



Practice of Epidemiology

Optimal Use of Available Claims to Identify a Medicare Population Free of Coronary Heart Disease

Shia T. Kent*, Monika M. Safford, Hong Zhao, Emily B. Levitan, Jeffrey R. Curtis, Ryan D. Kilpatrick, Meredith L. Kilgore, and Paul Muntner

* Correspondence to Dr. Shia T. Kent, Department of Epidemiology, School of Public Health, University of Alabama at Birmingham, 1700 University Boulevard, LHL 446, Birmingham, AL 39294 (e-mail: shia@uab.edu).

Initially submitted September 11, 2014; accepted for publication April 24, 2015.

We examined claims-based approaches for identifying a study population free of coronary heart disease (CHD) using data from 8,937 US blacks and whites enrolled during 2003–2007 in a prospective cohort study linked to Medicare claims. Our goal was to minimize the percentage of persons at study entry with self-reported CHD (previous myocardial infarction or coronary revascularization). We assembled 6 cohorts without CHD claims by requiring 6 months, 1 year, or 2 years of continuous Medicare fee-for-service insurance coverage prior to study entry and using either a fixed-window or all-available look-back period. We examined adding CHD-related claims to our “base algorithm,” which included claims for myocardial infarction and coronary revascularization. Using a 6-month fixed-window look-back period, 17.8% of participants without claims in the base algorithm reported having CHD. This was reduced to 3.6% using an all-available look-back period and adding other CHD claims to the base algorithm. Among cohorts using all-available look-back periods, increasing the length of continuous coverage from 6 months to 1 or 2 years reduced the sample size available without lowering the percentage of persons with self-reported CHD. This analysis demonstrates approaches for developing a CHD-free cohort using Medicare claims.

algorithms; bias (epidemiology); coronary disease; epidemiologic methods; Medicare

Abbreviations: CHD, coronary heart disease; E/M, evaluation and management; ICD-9, *International Classification of Diseases, Ninth Revision*; REGARDS, Reasons for Geographic and Racial Differences in Stroke.

Administrative data, such as those obtained from health insurance claims, are increasingly being used for health-care utilization studies, pharmacovigilance, comparative effectiveness research, and other epidemiologic investigations (1, 2). Administrative claims provide a cost-effective approach for conducting research, since data have already been collected. However, the analysis of claims data also presents challenges. For example, data on patient comorbidity and treatment history are often obtained using medical claims during a “look-back” period of a specified amount of time prior to study entry (3, 4). Obtaining medical history from a look-back period results in left-censoring of events that occur prior to the availability of claims.

Using claims to study coronary heart disease (CHD) incidence demonstrates the challenges of analyzing left-censored administrative data. Patients with CHD have a markedly increased risk of future health events and higher utilization of

health-care services; thus, analyses of incident and recurrent events are often performed separately (5). In claims-based analyses, investigation of incident CHD events is often conducted after the exclusion of persons with a history of CHD identified during a look-back period (6–8). There are trade-offs in selecting the length of a look-back period to exclude those with prevalent disease. Minimum lengths of continuous insurance coverage prior to study entry are often required, so that an adequate length of time is available for identifying a history of CHD. Look-back periods with short coverage requirements (e.g., 6 months) are attractive, because few beneficiaries will be excluded for not having insurance coverage. However, studies using short look-back periods can fail to accurately identify persons with prevalent disease (9–11).

In epidemiologic cohort studies with primary data collection, left-censoring is addressed by using a survey administered at

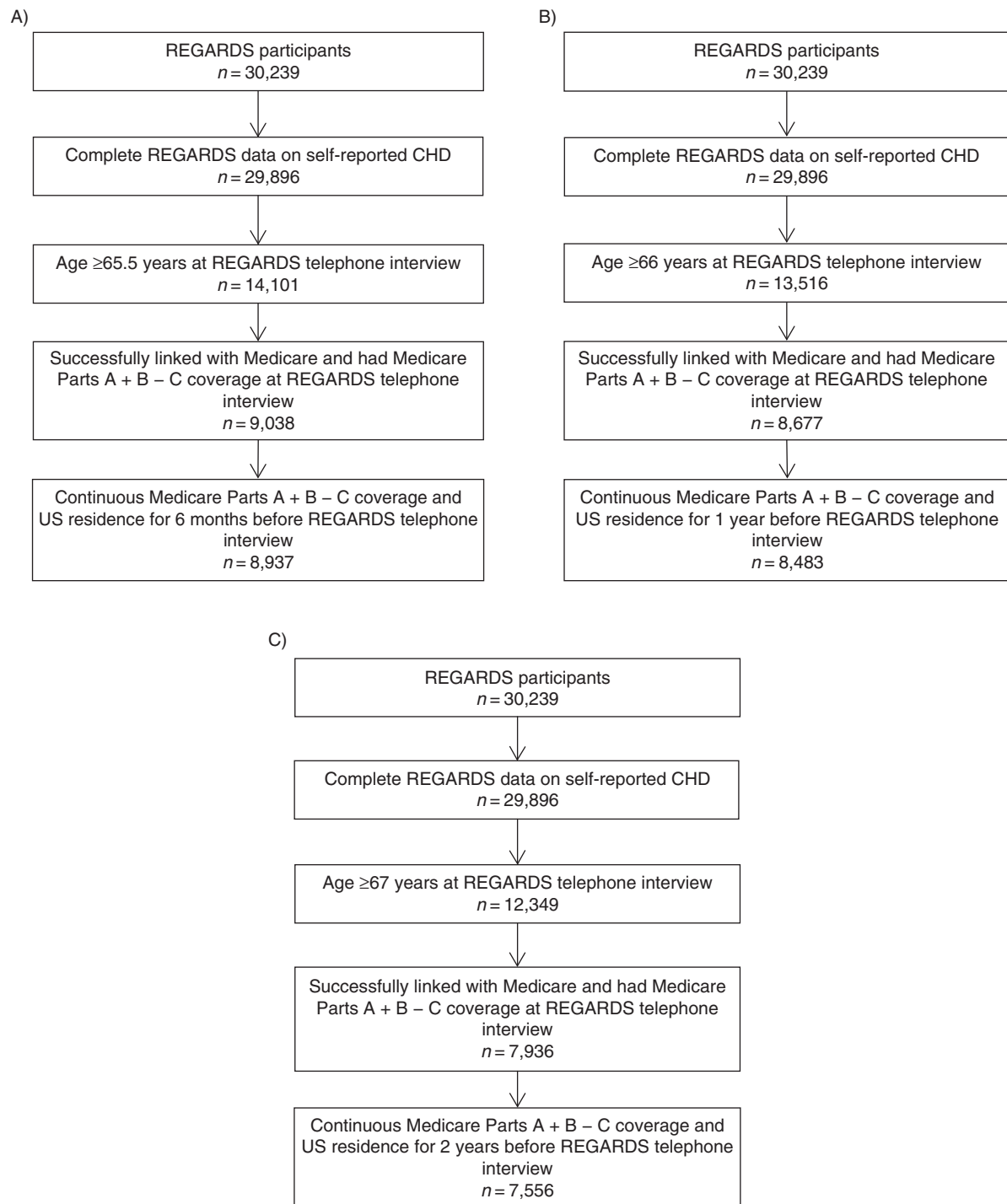


Figure 1. Selection of study populations without Medicare claims for coronary heart disease (CHD) using different lengths of required Medicare coverage (6 months (A), 1 year (B), or 2 years (C)) prior to the study telephone interview, Reasons for Geographic and Racial Differences in Stroke (REGARDS) Study, 1999–2007.

the time of enrollment, wherein participants are asked about their lifetime history of CHD. These studies can be linked with administrative data to develop and examine the performance of claims algorithms. We conducted analyses of

participants in the Reasons for Geographic and Racial Differences in Stroke (REGARDS) Study, a US population-based study in which data were previously linked with claims from Medicare, a federal health insurance program. Using

these data, we created CHD-free analytical samples by using look-back periods of varying durations and coverage requirements and by using several claims-based definitions of CHD. Our goal was to develop an approach for minimizing the percentage of persons with self-reported CHD, while examining the effect these approaches had on sample size.

METHODS

Data sources

The REGARDS Study was designed to investigate reasons underlying the higher rate of stroke mortality among US blacks as compared with whites and among residents of the southeastern United States as compared with other regions of the country. The study design and enrollment of participants has been described in detail elsewhere (12). In brief, 30,239 black and white men and women aged 45 years or older were recruited from all 48 contiguous US states and the District of Columbia between 2003 and 2007.

Medicare is a US federal health insurance program administered by the Centers for Medicare and Medicaid Services that covers persons who are aged 65 years or older, have disabilities, or have end-stage renal disease. Beneficiaries may receive coverage through fee-for-service or managed-care organizations (i.e., Medicare Part C, also known as Medicare Advantage). Medicare data used for the current analyses were derived from the beneficiary enrollment file and fee-for-service claims (i.e., Medicare Parts A (inpatient services) and B (outpatient services)) beginning in 1999. REGARDS data were linked to the beneficiary enrollment file by matching on Social Security number, sex, and birthdate. To allow for reporting or coding errors, we required 2 out of the 3 components of birthdate to be identical: 1) year and month, 2) year and day, or 3) month and day, with at most 1 year's difference.

The REGARDS study protocol was approved by the institutional review boards at the participating study centers, and all participants provided written informed consent. The Centers for Medicare and Medicaid Services and the Institutional Review Board at the University of Alabama at Birmingham (Birmingham, Alabama) approved the current analysis.

REGARDS data collection

Baseline data in the REGARDS Study were collected through a telephone interview followed by an in-home study visit. Self-reported information collected at baseline included age, race, sex, current cigarette smoking, and current medication use. Participants were defined as having reported CHD if, during the telephone interview, they responded "yes" to questions asking whether "a doctor or other health professional ever told you that you had a myocardial infarction or heart attack," "[you have] ever had coronary bypass surgery, such as graft, CABG [coronary artery bypass graft] or a bypass surgery on the arteries of your heart," or "[you have] ever had an angioplasty or stenting of a coronary artery with or without placing a coil in the artery to keep it open."

During the in-home study visit, a trained health professional measured blood pressure, height, and weight and conducted

venipuncture following a standardized protocol. Blood samples were sent to a central repository at the University of Vermont (Burlington, Vermont). Dyslipidemia was defined as having a total cholesterol concentration of ≥ 240 mg/dL, a low-density lipoprotein cholesterol concentration of ≥ 160 mg/dL, a high-density lipoprotein cholesterol concentration of ≤ 40 mg/dL, or use of lipid-lowering medication. Hypertension was defined as systolic blood pressure ≥ 140 mm Hg, diastolic blood pressure ≥ 90 mm Hg, or use of antihypertensive medication. Diabetes was defined as a fasting blood glucose concentration of ≥ 126 mg/dL, a nonfasting glucose concentration of ≥ 200 mg/dL, or use of antidiabetes medication.

CHD events following the baseline REGARDS Study visit were identified and adjudicated in an ancillary study using medical records and trained physicians (13, 14). Data on CHD outcomes were available through December 31, 2009.

Construction of the 6 cohorts without Medicare claims

We first created 3 samples of Medicare beneficiaries using the different lengths of required Medicare coverage (6 months, 1 year, or 2 years) prior to the REGARDS Study interview (Figure 1). We excluded participants who reported that they did not know their CHD status ($n = 343$). To have the sample better represent the general population and populations commonly used in Medicare analyses (15, 16), we required participants to be at least 65 years of age at the beginning of their required look-back periods (age 65.5, 66, or 67 years at the time of the REGARDS telephone interview for the 6-month, 1-year, and 2-year look-back periods, respectively). Because claims were not complete for persons with Medicare Part C coverage, we restricted the analyses to participants with Medicare Parts A and B coverage and excluded participants who had Part C coverage for the month in which their telephone interview occurred and during the 6-month, 1-year, or 2-year look-back period. The 6-month, 1-year, and 2-year Medicare coverage requirements resulted in samples consisting of 8,937, 8,483, and 7,556 participants, respectively.

Two approaches were used to identify CHD claims: a fixed-window look-back that included only claims that fell within the required Medicare coverage period (i.e., 6 months, 1 year, or 2 years) and an all-available look-back which included claims filed during the entire period in which a beneficiary had continuous Medicare Parts A and B coverage but not Part C coverage prior to the REGARDS Study interview (Table 1). We refer to the 6 cohorts using the length of required coverage (6 months, 1 year, or 2 years) and the look-back approach employed (fixed-window or all-available). For example, the "6-month all-available look-back cohort" required participants to have 6 months of continuous Medicare Parts A and B coverage but not Part C coverage prior to the REGARDS telephone interview and identified a history of CHD using CHD claims filed at any time prior to the REGARDS telephone interview.

A claims-based history of CHD was first identified using a "base algorithm" that included *International Classification of Diseases, Ninth Revision (ICD-9)*, codes for myocardial infarction (≥ 1 inpatient claim with an ICD-9 diagnosis of 410.x or 412.x or ≥ 2 physician evaluation and management (E/M) outpatient claims filed ≥ 7 days apart with an ICD-9 diagnosis code of 412.x) or ICD-9 and Current Procedural Terminology codes for coronary artery revascularization (≥ 1 inpatient or

Table 1. Cohort Designs Used to Obtain Study Populations Without Medicare Claims for Coronary Heart Disease, REGARDS Study, 1999–2007

| Look-Back Period | Amount of Medicare Coverage Required at Study Entry ^a | No. of Participants With Required Coverage | Duration of Time Used to Determine the Presence of CHD Claims ^{a,b} | No. of Participants Without CHD Claims ^b |
|------------------------|--|--|--|---|
| 6-month fixed-window | ≥6 months before index date | 8,937 | 6 months before index date | 8,603 |
| ≥6-month all-available | ≥6 months before index date | 8,937 | Maximum continuous coverage before index date | 7,682 |
| 1-year fixed-window | ≥1 year before index date | 8,483 | 1 year before index date | 7,908 |
| ≥1-year all-available | ≥1 year before index date | 8,483 | Maximum continuous coverage before index date | 7,255 |
| 2-year fixed-window | ≥2 years before index date | 7,556 | 2 years before index date | 6,756 |
| ≥2-year all-available | ≥2 years before index date | 7,556 | Maximum continuous coverage before index date | 6,411 |

Abbreviations: CHD, coronary heart disease; ICD-9, *International Classification of Diseases, Ninth Revision*; REGARDS, Reasons for Geographic and Racial Differences in Stroke.

^a For the current study, the REGARDS in-home visit was used as the index date.

^b CHD claims included those containing ICD-9 codes for myocardial infarction or ICD-9 and Current Procedural Terminology codes for coronary revascularization (see text for details).

outpatient claim containing ICD-9 procedure code 00.66 or codes 36.01–36.19 or Current Procedural Terminology codes 92980–92996 or 33510–33536, or ≥1 inpatient or physician E/M outpatient claim containing ICD-9 diagnosis code V45.81 or V45.82). Next, we expanded the claims-based definition of CHD history by adding, individually and in combination, a priori–selected claims to the base algorithm. We examined adding inpatient or physician E/M outpatient claims containing ICD-9 diagnosis codes for “other acute and subacute forms of ischemic heart disease” (ICD-9 code 411.x), “angina pectoris” (ICD-9 code 413.x), and “other forms of chronic ischemic heart disease” (ICD-9 code 414.x) and claims filed by a cardiologist (Medicare specialty code 06). We also examined adding physician E/M outpatient claims versus using only inpatient claims containing ICD-9 codes for acute myocardial infarction (ICD-9 code 410.x), as well as requiring only ≥1 E/M outpatient claim (vs. ≥2) with ICD-9 codes 410.x–414.x. Lastly, we examined adding claims obtained from a data-mining procedure used for automatic variable selection of administrative claims (17).

Statistical analysis

For each of the 6 cohorts without CHD claims in look-back periods using the base algorithm, we calculated participant characteristics overall and by self-reported CHD. We calculated the percentage of participants in each cohort with self-reported CHD after excluding those with CHD claims in Medicare using the base algorithm and after adding a priori–selected claims to the base algorithm. The algorithm that minimized the percentage of participants with self-reported CHD was designated the “expanded” algorithm. We then determined whether the addition of variables selected through data-mining resulted in a lower percentage of participants with self-reported CHD. To examine the impact of refining the claims algorithm to remove people with self-reported CHD, we calculated the incidence of CHD events for each of the 6 cohorts. We also calculated hazard ratios and 95% confidence intervals for associations between participant characteristics and CHD for these cohorts.

For the 6-month all-available look-back cohort, we evaluated the degree to which the maximum duration of the all-available look-back affected the percentage of persons with self-reported CHD by truncating the maximum duration of look-back at 1, 2, and 3 years. We then examined how the sample sizes of the potential CHD-free analytical samples were reduced due to using longer coverage requirements, all-available versus fixed-window look-back periods, and the expanded algorithm. Specifically, we calculated the percentage of participants excluded from the 6-month fixed-window look-back cohort with the base algorithm, when using the other 5 cohorts and the expanded algorithm. We categorized participants by the reason they were excluded.

In secondary analyses, we calculated the characteristics of participants we were unable to link to Medicare or who did not have 6 months or 1 year of continuous Medicare coverage. These groups were compared with participants without CHD claims in the expanded algorithm using the 6-month all-available look-back period and the 1-year fixed-window look-back period. Next, we calculated the type of self-reported CHD and other characteristics of participants with and without CHD claims, separately. We also calculated the types of CHD claims present and other characteristics among participants with CHD claims in the expanded algorithm using a 6-month all-available look-back period or a 1-year fixed-window look-back period, separately according to the presence or absence of self-reported CHD. We then calculated the percentage of participants with CHD claims who reported CHD by look-back duration, type of look-back (fixed vs. all-available), and type of algorithm (base vs. expanded). We also calculated the percentage of participants with self-reported CHD who had CHD claims. Analyses were conducted using SAS 9.3 (SAS Institute, Inc., Cary, North Carolina).

RESULTS

In each of the 6 assembled cohorts without CHD claims using the base algorithm, participants who reported having

Table 2. Characteristics of Participants Without Medicare Claims^a for Coronary Heart Disease According to 6 Different Look-Back Periods, REGARDS Study, 1999–2007

| Look-Back Period and Characteristic ^b | Overall | | Self-Reported CHD ^c | | | |
|--|------------|------|--------------------------------|------|------------|------|
| | No. | % | No | | Yes | |
| | | | No. | % | No. | % |
| 6-month fixed-window | | | | | | |
| Total | 8,603 | | 7,071 | | 1,532 | |
| Age, years ^d | 73.0 (5.7) | | 72.8 (5.7) | | 73.8 (5.9) | |
| Black race | 2,781 | 32.3 | 2,385 | 33.7 | 396 | 25.8 |
| Male sex | 4,180 | 48.6 | 3,129 | 44.3 | 1,051 | 68.6 |
| Current smoker | 796 | 9.3 | 638 | 9.1 | 158 | 10.3 |
| Dyslipidemia | 5,180 | 62.3 | 3,989 | 58.6 | 1,191 | 79.4 |
| Hypertension | 5,581 | 65.0 | 4,479 | 63.5 | 1,102 | 72.2 |
| Diabetes | 1,876 | 22.6 | 1,387 | 20.4 | 489 | 32.9 |
| ≥6-month all-available | | | | | | |
| Total | 7,682 | | 6,990 | | 692 | |
| Age, years | 72.8 (5.7) | | 72.8 (5.7) | | 73.1 (6.0) | |
| Black race | 2,558 | 33.3 | 2,352 | 33.6 | 206 | 29.8 |
| Male sex | 3,546 | 46.2 | 3,085 | 44.1 | 461 | 66.6 |
| Current smoker | 711 | 9.3 | 627 | 9.0 | 84 | 12.2 |
| Dyslipidemia | 4,452 | 60.1 | 3,945 | 58.6 | 507 | 75.6 |
| Hypertension | 4,929 | 64.3 | 4,422 | 63.4 | 507 | 73.6 |
| Diabetes | 1,567 | 21.2 | 1,362 | 20.2 | 205 | 30.8 |
| 1-year fixed-window | | | | | | |
| Total | 7,908 | | 6,669 | | 1,239 | |
| Age, years | 73.3 (5.6) | | 73.2 (5.6) | | 74.1 (5.8) | |
| Black race | 2,555 | 32.3 | 2,233 | 33.5 | 322 | 26.0 |
| Male sex | 3,791 | 47.9 | 2,944 | 44.1 | 847 | 68.4 |
| Current smoker | 720 | 9.1 | 591 | 8.9 | 129 | 10.5 |
| Dyslipidemia | 4,711 | 61.7 | 3,764 | 58.5 | 947 | 78.3 |
| Hypertension | 5,118 | 64.9 | 4,220 | 63.4 | 898 | 72.8 |
| Diabetes | 1,686 | 22.1 | 1,301 | 20.3 | 385 | 32.1 |

Table continues

CHD were older, were less likely to be black, and were more likely to be male, to be current smokers, and to have dyslipidemia, hypertension, or diabetes than their counterparts without self-reported CHD (Table 2).

Percentage with self-reported CHD among 6 cohorts without CHD claims

Using the base algorithm, the percentages of persons without CHD claims who reported having CHD were 17.8%, 15.7%, and 12.9% in the 6-month, 1-year, and 2-year fixed-window look-back cohorts, respectively (Tables 3–5). The percentage with self-reported CHD was lower when using an all-available look-back period. For both the fixed-window and all-available look-back periods, expanding the base algorithm by including ≥1 inpatient or physician E/M claim for any CHD-related diagnoses (ICD-9 codes 410–414) resulted in the lowest percentage of participants who reported

CHD. This algorithm was designated the “expanded algorithm.” The percentage of participants who reported having CHD was not lowered when we added variables selected through data-mining to the expanded algorithm (data not shown).

CHD incidence and associations with risk factors

The incidence of CHD was lower in cohorts using the expanded algorithm and all-available look-back periods (see Web Table 1, available at <http://aje.oxfordjournals.org/>). When using a fixed-window look-back, CHD incidence was lower when requiring longer look-back periods. For example, the CHD incidence rate was 9.9 per 1,000 person-years (95% confidence interval: 9.0, 10.8) using the base algorithm with a 6-month fixed-window look-back period and 6.6 per 1,000 person-years (95% confidence interval: 5.7, 7.4) using the expanded algorithm with a 6-month all-available look-back

Table 2. Continued

| Look-Back Period and Characteristic ^b | Overall | | Self-Reported CHD ^c | | | |
|--|------------|------|--------------------------------|------|------------|------|
| | No. | % | No | | Yes | |
| | | | No. | % | No. | % |
| 1-year all-available | | | | | | |
| Total | 7,255 | | 6,608 | | 647 | |
| Age, years | 73.2 (5.6) | | 73.2 (5.6) | | 73.5 (5.9) | |
| Black race | 2,402 | 33.1 | 2,207 | 33.4 | 195 | 30.1 |
| Male sex | 3,349 | 46.2 | 2,915 | 44.1 | 434 | 67.1 |
| Current smoker | 660 | 9.1 | 582 | 8.8 | 78 | 12.1 |
| Dyslipidemia | 4,202 | 60.0 | 3,731 | 58.6 | 471 | 75.2 |
| Hypertension | 4,659 | 64.4 | 4,178 | 63.4 | 481 | 74.6 |
| Diabetes | 1,470 | 21.0 | 1,280 | 20.1 | 190 | 30.5 |
| 2-year fixed-window | | | | | | |
| Total | 6,756 | | 5,886 | | 870 | |
| Age, years | 74.0 (5.4) | | 73.9 (5.4) | | 74.7 (5.6) | |
| Black race | 2,175 | 32.2 | 1,941 | 33.0 | 234 | 26.9 |
| Male sex | 3,189 | 47.2 | 2,590 | 44.0 | 599 | 68.9 |
| Current smoker | 586 | 8.7 | 499 | 8.5 | 87 | 10.0 |
| Dyslipidemia | 3,965 | 60.9 | 3,303 | 58.3 | 662 | 78.3 |
| Hypertension | 4,387 | 65.1 | 3,750 | 63.9 | 637 | 73.6 |
| Diabetes | 1,404 | 21.6 | 1,142 | 20.2 | 262 | 31.2 |
| 2-year all-available | | | | | | |
| Total | 6,411 | | 5,843 | | 568 | |
| Age, years | 73.9 (5.4) | | 73.9 (5.4) | | 74.2 (5.7) | |
| Black race | 2,092 | 32.6 | 1,923 | 32.9 | 169 | 29.8 |
| Male sex | 2,951 | 46.0 | 2,568 | 44.0 | 383 | 67.4 |
| Current smoker | 550 | 8.6 | 491 | 8.4 | 59 | 10.4 |
| Dyslipidemia | 3,694 | 59.8 | 3,277 | 58.2 | 417 | 75.8 |
| Hypertension | 4,147 | 64.8 | 3,723 | 63.9 | 424 | 74.9 |
| Diabetes | 1,292 | 20.9 | 1,126 | 20.0 | 166 | 30.2 |

Abbreviations: CHD, coronary heart disease; ICD-9, *International Classification of Diseases, Ninth Revision*; REGARDS, Reasons for Geographic and Racial Differences in Stroke.

^a CHD claims included those containing ICD-9 codes for myocardial infarction or ICD-9 and Current Procedural Terminology codes for coronary revascularization (see text for details).

^b Data on the characteristics shown were obtained from the REGARDS Study telephone interview and in-home visit.

^c Self-reported CHD was defined as reporting a previous myocardial infarction or coronary revascularization.

^d Data are presented as mean (standard deviation).

period. Hazard ratios for the associations of demographic and risk factors with CHD outcomes for the 6 cohorts are provided in Web Table 2.

Effect of shorter maximum look-back periods

Participants without CHD claims in the 6-month all-available look-back cohort using the expanded algorithm had a median look-back period of 4.0 years (25th–75th percentile range, 2.8–5.0 years; maximum, 7.8 years) (Figure 2). When the maximum duration of look-back was truncated at 1, 2, or 3 years, the percentages of participants who reported having CHD were 6.8%, 4.9%, and 4.2%, respectively.

Participants excluded due to coverage requirements and CHD claims

Figure 3 shows the percentage of participants without CHD claims using the base algorithm in the 6-month fixed-window look-back cohort who were excluded for 1) not meeting 1- or 2-year coverage requirements, 2) having CHD claims using the 1- or 2-year fixed-window look-back period, 3) having CHD claims using the all-available look-back period, or 4) having CHD claims using the expanded algorithm. The majority of participants who were excluded from this cohort for having additional CHD claims when 1 or 2 years of Medicare coverage were required did not have self-reported CHD. In contrast,

Table 3. Number of Participants Without Coronary Heart Disease (CHD) Claims in Medicare and Percentage Reporting CHD in 6-Month Look-Back Periods, by CHD History Claims Algorithm, REGARDS Study, 1999–2007

| CHD History Claims Algorithm | 6-Month Fixed-Window Look-Back Period | | ≥6-Month All-Available Look-Back Period | |
|---|---------------------------------------|------------------------------|---|------------------------------|
| | No. Without CHD Claims | % Reporting CHD ^a | No. Without CHD Claims | % Reporting CHD ^a |
| Base algorithm ^b | 8,603 | 17.8 | 7,682 | 9.0 |
| Base algorithm or ≥1 inpatient or ≥2 E/M outpatient claims with ICD-9 code 410 | 8,602 | 17.8 | 7,678 | 9.0 |
| Base algorithm or ≥1 inpatient or ≥2 E/M outpatient claims with ICD-9 code 410 or 411 | 8,588 | 17.8 | 7,609 | 8.8 |
| Base algorithm or ≥1 inpatient or ≥2 E/M outpatient claims with ICD-9 code 410, 411, or 413 | 8,556 | 17.6 | 7,431 | 8.2 |
| Base algorithm or ≥1 inpatient or ≥2 E/M outpatient claims with ICD-9 code 410, 411, 413, or 414 | 8,102 | 14.4 | 6,610 | 4.4 |
| Base algorithm or ≥1 inpatient or ≥1 E/M outpatient claim with ICD-9 code 412 | 8,565 | 17.5 | 7,622 | 8.6 |
| Base algorithm or ≥1 inpatient or ≥1 E/M outpatient claim with ICD-9 code 410 or 412 | 8,558 | 17.5 | 7,600 | 8.5 |
| Base algorithm or ≥1 inpatient or ≥1 E/M outpatient claim with ICD-9 code 410, 411, or 412 | 8,512 | 17.3 | 7,447 | 8.1 |
| Base algorithm or ≥1 inpatient or ≥1 E/M outpatient claim with ICD-9 code 410, 411, 412, or 413 | 8,383 | 16.7 | 7,118 | 7.4 |
| Base algorithm or ≥1 inpatient or ≥1 E/M outpatient claim with ICD-9 code 410, 411, 412, 413, or 414 (expanded algorithm) | 7,433 | 9.9 | 6,119 | 3.6 |
| Base algorithm or cardiologist visit | 7,269 | 12.8 | 5,715 | 5.7 |
| Base algorithm or ≥1 inpatient or ≥2 E/M outpatient claims with ICD-9 code 410, 411, 413, or 414 or cardiologist visit | 7,106 | 11.7 | 5,358 | 3.9 |

Abbreviations: CHD, coronary heart disease; E/M, evaluation and management; ICD-9, *International Classification of Diseases, Ninth Revision*; REGARDS, Reasons for Geographic and Racial Differences in Stroke.

^a Self-reported CHD was defined as reporting a previous myocardial infarction or coronary revascularization.

^b The base algorithm identified persons with CHD claims containing ICD-9 codes for myocardial infarction or ICD-9 and Current Procedural Terminology codes for coronary revascularization (see text for details).

the majority of participants excluded from this cohort for having additional CHD claims using the all-available look-back approach (vs. the fixed-window look-back approach) or the expanded algorithm (vs. the base algorithm) had self-reported CHD.

Secondary analyses

Participants who were ineligible for analyses due to unsuccessful Medicare linkage or lack of continuous Medicare coverage prior to the REGARDS telephone interview were similar to participants without CHD claims (Web Table 3). Among participants with self-reported CHD, those with CHD claims were more likely than those without CHD claims to report having a prior coronary artery bypass graft or percutaneous coronary intervention, but similar percentages reported a prior myocardial infarction (Web Table 4). Among participants who had CHD claims, those with self-reported CHD were consistently more likely than their counterparts without self-reported CHD to have each type of CHD claim, except for claims with ICD-9 code 413.0 (angina pectoris), and to have a CHD claim more proximal to their REGARDS Study interview (Web Table 5). For all 6 cohorts, over 90% of participants with CHD claims in the base algorithm reported a history of CHD (Web Table 6). However, 50%–75% of participants with CHD claims in the

expanded algorithm reported a history of CHD. Using the base algorithm and a fixed-window look-back, fewer than 50% of participants with self-reported CHD had CHD claims (Web Table 7). Using an all-available look-back period or an expanded algorithm increased this percentage to 60%–90%.

DISCUSSION

The current study used information from a large number of persons with self-reported medical histories and linked Medicare claims to evaluate approaches for identifying a cohort of persons who were free of CHD. The percentage of persons with self-reported CHD varied from 18% to less than 4%, depending on the claims included in a history-of-CHD algorithm, the length of required Medicare coverage prior to study entry, and the use of all available claims or claims filed during a fixed time window prior to study entry. Our analysis indicates that using all available claims, as opposed to the more commonly used fixed-window look-back approach, minimized the percentage of participants with CHD claims who reported CHD. In addition, we found that when an all-available look-back period was used, in this sample with a median of 4 years of maximum look-back, increasing the length of required continuous coverage beyond 6 months resulted in a smaller sample size

Table 4. Number of Participants Without Coronary Heart Disease (CHD) Claims in Medicare and Percentage Reporting CHD in 1-Year Look-Back Periods, by CHD Claims Algorithm, REGARDS Study, 1999–2007

| CHD History Claims Algorithm | 1-Year Fixed-Window Look-Back Period | | ≥1-Year All-Available Look-Back Period | |
|---|--------------------------------------|------------------------------|--|------------------------------|
| | No. Without CHD Claims | % Reporting CHD ^a | No. Without CHD Claims | % Reporting CHD ^a |
| Base algorithm ^b | 7,908 | 15.7 | 7,255 | 8.9 |
| Base algorithm or ≥1 inpatient or ≥2 E/M outpatient claims with ICD-9 code 410 | 7,905 | 15.6 | 7,251 | 8.9 |
| Base algorithm or ≥1 inpatient or ≥2 E/M outpatient claims with ICD-9 code 410 or 411 | 7,876 | 15.5 | 7,183 | 8.7 |
| Base algorithm or ≥1 inpatient or ≥2 E/M outpatient claims with ICD-9 code 410, 411, or 413 | 7,808 | 15.2 | 7,008 | 8.1 |
| Base algorithm or ≥1 inpatient or ≥2 E/M outpatient claims with ICD-9 code 410, 411, 413, or 414 | 7,090 | 9.7 | 6,203 | 4.2 |
| Base algorithm or ≥1 inpatient or ≥1 E/M outpatient claim with ICD-9 code 412 | 7,856 | 15.2 | 7,198 | 8.5 |
| Base algorithm or ≥1 inpatient or ≥1 E/M outpatient claim with ICD-9 code 410 or 412 | 7,845 | 15.2 | 7,176 | 8.4 |
| Base algorithm or ≥1 inpatient or ≥1 E/M outpatient claim with ICD-9 code 410, 411, or 412 | 7,778 | 14.8 | 7,026 | 8.0 |
| Base algorithm or ≥1 inpatient or ≥1 E/M outpatient claim with ICD-9 code 410, 411, 412, or 413 | 7,614 | 14.3 | 6,709 | 7.3 |
| Base algorithm or ≥1 inpatient or ≥1 E/M outpatient claim with ICD-9 code 410, 411, 412, 413, or 414 (expanded algorithm) | 6,593 | 6.8 | 5,740 | 3.4 |
| Base algorithm or cardiologist visit | 6,346 | 9.7 | 5,337 | 5.5 |
| Base algorithm or ≥1 inpatient or ≥2 E/M outpatient claims with ICD-9 code 410, 411, 413, or 414 or cardiologist visit | 6,091 | 7.8 | 4,992 | 3.7 |

Abbreviations: CHD, coronary heart disease; E/M, evaluation and management; ICD-9, *International Classification of Diseases, Ninth Revision*; REGARDS, Reasons for Geographic and Racial Differences in Stroke.

^a Self-reported CHD was defined as reporting a previous myocardial infarction or coronary revascularization.

^b The base algorithm identified persons with CHD claims containing ICD-9 codes for myocardial infarction or ICD-9 and Current Procedural Terminology codes for coronary revascularization (see text for details).

without reducing the percentage of participants with self-reported CHD.

Many claims-based analyses have used a fixed-window look-back period to identify comorbid conditions, such as CHD, ignoring older claims (4, 9, 11). As we demonstrate in the current study, using all available claims for the look-back period can identify a substantial percentage of beneficiaries who report having CHD but do not have CHD claims during a fixed window of Medicare coverage. The benefit of using an all-available-claims look-back approach depends on the lengths of continuous coverage in a data source. We found that even when we shortened the maximum coverage lengths, use of an all-available look-back obtained a sample with a similar percentage of participants who reported having CHD and excluded fewer participants, in comparison with fixed-window look-back periods. However, if an exposure of interest is related to beneficiaries' lengths of insurance coverage (e.g., age in Medicare), then participants may be excluded differentially. While this has not been previously examined, a simulation found that including a confounder obtained from an all-available look-back period rather than a fixed-window look-back period resulted in less bias in the exposure-outcome relationship (18).

Few studies have examined the impact of different lengths of continuous coverage required for a look-back period. In a

study of Canadian Ontario Health Insurance Plan beneficiaries, Tu et al. (19) found that extending the fixed-window look-back period from 1 year to 3 years reduced the percentage of persons who reported having CHD by 0.1%. However, beneficiaries were required to have at least 2 years of coverage for all analyses in this study, so the effect of different coverage requirements may have been minimized. We found that when we used an all-available look-back approach, increasing the length of required continuous coverage excluded a large number of participants but did not lower the percentage of participants reporting CHD. All-available look-back periods are primarily used when there is a fixed time point for the start of follow-up, such as when baseline data collection (e.g., the REGARDS telephone interview in the current study) or a particular diagnosis or event defines the study-entry time point. If there is no baseline data collection date or other fixed study-entry time point, a fixed-window look-back is often used. For these reasons, if a fixed-window look-back is used, our results suggest that requiring longer periods of continuous coverage will be useful for reducing the percentage of persons with CHD.

Few data are available on the validity of claims-based algorithms for identifying a population free of CHD at study entry. In the study of Ontario Health Insurance Plan beneficiaries, Tu et al. found that the percentage of beneficiaries who

Table 5. Number of Participants Without Coronary Heart Disease (CHD) Claims in Medicare and Percentage Reporting CHD in 2-Year Look-Back Periods, by CHD Claims Algorithm, REGARDS Study, 1999–2007

| CHD History Claims Algorithm | 2-Year Fixed-Window Look-Back Period | | ≥2-Year All-Available Look-Back Period | |
|---|--------------------------------------|------------------------------|--|------------------------------|
| | No. Without CHD Claims | % Reporting CHD ^a | No. Without CHD Claims | % Reporting CHD ^a |
| Base algorithm ^b | 6,756 | 12.9 | 6,411 | 8.9 |
| Base algorithm or ≥1 inpatient or ≥2 E/M outpatient claims with ICD-9 code 410 | 6,751 | 12.8 | 6,407 | 8.8 |
| Base algorithm or ≥1 inpatient or ≥2 E/M outpatient claims with ICD-9 code 410 or 411 | 6,714 | 12.7 | 6,343 | 8.7 |
| Base algorithm or ≥1 inpatient or ≥2 E/M outpatient claims with ICD-9 code 410, 411, or 413 | 6,618 | 12.3 | 6,177 | 8.0 |
| Base algorithm or ≥1 inpatient or ≥2 E/M outpatient claims with ICD-9 code 410, 411, 413, or 414 | 5,854 | 6.4 | 5,438 | 4.1 |
| Base algorithm or ≥1 inpatient or ≥1 E/M outpatient claim with ICD-9 code 412 | 6,700 | 12.4 | 6,357 | 8.4 |
| Base algorithm or ≥1 inpatient or ≥1 E/M outpatient claim with ICD-9 code 410 or 412 | 6,685 | 12.3 | 6,335 | 8.3 |
| Base algorithm or ≥1 inpatient or ≥1 E/M outpatient claim with ICD-9 code 410, 411, or 412 | 6,596 | 12.0 | 6,193 | 7.9 |
| Base algorithm or ≥1 inpatient or ≥1 E/M outpatient claim with ICD-9 code 410, 411, 412, or 413 | 6,391 | 11.3 | 5,893 | 7.2 |
| Base algorithm or ≥1 inpatient or ≥1 E/M outpatient claim with ICD-9 code 410, 411, 412, 413, or 414 (expanded algorithm) | 5,446 | 4.9 | 4,999 | 3.4 |
| Base algorithm or cardiologist visit | 5,199 | 7.7 | 4,646 | 5.6 |
| Base algorithm or ≥1 inpatient or ≥2 E/M outpatient claims with ICD-9 code 410, 411, 413, or 414 or cardiologist visit | 4,904 | 5.5 | 4,325 | 3.6 |

Abbreviations: CHD, coronary heart disease; E/M, evaluation and management; ICD-9, *International Classification of Diseases, Ninth Revision*; REGARDS, Reasons for Geographic and Racial Differences in Stroke.

^a Self-reported CHD was defined as reporting a previous myocardial infarction or coronary revascularization.

^b The base algorithm identified persons with CHD claims containing ICD-9 codes for myocardial infarction or ICD-9 and Current Procedural Terminology codes for coronary revascularization (see text for details).

reported having CHD changed by approximately 1% when the history-of-CHD algorithm was varied (19). Other investigators have calculated the kappa statistic to examine agreement between CHD algorithms and a gold standard (e.g., self-report or record review) but did not specifically look at the percentage of beneficiaries without CHD claims who were CHD-free by the gold standard (i.e., the negative predictive value) (9, 10). Additionally, the results of these studies vary not only because of differing CHD claims algorithms but also because of different data sources, study populations (e.g., a general population aged ≥20 years in a Canadian province (19) vs. a general US population sample of black and white Medicare enrollees aged ≥65 years (current study)), calendar year time periods, and continuous coverage requirements. In the current study, the percentage of participants who reported having CHD was reduced by expanding the CHD algorithm from one that included inpatient and outpatient claims for myocardial infarction and revascularization to an algorithm that also included claims for angina pectoris and other ischemic heart disease codes. Although the expanded algorithm excluded many more participants and resulted in a smaller cohort, most of the additional participants excluded reported having CHD.

While the primary purpose of this study was to identify a population free of CHD using Medicare claims, we also

examined the prevalence of self-reported CHD among participants with CHD claims. Regardless of the duration of the look-back period or the use of a fixed-window or all-available look-back, over 90% of participants with CHD claims in the base algorithm reported a history of CHD. This algorithm may be useful for identifying Medicare beneficiaries who are likely to have a history of CHD.

There are many strengths associated with using the REGARDS Study to examine approaches for identifying a CHD-free population in claims data. REGARDS participants were community-dwelling and resided in all 48 contiguous states and the District of Columbia. The population included in the current analysis was similar to the overall REGARDS cohort. The large sample size of REGARDS participants linked with Medicare claims allowed for stable prevalence estimates. However, this study also had limitations. We focused on Medicare beneficiaries aged ≥65 years, and the generalizability of our algorithms and approaches for defining look-back periods to other sources of claims is not known. Commercial and Medicaid insurance databases usually consist of younger populations, who would be likely to have a lower percentage of persons with CHD. In addition, patients in these databases often have shorter durations of continuous coverage. Another factor that may affect generalizability is

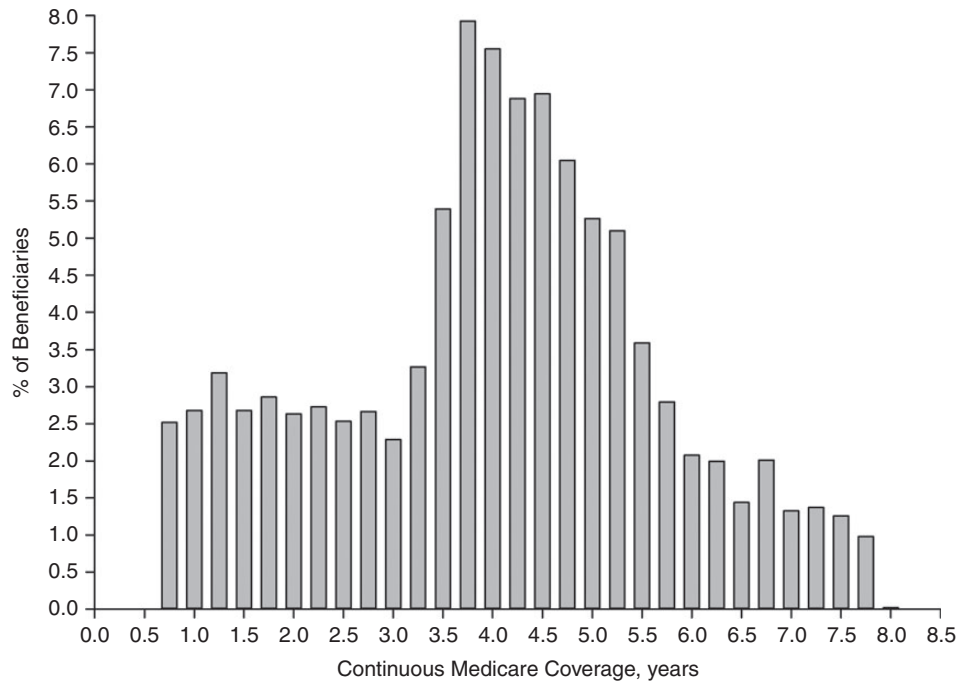


Figure 2. Distribution of participants without Medicare claims for coronary heart disease using the 6-month all-available look-back period and the expanded algorithm, by length of continuous Medicare coverage prior to the study telephone interview, Reasons for Geographic and Racial Differences in Stroke (REGARDS) Study, 1999–2007.

that approximately 20% of REGARDS participants with Medicare coverage were enrolled in Medicare Part C during the study period. In the general Medicare population, the

proportion of beneficiaries with Part C coverage is higher in western and upper midwestern states, and the proportion is increasing with time (20). Use of chart review to identify

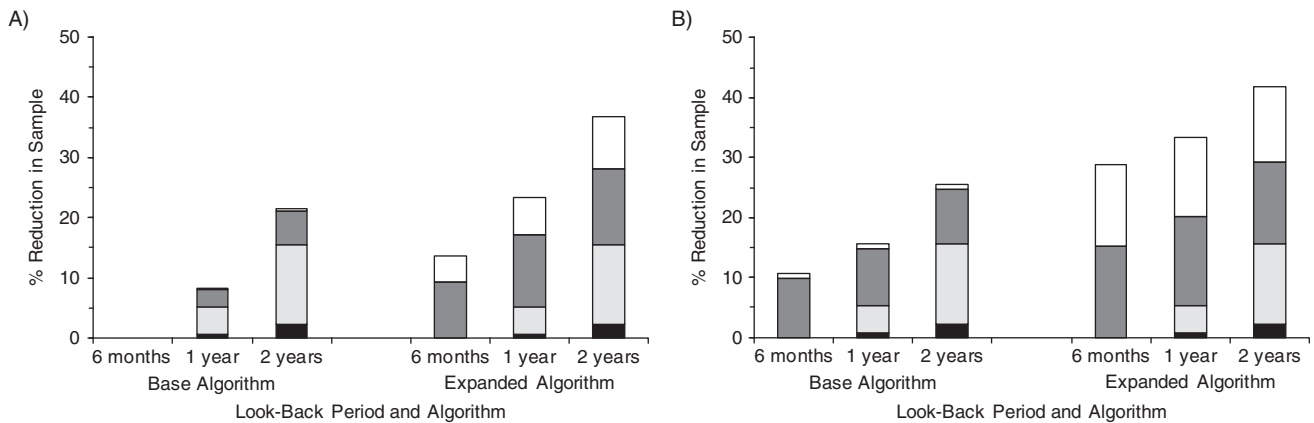


Figure 3. Percentage of the study sample excluded from the 6-month fixed-window look-back cohort, by duration of look-back period, look-back approach (fixed-window or all-available), and type of coronary heart disease (CHD) history claims algorithm, Reasons for Geographic and Racial Differences in Stroke (REGARDS) Study, 1999–2007. Black bar segments represent the percentage of participants who were excluded for not meeting the coverage requirement and who reported having CHD. Light gray bar segments represent the percentage of participants who were excluded for not meeting the coverage requirement and who reported no CHD. Dark gray bar segments represent the percentage of participants who were excluded for having CHD Medicare claims and who reported CHD. White bar segments represent the percentage of participants who were excluded for having CHD Medicare claims and who reported no CHD. The base algorithm identified persons with CHD claims containing *International Classification of Diseases, Ninth Revision* (ICD-9), codes for myocardial infarction or ICD-9 and Current Procedural Terminology codes for coronary revascularization (see text for details). The expanded algorithm identified persons with CHD claims in the base algorithm or ≥ 1 inpatient or physician evaluation and management outpatient claim with ICD-9 codes 410–414.

CHD at baseline was not feasible. Therefore, we relied on self-reports to define a history of CHD in REGARDS. The accuracy of self-reporting may be subject to recall bias and can vary by population and the amount of time between the event and the survey (21, 22). However, a high percentage of myocardial infarction and coronary artery bypass graft procedures are accurately recalled (22, 23). In addition, self-reporting can be useful in determining baseline disease status and is commonly used for this purpose in large clinical trials and cohort studies (5, 24–26). We confirmed in our study that self-reported CHD among participants without CHD claims was associated with traditional CHD risk factors and treatments and with higher CHD incidence.

In conclusion, the current study provides guidance on how to obtain an analytical sample free of CHD in administrative data, minimizing persons with self-reported CHD and maximizing the sample retained. For studies with a defined study entry date, such as a baseline data collection point or disease diagnosis, the choice of the look-back period depends on the duration of continuous coverage available prior to study entry. When the median duration of coverage prior to study entry is long enough, 6 months of continuous coverage and use of all available claims prior to study entry can help one obtain an analytical sample that has a low percentage of beneficiaries with CHD. However, if a fixed-window look-back period is used, our results suggest that requiring at least 1 year of continuous coverage may minimize the number of beneficiaries misclassified as CHD-free.

ACKNOWLEDGMENTS

Author affiliations: Departments of Epidemiology (Shia T. Kent, Hong Zhao, Emily B. Levitan, Paul Muntner) and Health Care Organization and Policy (Meredith L. Kilgore), School of Public Health, University of Alabama at Birmingham, Birmingham, Alabama; Divisions of Preventive Medicine (Monika M. Safford) and Clinical Immunology and Rheumatology (Jeffrey R. Curtis), Department of Medicine, School of Medicine, University of Alabama at Birmingham, Birmingham, Alabama; and Department of Global Drug Safety, Baxalta US, Inc., Cambridge, Massachusetts (Ryan D. Kilpatrick).

All authors contributed equally to this article.

Collection of data on myocardial infarction outcomes in the Reasons for Geographic and Racial Differences in Stroke (REGARDS) Study, REGARDS-MI, was supported by National Institutes of Health grants U01 NS041588, R01 HL080477, and K24 HL111154. Additional project support was provided by National Institutes of Health grant T32 HL00745733 and a contract for an academic/industry collaboration between the University of Alabama at Birmingham and Amgen, Inc. (Thousand Oaks, California).

We thank the other investigators and staff of the REGARDS Study for their valuable contributions. A full list of participating REGARDS investigators and institutions can be found at <http://www.regardsstudy.org>.

This work was presented at one of the American Heart Association's 2014 Scientific Sessions ("Epidemiology and

Prevention & Nutrition, Physical Activity, and Metabolism," San Francisco, California, March 18–21, 2014).

S.T.K., M.M.S., H.Z., E.B.L., J.R.C., M.L.K., and P.M. have received research support from Amgen, Inc. J.R.K. is a consultant for Amgen. M.M.S. also consults for diaDexus, Inc. (South San Francisco, California) and has received an educational grant from Medscape (Medscape, Inc., New York, New York). R.K. is a past employee of Amgen and GlaxoSmithKline (London, United Kingdom) and is a current employee of Baxalta US.

REFERENCES

- Gavriellov-Yusim N, Friger M. Use of administrative medical databases in population-based research. *J Epidemiol Community Health*. 2014;68(3):283–287.
- Jain SH, Rosenblatt M, Duke J. Is big data the new frontier for academic-industry collaboration? *JAMA*. 2014;311(21):2171–2172.
- Schneider KM, O'Donnell BE, Dean D. Prevalence of multiple chronic conditions in the United States' Medicare population. *Health Qual Life Outcomes*. 2009;7:82.
- Klabunde CN, Warren JL, Legler JM. Assessing comorbidity using claims data: an overview. *Med Care*. 2002;40(8 suppl):IV-26–IV-35.
- Kolansky DM. Acute coronary syndromes: morbidity, mortality, and pharmaco-economic burden. *Am J Manag Care*. 2009;15(2 suppl):S36–S41.
- Birman-Deych E, Waterman AD, Yan Y, et al. Accuracy of ICD-9-CM codes for identifying cardiovascular and stroke risk factors. *Med Care*. 2005;43(5):480–485.
- Duan Y, Mo J, Klein R, et al. Age-related macular degeneration is associated with incident myocardial infarction among elderly Americans. *Ophthalmology*. 2007;114(4):732–737.
- Luchsinger JA, Pablos-Méndez A, Knirsch C, et al. Relation of antibiotic use to risk of myocardial infarction in the general population. *Am J Cardiol*. 2002;89(1):18–21.
- Borzecki AM, Wong AT, Hickey EC, et al. Identifying hypertension-related comorbidities from administrative data: what's the optimal approach? *Am J Med Qual*. 2004;19(5):201–206.
- Robinson JR, Young TK, Roos LL, et al. Estimating the burden of disease. Comparing administrative data and self-reports. *Med Care*. 1997;35(9):932–947.
- Singh JA. Accuracy of Veterans Affairs databases for diagnoses of chronic diseases. *Prev Chronic Dis*. 2009;6(4):A126.
- Howard VJ, Cushman M, Pulley L, et al. The Reasons for Geographic and Racial Differences in Stroke Study: objectives and design. *Neuroepidemiology*. 2005;25(3):135–143.
- Safford MM, Brown TM, Muntner PM, et al. Association of race and sex with risk of incident acute coronary heart disease events. *JAMA*. 2012;308(17):1768–1774.
- Luepker RV, Apple FS, Christenson RH, et al. Case definitions for acute coronary heart disease in epidemiology and clinical research studies: a statement from the AHA Council on Epidemiology and Prevention; AHA Statistics Committee; World Heart Federation Council on Epidemiology and Prevention; the European Society of Cardiology Working Group on Epidemiology and Prevention; Centers for Disease Control and Prevention; and the National Heart, Lung, and Blood Institute. *Circulation*. 2003;108(20):2543–2549.

15. Krumholz HM, Chen J, Wang Y, et al. Comparing AMI mortality among hospitals in patients 65 years of age and older: evaluating methods of risk adjustment. *Circulation*. 1999; 99(23):2986–2992.
16. Setoguchi S, Solomon DH, Levin R, et al. Gender differences in the management and prognosis of myocardial infarction among patients ≥ 65 years of age. *Am J Cardiol*. 2008;101(11): 1531–1536.
17. Schneeweiss S, Rassen JA, Glynn RJ, et al. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. *Epidemiology*. 2009;20(4):512–522.
18. Brunelli SM, Gagne JJ, Huybrechts KF, et al. Estimation using all available covariate information versus a fixed look-back window for dichotomous covariates. *Pharmacoepidemiol Drug Saf*. 2013;22(5):542–550.
19. Tu K, Mitiku T, Lee DS, et al. Validation of physician billing and hospitalization data to identify patients with ischemic heart disease using data from the Electronic Medical Record Administrative data Linked Database (EMRALD). *Can J Cardiol*. 2010;26(7):e225–e228.
20. The Henry J. Kaiser Family Foundation. Medicare Advantage. <http://kff.org/medicare/fact-sheet/medicare-advantage-fact-sheet/>. Published May 1, 2014. Accessed March 19, 2015.
21. Bradburn NM, Rips LJ, Shevell SK. Answering autobiographical questions: the impact of memory and inference on surveys. *Science*. 1987;236(4798): 157–161.
22. Joshi R, Turnbull F. Validity of self-reported cardiovascular disease. *Intern Med J*. 2009;39(1):5–6.
23. Barr EL, Tonkin AM, Welborn TA, et al. Validity of self-reported cardiovascular disease events in comparison to medical record adjudication and a statewide hospital morbidity database: the AusDiab Study. *Intern Med J*. 2009;39(1): 49–53.
24. Okura Y, Urban LH, Mahoney DW, et al. Agreement between self-report questionnaires and medical record data was substantial for diabetes, hypertension, myocardial infarction and stroke but not for heart failure. *J Clin Epidemiol*. 2004; 57(10):1096–1103.
25. Rexrode KM, Lee IM, Cook NR, et al. Baseline characteristics of participants in the Women’s Health Study. *J Womens Health Gend Based Med*. 2000;9(1):19–27.
26. Haapanen N, Miilunpalo S, Pasanen M, et al. Agreement between questionnaire data and medical records of chronic diseases in middle-aged and elderly Finnish men and women. *Am J Epidemiol*. 1997;145(8):762–769.