



Completeness of cohort-linked U.S. Medicare data: An example from the Agricultural Health Study (1999–2016)

Christine G. Parks^{a,*}, Srishti Shrestha^b, Stuart Long^c, Thomas Flottemesch^d, Sarah Woodruff^c, Honglei Chen^e, Gabriella Andreotti^f, Jonathan N. Hofmann^f, Laura E. Beane Freeman^f, Dale P. Sandler^a

^a Epidemiology Branch, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC, USA

^b Formerly of the Epidemiology Branch, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC, USA

^c Westat, Rockville, MD, USA

^d DLH, Silver Spring, MD, USA

^e Department of Epidemiology and Biostatistics, College of Human Medicine, Michigan State University, East Lansing, MI, USA

^f Occupational and Environmental Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA

ARTICLE INFO

Keywords:
Cohort study
Medicare
Linked data
Farmers
Rural

ABSTRACT

Medicare Fee for Service (FFS) claims data, including inpatient (Part A) and outpatient (Part B) services, provide a valuable resource for research on older adults (≥ 65 year) in linked U.S. cohorts. Here we describe our experience linking the Agricultural Health Study cohort, including 47,501 licensed pesticide applicators and spouses from North Carolina (NC) and Iowa (IA) to Medicare claims data from 1999 to 2016. Given increased Part C (i.e., managed care/Medicare Advantage) enrollment during this period, and a resulting lack of available Part C claims data prior to 2015, we also explored potential for informative missingness. We compared those with partial or limited/no FFS to those with complete FFS coverage (i.e., ≥ 11 months per year parts AB, but not C, throughout Medicare enrollment) in relation to baseline farm size, general pesticide use, and mortality, in logistic regression models adjusted for age, sex, race, education, and smoking, and stratified by state. While 46,689 participants (98%) were linked to Medicare IDs, only 33,487 (70%) had complete FFS, 9353 (20%) had partial FFS (≥ 1 year FFS but not complete), and 3849 (8%) had limited/no FFS (Part A or Part C-only). Incomplete FFS was more common in NC, mostly due to Part C, and was associated with farm characteristics, pesticide use, and mortality. These findings indicate that, in addition to reduced sample size in analyses limited to complete FFS, missingness may not be random. The potential impact of incomplete FFS data and changes in coverage type need to be considered when planning linked analyses and interpreting results.

1. Introduction

Large prospective cohort studies are valuable for investigating risk factors for multiple health outcomes over time, however case ascertainment through active follow-up may be incomplete and certain outcomes of interests may not be included on questionnaires. Furthermore, surveys may be costly or burdensome for participants (especially in the elderly), and validation using medical records may be difficult or infeasible. In the U.S., where few disease registries exist other than for cancer, linkage to administrative claims data from the Centers for

Medicare & Medicaid Services (CMS) is an appealing strategy to enhance research in older populations. Individuals ages ≥ 65 years who qualify for benefits under the Social Security Age and Survivor Benefits program are eligible for Medicare. Coverage automatically includes inpatient hospital insurance (Part A); most individuals choose additional coverage for outpatient care (Part B), though some maintain only Part A due to other insurance options such as employer-based or Veterans benefits. Some may enroll in managed care (Part C), which covers inpatient and outpatient care. Parts A and B together, without Part C, are known as traditional Fee for Service (FFS) coverage.

Abbreviations: CMS, Centers for Medicare and Medicaid Services; FFS, Fee for service; SSN, Social Security Number; MMA, Medicare Modernization Act; ACA, Affordable Care Act.

* Corresponding author.

E-mail addresses: parks1@mail.nih.gov, Parks1@niehs.nih.gov (C.G. Parks).

<https://doi.org/10.1016/j.pmedr.2022.101766>

Received 16 July 2021; Received in revised form 5 March 2022; Accepted 13 March 2022

Available online 15 March 2022

2211-3355/Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Medicare claims data are widely used in health services research (Mues et al., 2017), and increasingly, linked epidemiologic cohorts (Hlatky et al., 2014; Jain et al., 2015; Kucharska-Newton et al., 2016; Lakshminarayan et al., 2014; Setiawan et al., 2015; Ton et al., 2014; Virnig et al., 2010; Xie et al., 2016; Zhou et al., 2015). Cohort studies typically use medical encounter claims, including diagnostic codes, to confirm self-reported diagnoses or complement existing data collected to identify cases (Jain et al., 2015; Lakshminarayan et al., 2014; Ton et al., 2014; Virnig et al., 2010), to ascertain outcomes not included in primary data collection (Johnson et al., 2018), and for comparisons with population data (Kalbaugh et al., 2017). Until recently (2015), claims data were limited to those processed under FFS insurance, and not for services covered through Part C (i.e., managed care, recently known as Medicare Advantage), which expanded coincident with programmatic changes introduced by the Medicare Modernization Act (MMA) in 2006 and the Affordable Care Act (ACA) following 2011 (Afendulis et al., 2012; Jacobson et al., 2015; McGuire et al., 2011). Missing health outcomes data in cohort studies can decrease statistical power, impact generalizability, and bias findings due to informative non-response, when known and unknown factors impacting case ascertainment are related to exposures, covariates, and outcomes. In analyses of Medicare-linked cohort data, missing or incomplete claims data may be a concern when missingness is related to differences in health and disease risk factors. In addition to temporal differences, patterns of FFS enrollment are known to vary by socioeconomic factors and health status (Meyers et al., 2019; Park et al., 2020; Rahman et al., 2015). These differences are particularly relevant to cohort studies with extended follow-up and in different geographic regions, which adds variability in access to managed care programs (Meyers et al., 2019), especially in rural areas (Chan et al., 2006; Geissler, 2020; Polinski et al., 2014).

The Agricultural Health Study (AHS) is a prospective cohort of over 89,000 licensed pesticide applicators (mostly private applicators/farmers) and their spouses, enrolled from 1993 to 1997. The cohort provides opportunities to investigate the chronic health effects of pesticide use; registry linkages have facilitated AHS research on cancer, end-stage renal disease, and mortality. Studies of other outcomes have relied on self-report, sometimes validated by medical records or evaluation by a healthcare provider. By 2015, more than half of participants were age 65 and older, providing an opportunity to supplement primary data collection with linkage to Medicare data. Here we describe the linkage of AHS to Medicare data from 1999 to 2016. During this time period, national rates of Part C enrollment more than doubled, to a high of 31% in 2016 (Jacobson et al., 2017). Thus, we also explored the impact of programmatic changes in Medicare on FFS coverage in the AHS and potential informative missingness in available claims data, considering baseline covariates, farming status and general pesticide use, and mortality.

2. Methods

2.1. Design and sample

The design of the AHS has been described previously (Alavanja et al., 1996). Private pesticide applicators (N = 52,394; mostly white male farmers) were recruited at licensing agencies in North Carolina (NC) and Iowa (IA), along with a sample (N = 4916) of commercial applicators in IA. Enrolled private applicators were given another questionnaire to complete at home (“take-home”, 44% response), and if married, were given a questionnaire for their spouse (75% response; N = 32,345). The first follow-up (1999–2005) was completed by 67%; subsequent follow-ups (2005–2010, 2013–2015) were limited to private applicators and spouses who completed the take-home or first follow-up (response 60–66%), and follow-up is ongoing (2019–2021). Study protocols, with implied consent for passive and active follow-up, were approved by the relevant institutional review boards, as was the linkage approved by CMS. AHS data used in this study (releaseP1REL201701.00,

P2REL201701.00) included demographics, smoking, self-reported diagnosis with various health conditions. Study questionnaires included health conditions and pesticide use (<https://aghealth.nih.gov/collaboration/questionnaires.html>). For this study, we identified five types of chronic health conditions based on enrollment questionnaires: (1) non-malignant lung disease (i.e., asthma, emphysema, chronic obstructive pulmonary disease or bronchitis), (2) cardiovascular (heart disease, myocardial infarction or stroke), (3) Hodgkin lymphoma, non-Hodgkin lymphoma, leukemia, melanoma, other cancer except non-melanoma skin cancer), (4) diabetes (not gestational), and (5) depression. Participants were grouped as having none, one, and two or more types of chronic conditions. In private applicators, data included number of acres planted/livestock raised, years and days per year mixed or applied pesticides during their lifetime, and at the first follow-up whether they lived or worked on a farm, and hours per day mixed/applied pesticides.

2.2. Performing the linkage

The linkage process and counts are described in Fig. 1. Participants were eligible if they were \geq age 65 between 1999 and 2016 (N = 47,501). AHS identifiers were provided to CMS for those with complete social security number (SSN; N = 45,063), and those lacking SSN (N = 2444) for matching by last name, gender, date of birth, and most recent zip code updated by annual study mailings. Six participants were excluded based on CMS date of death < age 65. CMS flagged exact matches for 43,792 (97%) with SSN and 2136 (87%) of those lacking SSN. We identified 761 inexact SSN matches as fuzzy-DOB matched. If zip code and gender matched, a fuzzy match allowed a 1-unit difference in date; most exactly matched by month/year (n = 488) or day/month (n = 114); some (n = 23) were confirmed by date of death. Those with multiple CMS-IDs (n = 42) were evaluated by the same criteria and retained for 26 individuals. Only 812 (1.7%) of the 47,501 eligible sample were unmatched.

2.3. Medicare data and derived variables

Using monthly enrollment data in the Master Beneficiary Summary File for each year, we defined three categories of FFS coverage (i.e., since age 65): (1) Complete FFS, ≥ 11 months per year of AB insurance, without >1 month of Part C or Part A-only, for the total time enrolled in Medicare. (2) Partial FFS, one or more years of continuous FFS but not meeting the definition of complete FFS due to enrollment in >1 month of Part C or Part A-only, and (3) limited/no FFS, Part A-only or Part C enrollment, and <1 year continuous FFS. Another variable identifies those with incomplete (partial or limited/no) FFS due to having either enrollment in Part C (any), Part A-only, and other reasons (e.g., gaps in coverage).

2.4. Analyses

Missing AHS race data were filled in using CMS race (N = 875). CMS date of death was used to determine the end of Medicare coverage. We examined participant age, type (private or commercial applicator, spouse), gender, race (white, non-white; non-white being mainly but not all African American in NC), state, education, smoking (never, past, current), and most recent AHS survey, by FFS coverage (complete, partial, or limited/no). We then compared those with partial or limited/no FFS coverage to complete FFS in multinomial logistic regression models, calculating odds ratios (OR) and 95% confidence intervals (CI) adjusted for continuous age and age-squared (SAS version 9.4, Cary, NC) to examine associations with gender, race/ethnicity, education, and smoking in mutually adjusted models limited to those with complete covariate data (for self-reported chronic diseases we also included a category for missing questionnaire data on health status). Models were stratified by state; differences were evaluated based on including

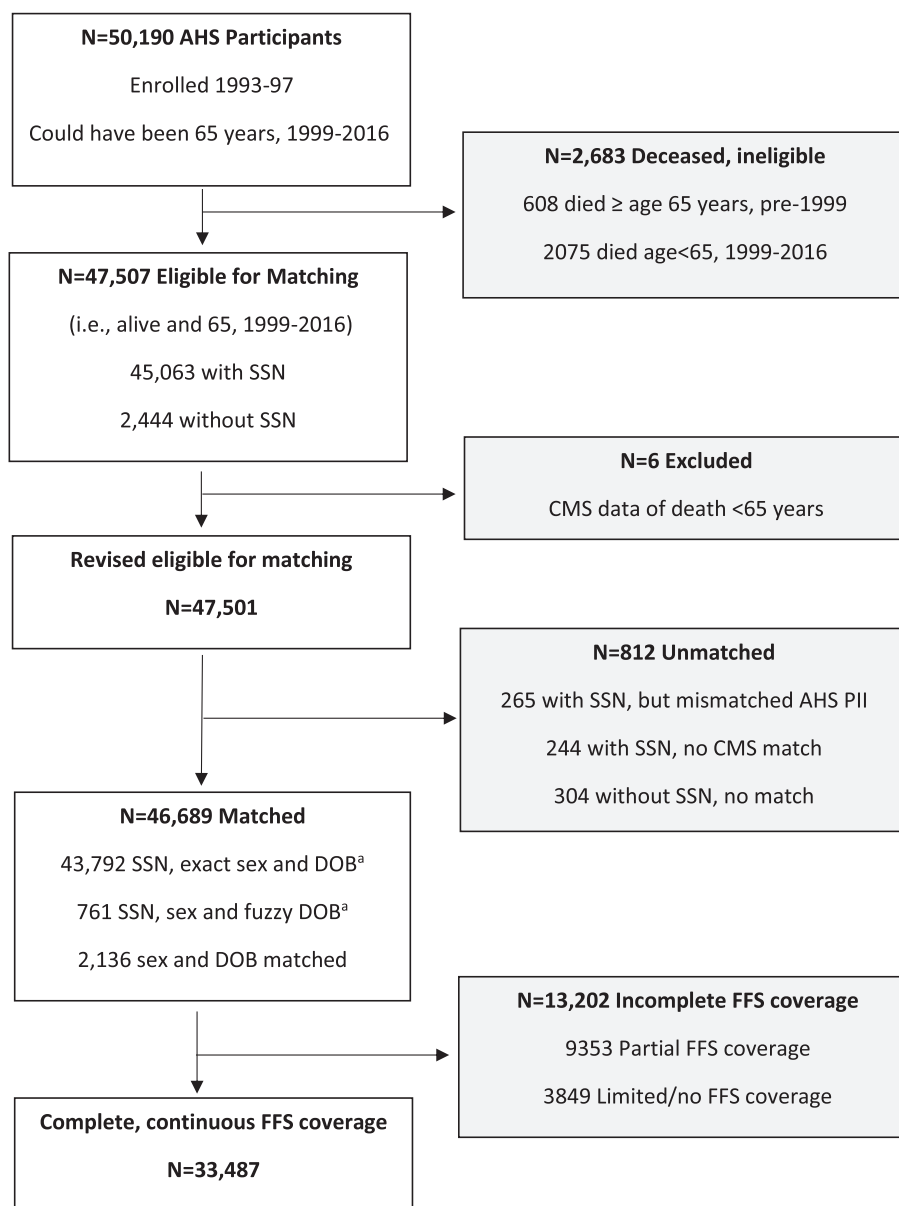


Fig. 1. AHS-Medicare Linkage (1999–2016): participants matched to CMS records and sample Ns; SSN = social security number, DOB = date of birth, PII = personally identifying information ^aOf those exactly matched by sex and DOB, 95% were spouses and 99% were from NC. Of those with fuzzy DOB matching, 23% were spouses and 83% were from NC.

product term (state by exposure). To explore potential informative missingness in FFS data among private applicators (most pesticide users in the AHS), we also compared those with incomplete (partial or limited/no) versus complete FFS, examining numbers of acres and livestock raised, total years and days per year personally mixed or applied pesticides, and farm work, residence, and hours per day pesticide at the first follow-up. Because reasons for enrollment in Parts C and A may differ, especially depending on non-farm employment, we also ran models among those with incomplete data due to Part C. Last, we examined the odds of mortality for those with partial and limited/no FFS versus complete coverage adjusting for the same covariates.

3. Results

Of 47,501 eligible participants, 46,689 (98%) were matched to a CMS ID (Fig. 1): 33,487 (70%) had complete FFS (1 or more years Part AB insurance ≥ 11 months/year, without >1 month Part C or Part A-only) since age 65; 9353 (20%) had partial FFS (≥ 1 years continuous

FFS, but with >1 month in Part C or Part A-only), and 3849 (8%) had limited/no FFS (>1 month Part A or Part C only, or <1 year FFS). FFS coverage varied by age and state (Table 1); participants with partial or limited/no FFS were more often from NC (47% versus 35% with complete FFS). Although 39% of the eligible sample was from NC, 90% of unmatched participants were from NC. Supplemental Table 1 shows that those ineligible for matching were also more often from NC, including those deceased before 1999 or age 65 (69% and 47%, respectively); altogether, a greater proportion of cohort participants from Iowa (74%) versus NC (58%) could be linked to Medicare data with complete FFS coverage.

State stratified covariate frequencies are shown in Supplemental Table 2, with adjusted ORs in Table 2. Compared to complete FFS coverage, the odds of partial FFS were higher for ages 50–59 years at AHS enrollment and lower for ≥ 70 (versus < 50), while odds of limited/no FFS decreased with increasing age, with similar patterns by state. In IA, odds of partial and limited/no FFS were higher in commercial (versus private) applicators, and lower for females (not seen in

Table 1

AHS-Medicare linkage: eligible sample (alive and 65, 1999–2016), by matching status and FFS coverage.

| Characteristic | Eligible sample | Matched, continuous FFS coverage ^a | | | Unmatched ^b |
|---|---------------------|---|-------------------------|----------------------------|------------------------|
| | Total N = 47,501 | Complete FFS N = 33,487 | Partial FFS N = 9353 | Limited/no FFS N = 3849 | No CMS ID N = 812 |
| At AHS enrollment | | N (%) | N (%) | N (%) | N (%) |
| Participant age | | | | | |
| 40–49 | 13,590 (29) | 8992 (27) | 1824 (20) | 2578 (67) | 196 (24) |
| 50–59 | 18,423 (39) | 12,338 (37) | 4787 (51) | 998 (26) | 300 (37) |
| 60–69 | 11,957 (25) | 9193 (27) | 2346 (25) | 207 (5) | 211 (26) |
| 70+ | 3531 (7) | 2964 (9) | 396 (4) | 66 (2) | 105 (13) |
| State | | | | | |
| Iowa | 28,931 (61) | 21,840 (65) | 4958 (53) | 2055 (53) | 78 (10) |
| North Carolina | 18,570 (39) | 11,647 (35) | 4395 (47) | 1794 (47) | 734 (90) |
| Participant type | | | | | |
| Private applicator | 28,507 (60) | 20,186 (60) | 5664 (59) | 2338 (61) | 419 (52) |
| Commercial applicator | 1457 (3) | 816 (2) | 414 (4) | 221 (6) | 6 (1) |
| Spouse | 17,537 (37) | 12,485 (37) | 3375 (36) | 1290 (34) | 387 (48) |
| Sex | | | | | |
| Male | 29,256 (62) | 20,515 (61) | 5839 (62) | 2490 (65) | 412 (51) |
| Female | 18,245 (38) | 12,972 (39) | 3514 (38) | 1359 (35) | 400 (49) |
| Race/ethnicity ^b | | | | | |
| White | 46,172 (97) | 32,797 (98) | 8980 (96) | 3732 (97) | 663 (82) |
| Non-white | 1273 (3) | 681 (2) | 373 (4) | 114 (3) | 105 (13) |
| Missing | 56 (0) | NS | NS | NS | 44 (5) |
| Education | | | | | |
| <High school | 5316 (11) | 3733 (11) | 1076 (12) | 285 (7) | 222 (27) |
| High school | 21,160 (45) | 15,287 (46) | 4071 (44) | 1519 (39) | 283 (35) |
| Some college | 9951 (21) | 7015 (21) | 1859 (20) | 962 (25) | 115 (14) |
| College graduate | 7475 (16) | 4995 (15) | 1598 (17) | 792 (21) | 90 (11) |
| Missing | 3599 (8) | 2457 (7) | 749 (8) | 291 (8) | 102 (13) |
| Smoking | | | | | |
| Never | 25,546 (54) | 18,185 (54) | 4953 (53) | 2016 (52) | 392 (48) |
| Past | 15,360 (32) | 10,846 (32) | 3145 (34) | 1147 (30) | 222 (27) |
| Current | 5571 (12) | 3758 (11) | 1058 (11) | 612 (16) | 143 (18) |
| Missing | 1024 (2) | 698 (2) | 197 (2) | 74 (2) | 55 (7) |
| Most recent AHS Survey^c | | | | | |
| Enrollment only | 6493 (14) | 4620 (14) | 1068 (11) | 553 (14) | 252 (31) |
| Take-home (appl) | 1876 (4) | 1458 (4) | 261 (3) | 129 (3) | 28 (3) |
| Phase 2 (1999–2003) | 6602 (14) | 4864 (15) | 1083 (12) | 528 (14) | 127 (16) |
| Phase 3 (2005–2010) | 6284 (13) | 4355 (13) | 1262 (13) | 532 (14) | 135 (17) |
| Phase 4 (2013–2015) | 26,246 (55) | 18,190 (54) | 5679 (61) | 2107 (55) | 270 (33) |

FFS = fee for service (part AB, but not C); NS = Not shown due to <10 individuals per category.

^a Eligible if alive and 65, 1999–2016; Complete FFS coverage starting age 65 is defined as having all years with ≥ 11 months continuous Part AB per year, ≥ 1 year, and never Part C; Partial FFS includes ≥ 1 year continuous, but not complete FFS, i.e., at least some Part A-only or Part C (see [Table 2](#)); Limited/no FFS includes short-term FFS (<11 months per year), Part A-only or Part C and other scenarios, including those with a CMS ID but no records (N = 51).

^b Non-white includes 1017 black, 256 other; Race is based on AHS self-report except where missing when filled in using CMS race (N = 865; 605 Complete, 190 Partial, 80 limited/no FFS); 608 (92%) white, 52 black, 21 other.

^c Only applicators were eligible for take-home. After Phase 2, follow-up was limited to private applicators and spouses with at least 1 study activity prior to Phase 3 (i.e., take-home or Phase 2 survey).

Table 2

AHS-Medicare linkage (1999–2016): participant characteristics associated with partial and limited/no FFS coverage compared to complete lifetime FFS coverage, by state.

| | Iowa | | North Carolina | | Interactions by state | |
|---|--------------------------|-----------------------------|--------------------------|-----------------------------|--------------------------|-----------------------------|
| | Partial FFS ^a | Limited/no FFS ^a | Partial FFS ^a | Limited/no FFS ^a | Partial FFS ^a | Limited/no FFS ^a |
| | OR ^b (95% CI) | OR ^b (95% CI) | OR ^b (95% CI) | OR ^b (95% CI) | P-value ^b | P-value ^b |
| AHS enrollment age | | | | | | |
| 40 to 49 | Referent | Referent | Referent | Referent | | |
| 50 to 59 | 1.95 (1.80, 2.12) | 0.29 (0.26, 0.32) | 1.76 (1.59, 1.95) | 0.27 (0.23, 0.30) | 0.12 | 0.32 |
| 60 to 69 | 1.28 (1.17, 1.41) | 0.06 (0.04, 0.07) | 1.07 (0.96, 1.19) | 0.09 (0.07, 0.11) | 0.01 | 0.004 |
| 70+ | 0.72 (0.60, 0.87) | 0.06 (0.04, 0.10) | 0.51 (0.43, 0.60) | 0.06 (0.05, 0.09) | 0.006 | 0.94 |
| Participant type | | | | | | |
| Private Applicator | Referent | Referent | Referent | Referent | | |
| Commercial Applicator | 2.36 (2.07, 2.69) | 2.00 (1.68, 2.33) | NA | NA | NA | NA |
| Spouse | 1.04 (0.76, 1.42) | 1.47 (0.93, 2.33) | 1.21 (0.98, 1.49) | 0.96 (0.71, 1.29) | 0.43 | 0.13 |
| Sex | | | | | | |
| Male | Referent | Referent | Referent | Referent | | |
| Female | 0.78 (0.72, 0.83) | 0.84 (0.75, 0.93) | 1.08 (1.00, 1.18) | 1.02 (0.90, 1.15) | <0.001 | 0.018 |
| Race/ethnicity | | | | | | |
| White | Referent | | Referent | Referent | | |
| Non-white | 1.76 (0.89, 3.49) | 1.28 (0.37, 4.45) | 1.75 (1.50, 2.04) | 1.10 (0.85, 1.42) | 0.98 | 0.81 |
| Education | | | | | | |
| <HS | 1.05 (0.93, 1.20) | 1.12 (0.86, 1.46) | 1.07 (0.96, 1.19) | 1.08 (0.91, 1.29) | 0.88 | 0.83 |
| HS | Referent | Referent | Referent | Referent | | |
| Some college | 1.05 (0.97, 1.14) | 0.95 (0.85, 1.07) | 1.10 (0.99, 1.22) | 0.94 (0.81, 1.08) | 0.47 | 0.87 |
| College Grad | 1.02 (0.93, 1.12) | 0.97 (0.86, 1.43) | 1.81 (1.64, 2.01) | 1.07 (0.92, 1.25) | <0.001 | 0.35 |
| Smoking history | | | | | | |
| Never | Referent | Referent | Referent | Referent | | |
| Past | 1.03 (0.95, 1.10) | 1.02 (0.91, 1.14) | 0.93 (0.85, 1.01) | 0.95 (0.83, 1.08) | 0.09 | 0.29 |
| Current | 1.05 (0.93, 1.17) | 1.05 (0.90, 1.23) | 0.76 (0.68, 0.86) | 0.99 (0.85, 1.15) | <0.001 | 0.43 |
| Self-reported conditions^c | | | | | | |
| None | Referent | Referent | Referent | Referent | | |
| One | 0.90 (0.83, 0.97) | 1.05 (0.93, 1.20) | 0.94 (0.86, 1.03) | 0.96 (0.83, 1.11) | 0.44 | 0.37 |
| Two or more | 0.83 (0.70, 0.97) | 0.97 (0.73, 1.31) | 0.66 (0.56, 1.38) | 0.79 (0.60, 1.05) | 0.06 | 0.33 |
| Missing | 0.92 (0.75, 1.15) | 1.00 (0.70, 1.42) | 1.21 (1.06, 1.38) | 1.22 (0.99, 1.51) | 0.04 | 0.33 |

FFS = fee for service (part AB, but not C); NA is not applicable, as no commercial applicators enrolled in North Carolina.

^a Total Ns and frequencies are shown in [Supplemental Table 2](#). Complete FFS coverage starting age 65 was defined as ≥ 11 months continuous Part AB per year, ≥ 1 year, never Part C, throughout the duration of Medicare enrollment; Partial FFS includes ≥ 1 year continuous, but not complete FFS.^b Multinomial logistic regression used to calculate Odds Ratio (OR) and 95% Confidence Interval (CI), models were unadjusted for categorical age, and for all other variables adjusted for continuous age and age-squared, and mutually adjusted for sex, race/ethnicity, education, and smoking excluding missings. P-values show interaction terms by state for each level of the outcome.^c Summary of representative chronic conditions including self-reported doctor diagnosis of cancer (except non-melanoma skin), cardiovascular disease (heart disease or stroke), diabetes (except gestational), chronic lung disease, depression.

Table 3

AHS-Medicare linkage (1999–2016): Incomplete (partial or limited/no FFS) versus complete FFS in private applicators, in relation to enrollment and follow-up farm experience and pesticide use, by state.

| Characteristic | Iowa | | North Carolina | | Interaction |
|---|-----------------|--------------|-----------------|--------------|-------------|
| | OR ^b | (95% CI) | OR ^b | (95% CI) | P-value |
| At AHS enrollment | | | | | |
| Owned or worked on a farm In the past 12 months: | 0.75 | (0.52, 1.06) | 0.87 | (0.76, 1.00) | 0.414 |
| Number of acres planted | | | | | |
| None/didn't work on farm | 1.04 | (0.75, 1.44) | 1.09 | (0.92, 1.30) | 0.782 |
| <50 acres | 1.36 | (1.07, 1.44) | 1.38 | (1.22, 1.56) | 0.902 |
| 50–199 | Ref | | Ref | | |
| 200–499 | 0.70 | (0.63, 0.78) | 0.62 | (0.52, 0.73) | 0.192 |
| 500+ | 0.52 | (0.47, 0.58) | 0.46 | (0.38, 0.54) | 0.183 |
| Number of livestock raised | | | | | |
| None/didn't work on farm | 0.80 | (0.72, 0.90) | 0.68 | (0.61, 0.75) | 0.026 |
| <100 | Ref | | Ref | | |
| 100–499 | 0.84 | (0.75, 0.93) | 0.78 | (0.64, 0.94) | 0.515 |
| 500+ | 0.73 | (0.66, 0.82) | 0.41 | (0.32, 0.53) | <0.0001 |
| Lifetime mixed/applied pesticides: | | | | | |
| Total years | | | | | |
| <11 | Ref | | Ref | | |
| 11–20 | 0.98 | (0.87, 1.11) | 0.92 | (0.82, 1.03) | 0.427 |
| 21–30 | 0.85 | (0.76, 0.99) | 0.87 | (0.77, 0.98) | 0.797 |
| >30 | 0.85 | (0.75, 0.99) | 0.70 | (0.62, 0.80) | 0.042 |
| Days per year | | | | | |
| <5 | Ref | | Ref | | |
| 5–9 | 1.00 | (0.89, 1.11) | 0.90 | (0.79, 1.02) | 0.212 |
| 10–19 | 0.88 | (0.79, 0.99) | 0.72 | (0.64, 0.81) | 0.014 |
| 20+ | 0.87 | (0.77, 0.99) | 0.59 | (0.53, 0.67) | <0.0001 |
| At 1999–2003 follow-up | | | | | |
| Did not live or work on farm | Ref | | Ref | | |
| Lived on a farm, <i>did not work</i> | 0.90 | (0.72, 1.14) | 0.84 | (0.71, 0.99) | 0.596 |
| Worked on a farm, <i>did not live</i> | 0.74 | (0.59, 0.93) | 0.92 | (0.75, 1.11) | 0.172 |
| Lived and worked on farm | 0.66 | (0.54, 0.80) | 0.88 | (0.76, 1.00) | 0.018 |
| Mixed/applied pesticides | | | | | |
| Hours per day | | | | | |
| None | Ref | | Ref | | |
| 1–2 | 1.15 | (0.98, 1.35) | 1.10 | (0.96, 1.26) | 0.679 |
| 3–5 | 1.16 | (0.99, 1.36) | 1.15 | (0.99, 1.34) | 0.952 |
| 6–8 | 0.97 | (0.84, 1.12) | 0.80 | (0.68, 0.95) | 0.093 |
| >8 | 0.83 | (0.72, 0.97) | 0.69 | (0.54, 0.89) | 0.216 |

FFS = fee for service (part AB, but not C).

^aLifetime complete FFS coverage starting age 65 was defined as ≥ 11 months continuous Part AB per year, ≥ 1 year, never Part C; Partial FFS includes ≥ 1 year continuous but not complete FFS (see Table 3).^b Logistic regression models used to calculate Odds Ratio (OR) and 95% Confidence Interval (CI), adjusted for AHS enrollment age and age-squared, sex, race, education, and smoking.

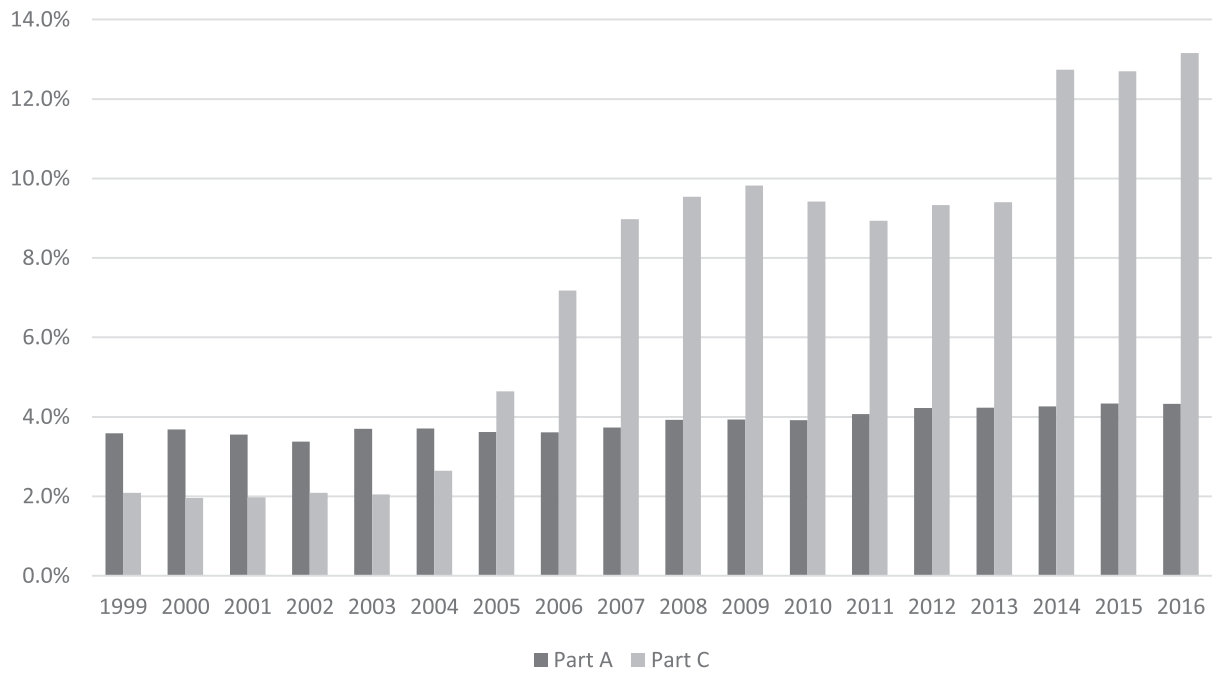


Fig. 2. Percent of total AHS-Medicare linked participants enrolled in Part C and Part A (1999–2016). Dark bars indicate Part A and grey bars indicate Part C enrollment.

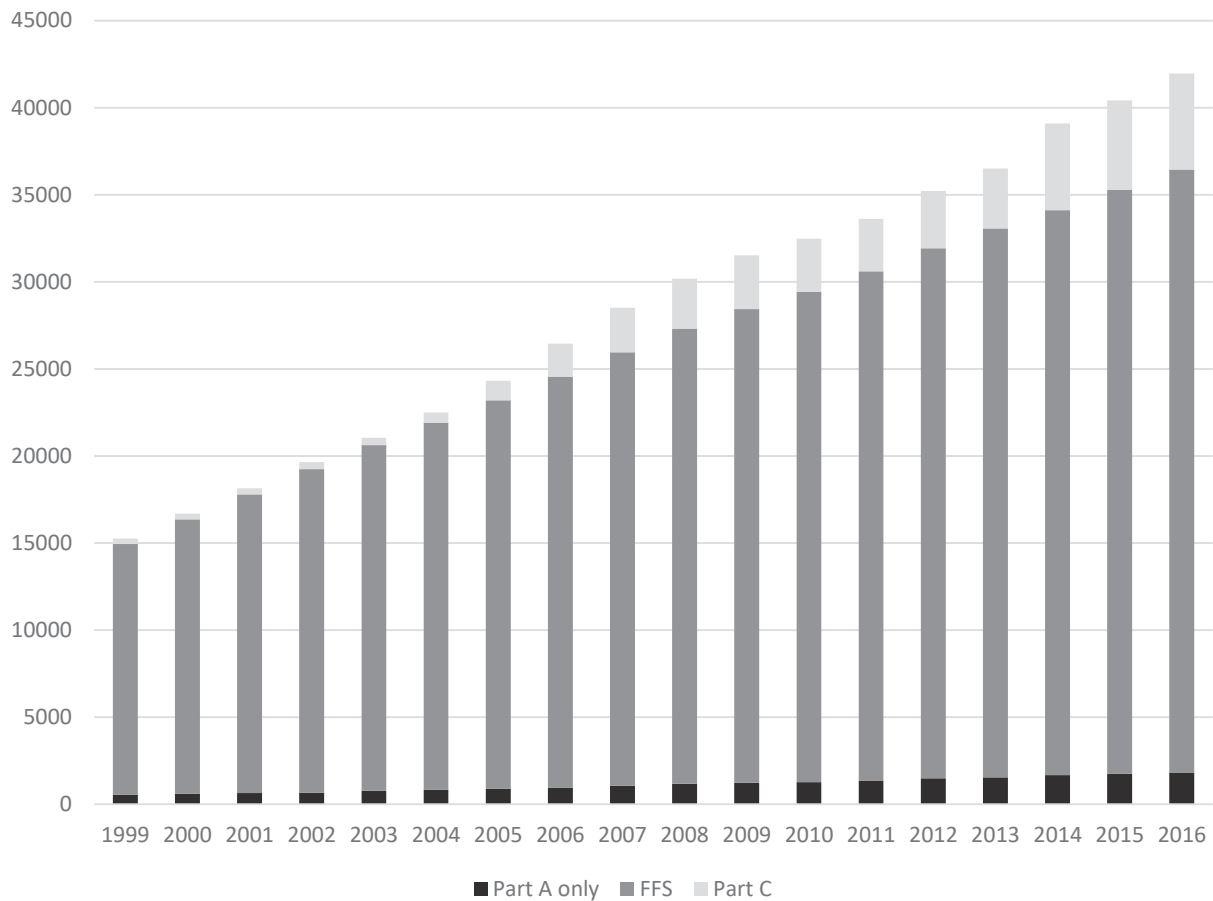


Fig. 3. Total number of Medicare-linked AHS participants with FFS, Part A and Part C by year (1999–2016). The combined height of the bars shows the increasing number of individuals entering Medicare, with dark grey indicating the majority with 11 or more months FFS per year. Light grey shows the growing number with Part C, and a smaller but increasing number with Part-A only in the dark color.

Table 4
Mortality in AHS private applicators with partial or limited/no FFS versus complete FFS coverage, by state.

| | Complete FFS ^b | Partial FFS ^b | Odds Ratio ^{c,d} (95% CI) | Limited/no FFS ^b | Odds Ratio ^{c,e} (95% CI) |
|---|---------------------------|--------------------------|------------------------------------|-----------------------------|------------------------------------|
| Total N and N (%) deceased participants, by state | | | | | |
| Iowa | N = 12,008 | N = 2726 | | N = 1103 | |
| Deceased ^d | 2979 (25) | 457 (17) | 0.66 (0.58, 0.74) | 100 (9) | 1.37 (1.06, 1.72) |
| North Carolina | N = 7016 | N = 2,471 | | N = 1106 | |
| Deceased ^d | 3132 (45) | 608 (25) | 0.40 (0.31, 0.46) | 185 (17) | 1.00 (0.81, 1.23) |

^a Died while enrolled in Medicare, based on linkage to national death index.

^b Lifetime complete FFS coverage starting age 65 was defined as ≥ 11 months continuous Part AB per year, ≥ 1 year, never Part C or Part A only; Partial FFS includes ≥ 1 year continuous FFS and some Part A-only or Part C.

^c Multinomial logistic regression, adjusting for baseline age, age-squared, state, race-ethnicity, education, and smoking history.

^d P-value for interaction by state < 0.0001.

^e P-value for interaction by state = 0.068.

NC). In both states, the odds of partial FFS were higher in non-whites. In NC, the odds of partial FFS coverage were increased with a college degree (versus high school) but decreased among current (versus never) smokers. Having two or more (versus none) of the chronic conditions was inversely associated with partial FFS. In NC, the odds of incomplete FFS coverage were greater in those who were also missing questionnaire data on health conditions. Patterns of FFS coverage varied over time, and by state (Supplemental Table 3). More NC participants enrolled due to disability/ESRD (8–10%) than IA (3–4%). In IA, 52% of partial FFS was due to Part C enrollment, compared to 38% of those with limited/no FFS. By contrast, Part C was the dominant reason for partial and limited FFS in NC (83 and 73% respectively). Utilization was similar by FFS status and by state, except for higher emergency room use in NC.

Fig. 2 illustrates the percent of Medicare-linked cohort-participants enrolled in Part C and Part-A only, by year. Part C enrollment increased from 2% in 1999–2003 to 9.8% by 2009 following the MMA in 2005, rising to 12.7% in 2014 following implementation of the ACA. Part-A enrollment increased slightly from 3.6% in 1999 to 4.3% by 2014. Increased Part C was driven by changes in NC (not shown); by 2016, 30% of NC participants were enrolled in Part C. Fig. 3 shows the total number of participants with 11 months of FFS coverage by year, as well as those with Part C, or Part A-only.

In private applicators, despite differences in farm characteristics by state (Supplemental Table 4), we saw similar patterns of association by farm size and pesticide use (Table 4). The odds of incomplete FFS were increased among participants with smaller farms and decreased with greater pesticide use. Odds of incomplete coverage were lower for those in the first follow-up who lived or worked on a farm (versus neither), and those with more hours per day pesticide use. Similar results were seen in sensitivity analyses focused on those with Part C (Supplemental Table 5), except for a loss in heterogeneity by state for the strength of associations with greater pesticide use.

During follow-up (1999–2016), 32% with complete FFS, 21% with partial, and 13% with limited/no FFS coverage died (Table 4). Compared to complete FFS, those with partial FFS had decreased mortality odds, especially in NC, while in IA those with limited/no FFS had increased mortality.

4. Discussion

In the Agricultural Health Study, many of the 89,000 enrolled licensed pesticide applicators and their spouses have remained active participants over the past two decades. As in many aging cohorts over extended time periods, loss to follow-up is a growing concern that may impact the internal validity, generalizability, or feasibility of research based solely on questionnaire data. Linkage to U.S. Medicare claims data provides a complementary resource for health outcomes research in the AHS. The vast majority of the eligible sample (65 and alive, 1999–2016) was matched to Medicare identifiers, comparable to other cohort linkages (Kucharska-Newton et al., 2016; Virnig et al., 2010; Xie et al., 2016). Most coverage was prior to 2015, when available claims data are limited to traditional FFS coverage and Part A insurance. Only 70% of the linked sample had complete FFS coverage throughout Medicare enrollment, effectively reducing sample size for some analyses.

Incomplete FFS coverage may pose threats to internal and external validity. We saw differences by age, state, sociodemographic factors, baseline health conditions (potential adjustment factors available on all participants). Prior studies show that Part C enrollment varies by socioeconomic factors and preexisting chronic diseases (Meyers et al., 2019; Morgan et al., 1997; Park et al., 2020; Rahman et al., 2015). There is also evidence of lower mortality associated with Part C, which may diminish over time (Krumholz et al., 2015; Newhouse et al., 2019). In the AHS, incomplete FFS (and Part C) coverage was more common in NC, and in adjusted models was associated with smaller farm size and less pesticide use. Mortality odds were decreased among those with partial coverage in both states, while those with limited or no coverage

had greater mortality odds in IA. Analyses of linked cohorts are often limited to those with complete FFS coverage (i.e., at least 11 months continuous FFS and no Part C insurance). Our findings suggest this approach will over-represent IA participants, those with more farming exposures and pesticide use, and with increased mortality odds. Impacts on specific analyses, however, will depend on both the exposures and outcomes of interest.

We confirmed expected shifts in FFS enrollment in 2006–7, following the MMA, and in 2011–14 with implementation of the ACA, impacting the completeness of data for older participants who may have enrolled in managed care, and for younger participants who may also opt out of Part B (e.g., due to other sources of coverage). Greater uptake of Part C in NC participants was consistent with greater use of cost-saving programs in NC (Dranove et al., 1998; Ortiz et al., 2015), while increases by non-white race/ethnicity are also consistent with recent national trends (Meyers et al., 2021). Our findings highlight the importance of considering programmatic changes to Medicare in making comparisons within cohorts over time. In the AHS (and other cohorts), comparisons of Medicare-based findings with the general population must also account for the unique features of the study population, such as state and rurality. Rural residents may have differential access to healthcare, resulting in differing utilization, for example emergency room services (Toth et al., 2015).

At age 65, Part A insurance coverage is automatic, requires no premiums, and may be paired with other types of insurance (e.g., employment or Veteran's benefits). Some lacking financial means may decline Part B, while others may choose Part C given financial subsidies. Studies have shown healthier individuals tend to enroll and stay in managed care (Mello et al., 2003; Meyers et al., 2019; Newhouse et al., 2019; Park et al., 2020; Rahman et al., 2015), while leaving may be related to socioeconomic factors either contributing to or resulting from health disparities. In the AHS, 1 in 5 participants with partial FFS lacked continuous coverage, which could also reflect non-payment of premiums (these details were beyond the scope of the current investigation). Differences in FFS coverage by farm size and pesticide use could reflect off-farm or spousal employment or other socioeconomic factors, but we lacked questionnaire data on income, outside employment, and other insurance coverage.

Analyses of Medicare-linked cohorts often exclude those with Part C coverage; at extra cost, Part C claims data are now available for studies starting in 2015. Few studies describe frequencies of Part A-only coverage, which may provide some data for case ascertainment based on hospitalization, but not outpatient care. Having a common denominator is important when comparing rates within a study (Setiawan et al., 2016), or with other populations such as the Medicare 5% sample, which may benefit from the application of weights to normalize by demographic factors (DiMartino et al., 2009; Parikh et al., 2017). While many studies acknowledge the limitation of missing claims data, only a few present the systematic comparisons needed to evaluate potential biases. One study compared cohort participants with FFS (63%) to those with managed care only (37%), describing differences by state and race/ethnicity similar to those in the AHS (Setiawan et al., 2015). Association studies may use a nested case-control design, matching controls to cases by duration of FFS, including those with at least some but not complete FFS (Noureddin et al., 2020). Some claims data are available for participants with incomplete FFS, including partial years with FFS or Part A data on inpatient diagnoses. Neither restricting nor matching solves the problem of informative missingness, however, which may instead be addressed using strategies such as inverse probability weighting (IPW) or propensity scores to account for determinants of missing FFS data (Bergquist et al., 2018). A large portion of AHS participants enrolled in Medicare prior to 1999, resulting in left truncation, another type of missing data that impacts ascertainment of incident outcomes. Restricting analyses to those who enrolled in Medicare \geq 1999 may mitigate this somewhat.

Our study has limitations. Findings may not be generalizable due to

the largely rural population and farm families represented by this cohort; however, they provide novel information on linking both within a rural setting and with an occupational cohort that is largely self-employed. While they may not apply to studies in other countries, some findings may exemplify universal challenges in data linkages. AHS participants are primarily white male farmers and their spouses. Matches were not made for 2% of eligible participants, mostly non-white, older, or female participants from NC. Participants from NC and spouses were more likely to require matching by characteristics other than Social Security number (not shown), consistent with other studies showing differences for women and minorities, and those with insufficient work histories to qualify for Social Security (Miller et al., 2017). In the AHS, similar demographic factors have been related to lost to follow-up (Rinsky et al., 2017), which has been addressed using IPW in studies of survey-based outcomes. We replicated expected associations with mortality after adjusting for demographics and baseline smoking, but did not investigate the role of baseline health covariates or diagnoses in linked claims data. Our investigation here simply describes the potential for bias due to missing claims data, providing background for future bias assessment and mitigation strategies that will need to account for heterogeneity by state, age and other covariates.

5. Conclusion

Cohort linkage to Medicare claims data provides important opportunities to confirm self-reported diagnoses and ascertain cases in participants lost to active follow-up or for outcomes not included on surveys. Our findings revealed patterned missingness in FFS data, which also may be a concern in other cohorts. The impact of programmatic changes to Medicare warrants further consideration in the design and interpretation of research in linked cohorts.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

This work was supported by the Intramural Research Program of the National Institutes of Health (NIH), National Cancer Institute, Division of Cancer Epidemiology and Genetics (Z01CP010119), and the NIH/National Institute of Environmental Health (Z01ES049030).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.pmedr.2022.101766>.

References

- Afendulis, C.C., Landrum, M.B., Chernew, M.E., 2012. The impact of the Affordable Care Act on Medicare Advantage plan availability and enrollment. *Health services research* 47 (6), 2339–2352.
- Alavanja, M.C., Sandler, D.P., McMaster, S.B., Zahm, S.H., McDonnell, C.J., Lynch, C.F., Pennybacker, M., Rothman, N., Dosemeci, M., Bond, A.E., Blair, A., 1996. The Agricultural Health Study. *Environ Health Perspect* 104 (4), 362–369.
- Bergquist, S.L., McGuire, T.G., Layton, T.J., Rose, S., 2018. Sample Selection for Medicare Risk Adjustment Due to Systematically Missing Data. *Health services research* 53 (6), 4204–4223.
- Chan, L., Hart, L.G., Goodman, D.C., 2006. Geographic access to health care for rural Medicare beneficiaries. *The Journal of rural health : official journal of the American Rural Health Association and the National Rural Health Care Association* 22 (2), 140–146.
- DiMartino, L.D., Hammill, B.G., Curtis, L.H., Gottdiener, J.S., Manolio, T.A., Powe, N.R., Schulman, K.A., 2009. External validity of the cardiovascular health study: a comparison with the Medicare population. *Medical care* 47, 916–923.
- Dranove, D., J. Simon, C., White, W.D., 1998. Determinants of managed care penetration. *J Health Econ* 17 (6), 729–745.

- Geissler, K.H., 2020. Differences in referral patterns for rural primary care physicians from 2005 to 2016. *Health services research* 55 (1), 94–102.
- Hlatky, M.A., Ray, R.M., Burwen, D.R., Margolis, K.L., Johnson, K.C., Kucharska-Newton, A., Manson, J.E., Robinson, J.G., Safford, M.M., Allison, M., Assimes, T.L., Bavry, A.A., Berger, J., Cooper-DeHoff, R.M., Heckbert, S.R., Li, W., Liu, S., Martin, L.W., Perez, M.V., Tindle, H.A., Winkelmayer, W.C., Stefanick, M.L., 2014. Use of Medicare data to identify coronary heart disease outcomes in the Women's Health Initiative. *Circulation. Cardiovascular quality and outcomes* 7 (1), 157–162.
- Jacobson, G.A., Damico, A., Neuman, P., Gold, M., 2017. Medicare Advantage 2017 Spotlight: Enrollment Market Update. Kaiser Family Foundation, Henry J.
- Jacobson, G.A., Neuman, P., Damico, A., 2015. At least half of new Medicare advantage enrollees had switched from traditional Medicare during 2006–11. *Health affairs (Project Hope)* 34 (1), 48–55.
- Jain, S., Himali, J., Beiser, A., Ton, T.G.N., Kelly-Hayes, M., Biggs, M.L., Delaney, J.A.C., Rosano, C., Seshadri, S., Frank, S.A., 2015. Validation of secondary data sources to identify Parkinson disease against clinical diagnostic criteria. *American journal of epidemiology* 181 (3), 185–190.
- Johnson, E.L., Krauss, G.L., Lee, A.K., Schneider, A.L.C., Dearborn, J.L., Kucharska-Newton, A.M., Huang, J., Alonso, A., Gottesman, R.F., 2018. Association Between Midlife Risk Factors and Late-Onset Epilepsy: Results From the Atherosclerosis Risk in Communities Study. *JAMA neurology* 75, 1375–1382.
- Kalbaugh, C.A., Kucharska-Newton, A., Wruck, L., Lund, J.L., Selvin, E., Matsushita, K., Bengtson, L.G.S., Heiss, G., Loehr, L., 2017. Peripheral Artery Disease Prevalence and Incidence Estimated From Both Outpatient and Inpatient Settings Among Medicare Fee-for-Service Beneficiaries in the Atherosclerosis Risk in Communities (ARIC) Study. *Journal of the American Heart Association* 6 (5).
- Krumholz, H.M., Nuti, S.V., Downing, N.S., Normand, S.L., Wang, Y., 2015. Mortality, Hospitalizations, and Expenditures for the Medicare Population Aged 65 Years or Older, 1999–2013. *Jama* 314, 355–365.
- Kucharska-Newton, A.M., Heiss, G., Ni, H., Stearns, S.C., Puccinelli-Ortega, N., Wruck, L.M., Chambless, L., 2016. Identification of Heart Failure Events in Medicare Claims: The Atherosclerosis Risk in Communities (ARIC) Study. *Journal of cardiac failure* 22 (1), 48–55.
- Lakshminarayan, K., Larson, J.C., Virnig, B., Fuller, C., Allen, N.B., Limacher, M., Winkelmayer, W.C., Safford, M.M., Burwen, D.R., 2014. Comparison of Medicare claims versus physician adjudication for identifying stroke outcomes in the Women's Health Initiative. *Stroke; a journal of cerebral circulation* 45 (3), 815–821.
- McGuire, T.G., Newhouse, J.P., Sinaiko, A.D., 2011. An economic history of Medicare part C. *The Milbank quarterly* 89, 289–332.
- Mello, M.M., Stearns, S.C., Norton, E.C., Ricketts 3rd, T.C., 2003. Understanding biased selection in Medicare HMOs. *Health services research* 38, 961–992.
- Meyers, D.J., Belanger, E., Joyce, N., McHugh, J., Rahman, M., Mor, V., 2019. Analysis of Drivers of Disenrollment and Plan Switching Among Medicare Advantage Beneficiaries. *JAMA internal medicine* 179, 524–532.
- Meyers, D.J., Mor, V., Rahman, M., Trivedi, A.N., 2021. Growth In Medicare Advantage Greatest Among Black And Hispanic Enrollees. *Health affairs (Project Hope)* 40 (6), 945–950.
- Miller, E.A., McCarty, F.A., Parker, J.D., 2017. Racial and Ethnic Differences in a Linkage with the National Death Index. *Ethnicity & disease* 27, 77–84.
- Morgan, R.O., Virnig, B.A., DeVito, C.A., Persily, N.A., 1997. The Medicare-HMO revolving door—the healthy go in and the sick go out. *The New England journal of medicine* 337 (3), 169–175.
- Mues, K.E., Liede, A., Liu, J., Wetmore, J.B., Zaha, R., Bradbury, B.D., Collins, A.J., Gilbertson, D.T., 2017. Use of the Medicare database in epidemiologic and health services research: a valuable source of real-world evidence on the older and disabled populations in the US. *Clinical epidemiology* 9, 267–277.
- Newhouse, J.P., Price, M., McWilliams, J.M., Hsu, J., Souza, J., Landon, B.E., 2019. Adjusted Mortality Rates Are Lower For Medicare Advantage Than Traditional Medicare, But The Rates Converge Over Time. *Health affairs (Project Hope)* 38 (4), 554–560.
- Noureddin, M., Zelber-Sagi, S., Wilkens, L.R., Porcel, J., Boushey, C.J., Le Marchand, L., Rosen, H.R., Setiawan, V.W., 2020. Diet Associations With Nonalcoholic Fatty Liver Disease in an Ethnically Diverse Population: The Multiethnic Cohort. *Hepatology* 71 (6), 1940–1952.
- Ortiz, J., Hofler, R.A., Lin, Y.L., Berzon, R., 2015. Participation of Rural Health Care Providers in Accountable Care Organizations: Early Indications. *The health care manager* 34, 255–264.
- Parikh, K.S., Greiner, M.A., Wang, W., Min, Y.-I., Correa, A., Banahan, B.F., Curtis, L.H., Hernandez, A.F., O'Brien, E.C., Mentz, R.J., 2017. Representativeness of Medicare Participants in the Jackson Heart Study for African American Medicare Beneficiaries. *Epidemiology (Cambridge)* 28 (5), 740–746.
- Park, S., Fishman, P., White, L., Larson, E.B., Coe, N.B., 2020. Disease-Specific Plan Switching Between Traditional Medicare and Medicare Advantage. *The Permanente journal* 24.
- Polinski, J.M., Brookhart, M.A., Ayanian, J.Z., Katz, J.N., Kim, S.C., Lii, J., Tonner, C., Yelin, E., Solomon, D.H., 2014. Relationships between driving distance, rheumatoid arthritis diagnosis, and disease-modifying antirheumatic drug receipt. *Arthritis care & research* 66 (11), 1634–1643.
- Rahman, M., Keohane, L., Trivedi, A.N., Mor, V., 2015. High-Cost Patients Had Substantial Rates Of Leaving Medicare Advantage And Joining Traditional Medicare. *Health affairs (Project Hope)* 34 (10), 1675–1681.
- Rinsky, J.L., Richardson, D.B., Wing, S., Beard, J.D., Alavanja, M., Beane Freeman, L.E., Chen, H., Henneberger, P.K., Kamel, F., et al., 2017. Assessing the Potential for Bias From Nonresponse to a Study Follow-up Interview: An Example From the Agricultural Health Study. *American journal of epidemiology* 186, 395–404.
- Setiawan, V.W., Stram, D.O., Porcel, J., Lu, S.C., Le Marchand, L., Noureddin, M., 2016. Prevalence of chronic liver disease and cirrhosis by underlying cause in understudied ethnic groups: The multiethnic cohort. *Hepatology* 64 (6), 1969–1977.
- Setiawan, V.W., Virnig, B.A., Porcel, J., Henderson, B.E., Le Marchand, L., Wilkens, L.R., Monroe, K.R., 2015. Linking data from the Multiethnic Cohort Study to Medicare data: linkage results and application to chronic disease research. *American journal of epidemiology* 181, 917–919.
- Ton, T.G.N., Biggs, M.L., Comer, D., Curtis, L., Hu, S.-C., Thacker, E.L., Searles Nielsen, S., Delaney, J.A., Landsittel, D., Longstreth, W.T., Checkoway, H., Jain, S., 2014. Enhancing case ascertainment of Parkinson's disease using Medicare claims data in a population-based cohort: the Cardiovascular Health Study. *Pharmacoepidemiol Drug Saf* 23 (2), 119–127.
- Toth, M., Holmes, M., Van Houtven, C., Toles, M., Weinberger, M., Silberman, P., 2015. Rural Medicare Beneficiaries Have Fewer Follow-up Visits and Greater Emergency Department Use Postdischarge. *Medical care* 53, 800–808.
- Virnig, B., Durham, S.B., Folsom, A.R., Cerhan, J., 2010. Linking the Iowa Women's Health Study cohort to Medicare data: linkage results and application to hip fracture. *American journal of epidemiology* 172 (3), 327–333.
- Xie, F., Colantonio, L.D., Curtis, J.R., Safford, M.M., Levitan, E.B., Howard, G., Muntner, P., 2016. Linkage of a Population-Based Cohort With Primary Data Collection to Medicare Claims: The Reasons for Geographic and Racial Differences in Stroke Study. *American journal of epidemiology* 184 (7), 532–544.
- Zhou, L., Stearns, S.C., Thudium, E.M., Alburikan, K.A., Rodgers, J.E., 2015. Assessing Medicare Part D claim completeness using medication self-reports: the role of veteran status and Generic Drug Discount Programs. *Medical care* 53, 463–470.