predictive value; PPV, positive predictive value; RA, rheumatoid arthritis.

RX = Self-report that respondent is taking medication for arthritis. Doctor = Self-report the respondent has seen a doctor in the past two years for arthritis. Sensitivity is the proportion of people with a Medicare history of RA reporting that they have RA. Specificity is the proportion of people without a Medicare history of RA of all respondents not reporting RA. The NPV is the proportion of respondents not reporting RA who have no Medicare history of RA. The CDC method is the CDC case definition of a single RA diagnosis. The CCW method is the CCW criteria of two RA diagnoses no greater than 2 years apart. The strict method is a minimum of two RA claims no more than 2 years apart, with a minimum of one claim from a rheumatologist.

should consider these tradeoffs and generalizability limitations when using additional self-report questions.

A potential weakness of our study is that we excluded a group of HRS participants over 65 years of age who did not have linked Medicare claims and another group who did not have full non-HMO Parts A and B coverage during the 4 years of either lookback period. We examined differences in self-reported RA in the excluded groups and found small differences between those with and without full Parts A and B FFS coverage (Table 3). We do not have complete Medicare claims for those excluded; therefore, we cannot verify whether differences between self-reported RA represent differences in the presence of RA between groups. Finally, losses due to study dropout, death, and nonparticipation could also affect our results; however, we do not have incident self-report data on these respondents and cannot say how their absence would alter our findings.

The use of incident rather than prevalent self-reported RA may also limit our findings. The HRS asks about a lifetime history of RA (prevalent cases), whereas we tabulated changes in this lifetime self-report between waves to classify incident self-report. Our definition of incident case is not the same thing as asking whether someone had a new diagnosis of RA in the last 4 years, which could yield different responses and results.

Because our study sample was significantly reduced from the original HRS sample and because of the study limitations mentioned, we cannot generalize our findings to the entire US older adult population. However, on the basis of our results and prior research, we recommend that studies using self-reported RA diagnoses do so with caution. To improve the validity of these self-reports, we recommend survey designers revise self-reported RA questions as follows: 1) Ask whether respondents have taken a DMARD and include a brief list of common medications and 2) ask whether the respondent has seen a rheumatologist for RA in the past 2 years. These small changes to the current questions in the HRS and similar surveys could yield a significant improvement in the detection of RA using self-report and facilitate more rapid and accessible research of RA in national health datasets without the cost, time requirements, and sample restrictions imposed by linked Medicare data.

## AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it for intellectual content. Michael Booth had full access to the data in the study and takes responsibility for the integrity of the data and the accuracy of the analysis. All authors approved the final version for publication. No potential conflicts of interest relevant to this article were reported.

**Study conception and design.** All authors contributed to study conception and design.

**Acquisition of data.** Michael Booth

**Analysis and interpretation of data.** Michael Booth

## REFERENCES

- Kehoe R, Wu S-Y, Leske CM, Chylack LT. Comparing self-reported and physician-reported medical history. *Am J Epidemiol* 1994;139:813–8.
- Skinner KM, Miller DR, Lincoln E, Lee A, Kazis LE. Concordance between respondent self-reports and medical records for chronic conditions. *J Ambul Care Manage* 2005;28:102–10.
- Smith B, Chu LK, Smith TC, Amoroso PJ, Boyko EJ, Hooper TI, Gackstetter GD, et al. Challenges of self-reported medical conditions and electronic medical records among members of a large military cohort. *BMC Med Res Methodol* 2008;8:37.
- Nguyen Y, Salliot C, Gusto G, Descamps E, Mariette X, Boutron-Ruault M-C, et al. Improving accuracy of self-reported diagnoses of rheumatoid arthritis in the French prospective E3N-EPIC cohort: a validation study. *BMJ Open* 2019;9:e033536.
- O'Rourke JA, Ravichandran C, Howe YJ, Mullett JE, Keary CJ, Golas SB, et al. Accuracy of self-reported history of autoimmune disease: a pilot study. *PLoS One* 2019;14:e0216526.
- Videm V, Thomas R, Brown MA, Hoff M. Self-reported diagnosis of rheumatoid arthritis or ankylosing spondylitis has low accuracy: Data from the nord-trøndelag health study. *J Rheumatol* 2017;44:1134–41.
- Formica MK, McAlindon TE, Lash TL, Demissie S, Rosenberg L. Validity of self-reported rheumatoid arthritis in a large cohort: results from the black women's health study. *Arthritis Care Res* 2010;62:235–41.
- Oksanen T, Kivimäki M, Pentti J, Virtanen M, Klaukka T, Vahtera J. Self-report as an indicator of incident disease. *Ann Epidemiol* 2010;20:547–54.
- Cooper GS, Wither J, McKenzie T, Claudio JO, Bernatsky S, Fortin PR, et al. The prevalence and accuracy of self-reported history of 11 autoimmune diseases. *J Rheumatol* 2008;35:2001–4.
- Walitt BT, Constantinescu F, Katz JD, Weinstein A, Wang H, Hernandez RK, et al. Validation of self-report of rheumatoid arthritis and systemic lupus erythematosus: the Women's Health Initiative. *J Rheumatol* 2008;35:811–8.
- Karlon EW, Mandl LA, Hankinson SE, Grodstein F. Do breast-feeding and other reproductive factors influence future risk of rheumatoid arthritis? Results from the nurses' health study. *Arthritis Rheum* 2004;50:3458–67.
- Mikuls TR, Cerhan JR, Criswell LA, Merlino L, Mudano AS, Burma M, et al. Coffee, tea, and caffeine consumption and risk of rheumatoid arthritis: results from the Iowa Women's Health Study. *Arthritis Rheum* 2002;46:83–91.
- Ling S, Fried L, Garrett E, Hirsch R, Guralnik J, Hochberg M, et al. The accuracy of self-report of physician diagnosed rheumatoid arthritis in moderately to severely disabled older women. *J Rheumatol* 2000;27:1390–4.
- Karlon EW, Lee IM, Cook NR, Manson JA, Buring JE, Hennekens CH. Comparison of self-reported diagnosis of connective tissue disease with medical records in female health professionals: the Women's Health Cohort Study. *Am J Epidemiol* 1999;150:652–60.
- Kvien TK, Glennås A, Knudsrød OG, Smedstad LM. The validity of self-reported diagnosis of rheumatoid arthritis: results from a population survey followed by clinical examinations. *J Rheumatol* 1996;23:1866–71.
- Star VL, Scott JC, Sherwin R, Lane N, Nevitt MC, Hochberg MC. Validity of self-reported rheumatoid arthritis in elderly women. *J Rheumatol* 1996;23:1862–5.
- Rasooly I, Papageorgiou AC, Badley EM. Comparison of clinical and self reported diagnosis for rheumatology outpatients. *Ann Rheum Dis* 1995;54:850–2.
- American College of Rheumatology Subcommittee on Rheumatoid Arthritis Guidelines. Guidelines for the management of rheumatoid arthritis: 2002 update. *Arthritis Rheum* 2002;46:328–46.
- Lacaille D, Anis AH, Guh DP, Esdaile JM. Gaps in care for rheumatoid arthritis: a population study. *Arthritis Rheum* 2005;53:241–8.
- Sonnega A, Faul JD, Ofstedal MB, Langa KM, Phillips JW, Weir DR. Cohort profile: the Health and Retirement Study (HRS). *Int J Epidemiol* 2014;43:576–85.

21. Heeringa SG, Connor JH. Technical description of the Health and Retirement Survey sample design. Ann Arbor, Michigan: University of Michigan; 1195. URL: <https://hrs.isr.umich.edu/sites/default/files/biblio/HRSSAMP.pdf>.
22. Chung CP, Rohan P, Krishnaswami S, McPheeters ML. A systematic review of validated methods for identifying patients with rheumatoid arthritis using administrative or claims data. *Vaccine* 2013;31 Suppl 10:K41–61.
23. Murphy LB, Cisternas MG, Greenlund KJ, Giles W, Hannan C, Helmick CG. Defining arthritis for public health surveillance: methods and estimates in four US population health surveys. *Arthritis Care Res* 2017;69:356–67.
24. Helmick CG, Lawrence RC, Pollard RA, Lloyd E, Heyse SP. Arthritis and other rheumatic conditions: who is affected now, who will be affected later? [Original Article]. *Arthritis Care Res* 1995;8:203–11.
25. Chronic Conditions Data Warehouse. Condition categories. Baltimore, MD: Centers for Medicare & Medicaid Services; 2020 [accessed August 7, 2020]. URL: <https://www2.ccwdata.org/web/guest/condition-categories>.
26. Kim SY, Servi A, Polinski JM, Mogun H, Weinblatt ME, Katz JN, et al. Validation of rheumatoid arthritis diagnoses in health care utilization data. *Arthritis Res Ther* 2011;13:R32.
27. Wehby GL, Jones MP, Ullrich F, Lou Y, Wolinsky FD. Does the relationship of the proxy to the target person affect the concordance between survey reports and medicare claims measures of health services use? [Research Brief]. *Health Serv Res* 2016;51:314–27.
28. Losina E, Barrett J, Baron JA, Katz JN. Accuracy of Medicare claims data for rheumatologic diagnoses in total hip replacement recipients. *J Clin Epidemiol* 2003;56:515–9.