Published in final edited form as:

Pharmacoepidemiol Drug Saf. 2018 July ; 27(7): 740–750. doi:10.1002/pds.4421.

Development of algorithms for identifying fatal cardiovascular disease in Medicare claims

Fenglong Xie^{1,2}, Lisandro D. Colantonio¹, Jeffrey R. Curtis^{1,2}, Meredith L. Kilgore³, Emily B. Levitan¹, Keri L. Monda⁴, Monika M. Safford⁵, Ben Taylor⁴, Mark Woodward^{6,7,8}, Paul Muntner¹

¹Department of Epidemiology, School of Public Health, University of Alabama at Birmingham, Birmingham, AL, USA

²Department of Medicine, Division of Clinical Immunology and Rheumatology, School of Medicine, University of Alabama at Birmingham, Birmingham, AL, USA

³Department of Health Care Organization and Policy, School of Public Health, University of Alabama at Birmingham, Birmingham, AL, USA

⁴The Center for Observational Research, Amgen Inc. USA.

⁵Cornell University, New York, NY, USA

⁶The George Institute for Global Health, University of Oxford, UK

⁷The George Institute for Global Health, University of New South Wales, Sydney, Australia

⁸Department of Epidemiology, Johns Hopkins University, Baltimore, MD, USA

Abstract

Background—Cause of death is often not available in administrative claims data.

Objective—To develop claims-based algorithms to identify deaths due to fatal cardiovascular disease (CVD; i.e., fatal coronary heart disease [CHD] or stroke), CHD, and stroke.

Methods—Reasons for Geographic and Racial Differences in Stroke (REGARDS) study data were linked with Medicare claims to develop the algorithms. Events adjudicated by REGARDS study investigators were used as the gold standard. Stepwise selection was used to choose predictors from Medicare data for inclusion in the algorithms. C-index, sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) were used to assess algorithm performance. Net reclassification index (NRI) was used to compare the algorithms to an approach of classifying all deaths within 28 days following hospitalization for myocardial infarction and stroke to be fatal CVD.

Corresponding Author: Paul Muntner, PhD, Department of Epidemiology, School of Public Health, University of Alabama at Birmingham, 1700 University Boulevard, Suite 450, Birmingham, AL 35294, pmuntner@uab.edu, Phone: 205-975-8077, Fax: 205-975-7058.

Conflict of interest

F. Xie and L. D. Colantonio report no disclosure. E. B. Levitan, M. Kilgore, M. M. Safford, and P. Muntner receive research grant support from Amgen. J. R. Curtis receives research grant support from Amgen not related to this project. M. M. Safford and M. Woodward receive consulting fees from Amgen. E. B. Levitan has served on advisory boards for Amgen. E. B. Levitan has received consulting fees from Novartis. P. Muntner receives honoraria from Amgen. K. L. Monda and B. Taylor are employees of Amgen.

Results—Data from 2,685 REGARDS participants with linkage to Medicare, who died between 2003 and 2013, were analyzed. The C-index for discriminating fatal CVD from other causes of death was 0.87. Using a cut-point that provided the closest observed-to-predicted number of fatal CVD events, the sensitivity was 0.64, specificity 0.90, PPV 0.65 and NPV 0.90. The algorithms resulted in positive NRIs compared with using deaths within 28 days following hospitalization for myocardial infarction and stroke. Claims-based algorithms for discriminating fatal CHD and fatal stroke performed similarly to fatal CVD.

Conclusion—The claims-based algorithms developed to discriminate fatal CVD events from other causes of death performed better than the method of using hospital discharge diagnosis codes.

Keywords

algorithm; fatal cardiovascular disease; Medicare claims

Medicare is a health insurance program for US adults 65 years of age and younger adults who are disabled or who have end-stage renal disease [1]. In 2015, 55 million US adults received health insurance through the Medicare program [2]. Medicare claims data provide the opportunity to estimate disease burden and conduct comparative effectiveness research and safety studies with much larger sample sizes than cohorts that involve primary data collection. A limitation of Medicare claims data is the lack of information on cause of death. While assessing the association of exposures with non-fatal outcomes and all-cause mortality can provide useful information, cause of death is relevant for addressing many study questions.

Cardiovascular disease (CVD), including coronary heart disease (CHD) and stroke, is the leading cause of death in the US [3]. Many analyses of Medicare claims data have investigated CVD as an outcome, but have relied on hospital discharge diagnoses and discharge disposition (i.e., deceased or alive), an approach which fails to capture out-of-hospital CVD deaths and some in-hospital CVD deaths [4–6]. In 2013, 84% of Medicare beneficiaries were 65 years of age, and mortality is high in this population [7, 8]. Not having cause of death for Medicare beneficiaries can result in the inability to accurately estimate CVD event rates, as many events are fatal [9], and in biased exposure-outcome associations due to censoring some CVD-related deaths as non-events. The goal of the current analysis was to develop claims-based algorithms to discriminate deaths that are fatal CVD, fatal CHD and fatal stroke versus deaths from other causes.

Methods

Data source

We used data from the Reasons for Geographic and Racial Differences in Stroke (REGARDS) study, a US population-based cohort that enrolled 30,239 non-Hispanic white and black adults 45 years of age from all 48 contiguous US states and Washington DC between January 1, 2003 and October 31, 2007 [10]. Data from REGARDS study participants were linked to Medicare claims using social security numbers, with linkages confirmed using birthdate and sex [11]. The REGARDS study was approved by the

institutional review board at the University of Alabama at Birmingham and all participants provided written informed consent.

Study population

We included REGARDS participants who had their study data linked to Medicare claims and died between January 1, 2003 and December 31, 2013. We excluded participants whose death date recorded in REGARDS study did not match with that recorded in the Medicare beneficiary summary file; those who died before the age of 65.5 years; and those who lacked 182 consecutive days of Medicare fee-for-service coverage (Medicare Parts A and B but not Part C) prior to their death. We required Medicare fee-for-service coverage for 182 days prior to death, referred to as the “look-back” period, to identify claims that were used in the algorithms. We excluded Medicare beneficiaries <65 years of age at the start of the look-back period because they represent a select population that qualifies for Medicare due to disability or end-stage renal disease. Finally, we excluded REGARDS study participants with a hospital discharge diagnosis code for myocardial infarction (MI) or stroke, as defined below, within 28 days prior to their death for whom medical records could not be retrieved for adjudication by the REGARDS study investigators. These participants were excluded as it could not be determined whether or not their deaths were fatal CVD events.

CVD, CHD and stroke outcomes in REGARDS

Living REGARDS study participants or their proxies are contacted every six months via telephone to confirm vital status [9, 10]. Deaths are also identified through searches in the National Death Index. When deaths are identified, interviews with proxies are conducted, and death certificates and medical records for hospitalizations in the last year of life are retrieved. Causes of death are adjudicated by two physicians independently [9]. Table 1 provides the definitions of fatal CVD, fatal CHD and fatal stroke events in the REGARDS study.

Hospital discharge diagnosis code method

One approach to define fatal CVD, CHD, and stroke in Medicare is to use discharge diagnosis codes for CVD, CHD and stroke from inpatient claims within 28 days prior to a participant’s death date (hospital discharge diagnosis method). Using this approach, we defined fatal CHD as a death preceded by an inpatient claim with a hospital discharge diagnosis code for MI (Supplemental Table 1). Fatal hemorrhagic or ischemic stroke was defined by a death preceded by an inpatient claim with a hospital discharge diagnosis code for stroke (Supplemental Table 1). Fatal CVD was defined as fatal CHD or fatal stroke. We chose to include inpatient claims within 28 days prior to the death date to match the definition of fatal MI provided by the “Case definitions for acute CHD in epidemiology and clinical research studies” [12].

Statistical methods

Candidate predictors

All potential predictors considered in the development of algorithms were defined using Medicare data. Candidate predictors, pre-specified by the authors, included demographic

data (age at death, sex, race) from the Medicare beneficiary summary file and diagnosis and procedure codes, and current procedure terminology codes from Medicare claims within 28 days and, separately, at any time prior to death (Supplemental Table 1). Also, we evaluated a second set of candidate predictors using variables defined by the Agency for Healthcare Research and Quality Clinical Classifications Software (AHRQ CCS) [13]. We calculated summary statistics for the pre-specified potential predictors among deceased REGARDS study participants included in the analysis.

Approaches for developing algorithms for composite outcomes

In the REGARDS study, fatal CVD was defined as fatal MI, CHD death or fatal stroke and fatal CHD was defined as fatal MI or CHD death [12]. We evaluated three approaches for developing algorithms for the prediction of these composite outcomes: a direct approach and two indirect approaches (Figure 1).

Direct approach

Logistic regression models, with three rounds of variable selection, were used to develop algorithms for having a fatal CVD event. For each algorithm, all pre-specified candidate predictors were included in a single logistic regression model for the first round of stepwise selection and those with a p-value ≤ 0.05 were retained for the second round of stepwise selection. During the second round of variable selection, the AHRQ CCS categories were added to a logistic regression model with pre-specified candidate predictors retained following the first round of selection. Given the large number of AHRQ CCS categories, these were added to the model one at a time to avoid overfitting the model [14]. Pre-specified candidate predictors with p-values ≤ 0.05 in the first round and AHRQ CCS categories with p-values ≤ 0.05 in the second round of selection were included in a single logistic regression model for the third round of stepwise selection. Only candidate predictors with a p-value ≤ 0.05 in the third round of selection were included in the final algorithm. A separate algorithm was developed for fatal CHD using this approach.

Indirect approach 1

The first indirect approach used separate logistic regression models for fatal MI, CHD death and fatal stroke, with three rounds of selection to identify candidate predictors to retain for each outcome, as described above for the direct approach. Participants with a predicted probability higher than a pre-specified cut-point for fatal MI, CHD death or fatal stroke were defined as having a high predicted probability for fatal CVD. Participants with a predicted probability higher than a pre-specified cut-point for fatal MI or CHD death were defined as having a high predicted probability for fatal CHD. A description of how the pre-specified cut-points were selected is provided in the *Comparison of the Direct and Indirect Approaches* section below.

Indirect approach 2

The second indirect approach modeled the outcomes of fatal MI, CHD death and fatal stroke, separately, as described in the indirect approach 1 above. The predicted probabilities for each outcome from the three separate models were used as independent variables in a

logistic regression model with fatal CVD as the outcome. In a separate analysis, the predicted probabilities of fatal MI and CHD death were used to model fatal CHD as the outcome.

Comparison of the Direct and Indirect approaches

We identified the cut-point of predicted probability for each algorithm that resulted in the closest percentage of the population with predicted and observed events. Participants with a predicted probability above this cut-point were categorized as having a high probability of that outcome (i.e., a positive test result). To compare the direct and both indirect approaches for developing algorithms for the composite outcomes, we calculated sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). The C-index was calculated for the direct approach and the second indirect approach. Confidence intervals for the C-index were calculated using a bootstrap method with 1,000 replications [15]. The C-index cannot be calculated for the first indirect approach as this approach does not provide predicted probabilities for the composite outcomes.

Evaluation of final algorithms

We assessed the performance of the final algorithms using sensitivity, specificity, NPV and PPV and plotted the receiver operating characteristic curve. We assessed the improvement in discrimination for the final algorithms compared with hospital discharge diagnosis codes using the net reclassification index (NRI) [16, 17]. Choosing the cut-point of predicted probability that provides the closest number of predicted and observed events to define a high probability of an outcome may not meet the needs of all researchers. Therefore, we calculated test characteristics and discrimination using three additional cut-points for each outcome. These included cut-points that provided (1) the maximum specificity with a sensitivity 0.80; (2) the maximum NPV with a PPV 0.80, and (3) the maximum sum of sensitivity and specificity. All analyses were conducted using SAS 9.4 (SAS Institute Inc., Cary, NC).

Results

Among 20,403 REGARDS participants with data linked to Medicare, 4,327 died between January 1, 2003 and December 31, 2013 of whom 2,685 were included in current analyses (Figure 2). Among participants included in current analyses, mean age 79.6 years, 41.3% women, 32.8% black, 3.5% and 4.1 had a hospital discharge diagnosis code for MI and stroke, respectively, within 28 days prior to their death (Table 2). Based on adjudicated REGARDS study data, 608 (22.6%) deaths were classified as fatal CVD, 479 (17.8%) as fatal CHD and 142 (5.3%) as fatal stroke, with 13 participants having both a fatal CHD and a fatal stroke.

Comparison of the direct and indirect approaches for algorithm development

The predictors with the strongest associations with fatal CVD when using the direct approach were hospital discharge diagnosis codes for stroke and MI in any position within 28 days of death date, and for fatal CHD was hospital discharge diagnosis codes for MI in any position within 28 days of death date (Supplemental Table 2). When each component of

fatal CVD was modeled separately, the strongest predictors for fatal stroke and fatal MI were hospital discharge diagnosis codes for stroke and MI in any position within 28 days of death date respectively. The C-index for developing the fatal CVD and fatal CHD algorithms were similar using the direct approach and the second indirect approach (Table 3). For fatal CVD, sensitivity and PPV were higher when using the indirect approaches compared with the direct approach. For fatal CHD, sensitivity and PPV were within 0.02 when using the indirect approaches and the direct approach. Based on these results, the second indirect approach was used to model fatal CVD and fatal CHD.

Performance of final algorithms for fatal CVD, fatal CHD and fatal stroke

The C-index was >0.85 for fatal CVD, fatal CHD and fatal stroke (Figure 3). For each cut-point evaluated, the sensitivity was higher using the algorithms compared with hospital discharge diagnosis codes (Table 4). The specificity was lower using the algorithms versus hospital discharge diagnosis codes when defining high probability using cut-points resulting in the closest percentage of predicted and observed events, maximum specificity with sensitivity >0.80 or the maximum sum of sensitivity and specificity. When cut-points of predicted probability were chosen to achieve the highest NPV with PPV >0.80 , the specificity was similar using the algorithms and hospital discharge diagnosis codes. For each cut-point used to define a high probability of fatal CVD and fatal CHD, the algorithms resulted in a positive NRI compared to hospital discharge diagnosis codes (Table 4 and Supplemental Tables 3 and 4). When cut-points were chosen to provide the closest percentage of predicted and observed fatal CVD events, maximum specificity with sensitivity >0.80 or maximize the sum of sensitivity and specificity, the algorithm for fatal stroke resulted in a positive NRI. (Table 4 and Supplemental Table 5). When choosing a cut-point to achieve maximum NPV while maintaining a PPV >0.80 , the algorithm for fatal stroke resulted in a positive, non-significant NRI.

Discussion

We developed claims-based algorithms for discriminating fatal CVD, fatal CHD and fatal stroke, separately, versus other causes of death. Each algorithm showed good discrimination. The algorithms had a positive NRI compared with using hospital discharge diagnoses. We provided cut-points for identifying a high probability of fatal CVD, CHD or stroke based on the algorithms to obtain good calibration (i.e., the same percentage of observed and predicted events), high sensitivity, high PPV and maximizing the sum of sensitivity and specificity. Figure 4 and Supplemental Tables 6-8 provide details on how to use the algorithms.

The majority of Medicare beneficiaries are >65 years of age, consequently the mortality rate is high in this population [7, 8]. One-third of CVD events among Medicare beneficiaries are fatal [18]. Causes of death are available only for Medicare beneficiaries who died between 2006 and 2008 through a linkage with the National Death Index. Using these causes of death is not ideal as the agreement on cause of death between death certificates and clinician adjudicators is moderate (kappa statistic=0.54) [19, 20]. Studies assessing CVD risk in Medicare often analyze CVD hospitalizations as outcomes [4, 5, 21, 22]. This approach will

underestimate the rate of CVD as not all deaths are preceded by a claim with an inpatient diagnosis code for CVD. In the current study, only 28.5% of participants with a fatal CVD event had a hospital discharge diagnosis code for MI or stroke within 28 days of their death. Also risk factors may show different associations with fatal and nonfatal CVD events [9, 23].

The hospital discharge diagnosis method has been used previously to define fatal MI and stroke [24]. The sensitivity of this approach was low in the current study. The claims-based algorithms that we developed had higher sensitivity compared with this approach, while maintaining high specificity, and improving discrimination as indicated by a positive NRI. To minimize the misclassification of CVD events, we recommend using the algorithms developed herein rather than hospital discharge diagnoses.

Those interested in using the algorithms that we developed may have different objectives. Therefore, we evaluated thresholds with high calibration, sensitivity, and PPV and balancing sensitivity and specificity for defining a high probability of fatal CVD, fatal CHD and fatal stroke based on the algorithms. Multiple cut-points can be applied to determine fatal CVD outcomes based on the algorithms as sensitivity analyses to assess the robustness of findings.

Data from cohorts including adjudicated fatal and nonfatal CVD events have been used to develop the Pooled Cohort risk equations [25]. The Pooled Cohort risk equations estimate 10-year predicted risk for fatal and non-fatal CVD events and have been incorporated into clinical practice guidelines [25, 26]. Although investigators may be interested in calculating CVD events rates in claims data using the outcomes from the Pooled Cohort risk equations, this has not been possible previously. A high predicted probability for a fatal CVD event based on the algorithms developed in the current study can be combined with non-fatal outcomes to calculate CVD rates that resemble the outcome of the Pooled Cohort equations.

To avoid data inconsistencies, we excluded 130 REGARDS participants with conflicting death date in the REGARDS database and the Medicare beneficiary summary file. For 40 of these 130 participants, the death date only differed by one day; for 38 participants, Medicare recorded the death date as the last day of the same month that REGARDS adjudicated as the death date; for 17 participants, there was no death date recorded in the Medicare beneficiary summary file; and for the remaining 35 participants, there was no clear relationship between the death dates recorded in the REGARDS database and Medicare. Given that these 130 deaths only represented 3% of the total 4,327 with deaths linked in REGARDS and Medicare, these exclusions are unlikely to have had an effect on our algorithm.

The current study has several strengths including a relatively large sample size, including 2,685 deaths, of which 608 were fatal CVD events. Fatal events in the REGARDS study were adjudicated by two trained investigators with an adjudication committee resolving any disagreement. The REGARDS study enrolled adults from across the US and participants with Medicare fee-for-service coverage are comparable to all US adults aged 65 years or older with this insurance coverage [11]. The current study also has some limitations. The REGARDS study was restricted to blacks and whites and the algorithm needs to be evaluated in other race/ethnic groups. Also, we restricted the analysis to adults 65 years of

age. Although most deaths in the US occur among adults in this age group [27], the validity of the algorithms in populations <65 years of age needs to be tested. These algorithms should also be validated in external datasets.

In conclusion, we developed claims-based algorithms that can be used to identify fatal CVD, CHD and stroke events with high sensitivity and specificity. These algorithms can be used to investigate factors associated with fatal CVD, CHD and stroke among Medicare beneficiaries. Adding fatal events with a high probability of being CVD-related on these algorithms to non-fatal CVD events identified through claims data should provide a more accurate event rate and less biased exposure-outcome associations.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The work for this manuscript was funded by an industry /academic collaboration between Amgen Inc. and University of Alabama at Birmingham. REGARDS project is supported by a cooperative agreement (U01 NS041588) from the National Institute of Neurological Disorders and Stroke, National Institutes of Health, Department of Health and Human Services. Additional funding was provided by grants from the Agency for Healthcare Research and Quality (R01-HS-8517) and the National Institutes of Health (R01HL080477 and K24HL111154-K24).

References

1. Administration on Aging: Administration for Community Living. Profile of older Americans: 2013, Washington, DC: US Department of Health and Human Services; 2014 http://www.aoa.gov/Aging_Statistics/Profile/2013/index.aspx. (Accessed at April 25, 2017).
2. Nonell L: Total Number of Medicare Beneficiaries. Timeframe: 2015 2017 <http://kff.org/medicare/state-indicator/total-medicare-beneficiaries/?currentTimeframe=0&sortModel=%7B%22colId%22:%22Location%22,%22sort%22:%22asc%22%7D>. (Accessed at April 25, 2017).
3. Heron M: National Vital Statistics Report. 2015, 64(10). http://www.cdc.gov/nchs/data_access/Vitalstatsonline.htm. (Accessed at April 25, 2017).
4. Serban MC, Colantonio LD, Manthripragada AD, Monda KL, Bittner VA, Banach M, Chen L, Huang L, Dent R, Kent ST et al.: Statin Intolerance and Risk of Coronary Heart Events and All-Cause Mortality Following Myocardial Infarction. *Journal of the American College of Cardiology* 2017, 69(11):1386–1395. [PubMed: 28302290]
5. Zhang J, Xie F, Yun H, Chen L, Muntner P, Levitan EB, Safford MM, Kent ST, Osterman MT, Lewis JD et al.: Comparative effects of biologics on cardiovascular risk among older patients with rheumatoid arthritis. *Annals of the Rheumatic Diseases* 2016, 75(10): 1813–18. [PubMed: 26792814]
6. Spatz ES, Beckman AL, Wang Y, Desai NR, Krumholz HM: Geographic Variation in Trends and Disparities in Acute Myocardial Infarction Hospitalization and Mortality by Income Levels, 1999–2013. *JAMA Cardiology* 2016, 1(3):255–265. [PubMed: 27438103]
7. Henry J Kaiser Family Foundation: Distribution of Medicare Beneficiaries by Eligibility Category. *State Health Facts* 2013.
8. Roth DL, Skarupski KA, Crews DC, Howard VJ, Locher JL: Distinct age and self-rated health crossover mortality effects for African Americans: Evidence from a national cohort study. *Social Science & Medicine* 2016, 156:12–20. [PubMed: 27015163]
9. Safford MM, Brown TM, Muntner PM, Durant RW, Glasser S, Halanych JH, Shikany JM, Prineas RJ, Samdarshi T, Bittner VA et al.: Association of race and sex with risk of incident acute coronary

- heart disease events. JAMA : the Journal of the American Medical Association 2012, 308(17):1768–1774. [PubMed: 23117777]
10. Howard VJ, Cushman M, Pulley L, Gomez CR, Go RC, Prineas RJ, Graham A, Moy CS, Howard G: The reasons for geographic and racial differences in stroke study: objectives and design. Neuroepidemiology 2005, 25(3):135–143. [PubMed: 15990444]
 11. Xie F, Colantonio LD, Curtis JR, Safford MM, Levitan EB, Howard G, Muntner P: 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 2016, 184(7):532–544. [PubMed: 27651383]
 12. Luepker RV, Apple FS, Christenson RH, Crow RS, Fortmann SP, Goff D, Goldberg RJ, Hand MM, Jaffe AS, Julian DG 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. [PubMed: 14610011]
 13. Cost Healthcare and Project Utilization. Agency for Healthcare Research and Quality. Rockville M: Clinical Classifications Software (CCS) for ICD-9-CM <https://www.hcup-us.ahrq.gov/toolsoftware/ccs/ccs.jsp>. (Accessed at January 2017).
 14. Stoltzfus JC: Logistic regression: a brief primer. Academic Emergency Medicine : Official Journal of the Society for Academic Emergency Medicine 2011, 18(10):1099–1104. [PubMed: 21996075]
 15. Efron B, Halloran E, Holmes S: Bootstrap confidence levels for phylogenetic trees. Proceedings of the National Academy of Sciences of the United States of America 1996, 93(23):13429–13434. [PubMed: 8917608]
 16. Maarten LJG MV, Jacqueline WCM, Michael PJ, and Ewout SW: Net Reclassification Improvement: Computation, Interpretation, and Controversies. Annals of Internal Medicine 2014, 160:10.
 17. Pencina MJ, D'Agostino RB Sr., D'Agostino RB Jr., Vasan RS: Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. Statistics in Medicine 2008, 27(2):157–172; Discussion 207–112. [PubMed: 17569110]
 18. Pearce CA, Furberg CD, O'Meara ES, Psaty BM, Kuller L, Powe NR, Manolio T: Characteristics and baseline clinical predictors of future fatal versus nonfatal coronary heart disease events in older adults: the Cardiovascular Health Study. Circulation 2006, 113(18):2177–2185. [PubMed: 16651468]
 19. Halanych JH, Shuaib F, Parmar G, Tanikella R, Howard VJ, Roth DL, Prineas RJ, Safford MM: Agreement on cause of death between proxies, death certificates, and clinician adjudicators in the Reasons for Geographic and Racial Differences in Stroke (REGARDS) study. American journal of epidemiology 2011, 173(11):1319–1326. [PubMed: 21540327]
 20. Olubowale OT, Safford MM, Brown TM, Durant RW, Howard VJ, Gamboa C, Glasser SP, Rhodes JD, Levitan EB: Comparison of Expert Adjudicated Coronary Heart Disease and Cardiovascular Disease Mortality With the National Death Index: Results From the REasons for Geographic And Racial Differences in Stroke (REGARDS) Study. Journal of the American Heart Association 2017, 6(5).
 21. Brown TM, Deng L, Becker DJ, Bittner V, Levitan EB, Rosenson RS, Safford MM, Muntner P: Trends in mortality and recurrent coronary heart disease events after an acute myocardial infarction among Medicare beneficiaries, 2001–2009. American heart journal 2015, 170(2):249–255. [PubMed: 26299221]
 22. Graham DJ, Ouellet-Hellstrom R, MaCurdy TE, Ali F, Sholley C, Worrall C, Kelman JA: Risk of acute myocardial infarction, stroke, heart failure, and death in elderly Medicare patients treated with rosiglitazone or pioglitazone. JAMA : the journal of the American Medical Association 2010, 304(4):411–418. [PubMed: 20584880]
 23. Soliman EZ, Prineas RJ, Case LD, Russell G, Rosamond W, Rea T, Sotoodehnia N, Post WS, Siscovick D, Psaty BM et al.: Electrocardiographic and clinical predictors separating atherosclerotic sudden cardiac death from incident coronary heart disease. Heart 2011, 97(19):1597–1601. [PubMed: 21775508]

24. Matthew D Ritchey FL, Hilary K. Wall, Claudia A. Steiner, Cathleen Gillespie, Mary G. George, Janet S. Wright: Million Hearts: Description of the National Surveillance and Modeling Methodology Used to Monitor the Number of Cardiovascular Events Prevented During 2012–2016. *Journal of the American Heart Association* 2017, 6(5).
25. Goff DC Jr., Lloyd-Jones DM, Bennett G, Coady S, D'Agostino RB Sr., Gibbons R, Greenland P, Lackland DT, Levy D, O'Donnell CJ et al.: 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Journal of the American College of Cardiology* 2014, 63(25 Pt B):2935–2959. [PubMed: 24239921]
26. Stone NJ, Robinson JG, Lichtenstein AH, Goff DC Jr., Lloyd-Jones DM, Smith SC Jr., Blum C, Schwartz JS, Panel AACG: Treatment of blood cholesterol to reduce atherosclerotic cardiovascular disease risk in adults: synopsis of the 2013 American College of Cardiology/American Heart Association cholesterol guideline. *Annals of internal medicine* 2014, 160(5):339–343. [PubMed: 24474185]
27. Kenneth D. KochanekMA, Sherry L. Murphy, B. S., Jiaquan Xu, M.D., Betzaida Tejada-Vera, M. S., Division of Vita Statistics: Deaths: Final Data for 2014. *National Vital Statistics Reports* 2016, 65(4):122.
28. WHO: Stroke—1989. Recommendations on stroke prevention, diagnosis, and therapy. Report of the WHO Task Force on Stroke and other Cerebrovascular Disorders. *Stroke; a journal of cerebral circulation* 1989, 20(10):1407–1431.

Figure 1:

Direct and indirect approaches for developing the claims-based algorithms.

CHD: Coronary heart disease; CVD: Cardiovascular disease. MI: myocardial infarction.

Figure 2:
Flow chart for cohort selection.

Figure 3:
Receiver operating characteristic curve for fatal CVD, fatal CHD and fatal stroke.
CHD: Coronary heart disease; CVD: Cardiovascular disease.

Figure 4:

Steps to define fatal stroke, fatal CHD and fatal CVD using the claims-based algorithms.

AHRQ CCS category: Agency for Healthcare Research and Quality Clinical Classifications

Software; CHD: Coronary heart disease; CVD: Cardiovascular disease.

Table 1:

Definitions of fatal events in the REGARDS study.

Outcome	Description
Fatal CVD	Includes fatal CHD or fatal stroke
Fatal stroke	Was defined as adjudicated stroke followed by death within 28 days. A stroke was adjudicated following the World Health Organization (WHO) definition[28]. Events not meeting the WHO definition but characterized by stroke-related symptoms lasting less than 24 hours with neuroimaging consistent with acute infarct or hemorrhage were also classified as strokes
Fatal CHD	Includes a fatal MI or CHD death, in accordance to the 2003 “Case definitions for acute coronary heart disease in epidemiology and clinical research studies” [12].
Fatal MI	Was defined as adjudicated definite or probable MI followed by death within 28 days, following the 2003 “Case definitions for acute coronary heart disease in epidemiology and clinical research studies” [12]. A definite MI was defined by the presence of diagnostic enzyme elevations or diagnostic changes on the electrocardiogram. A probable MI was defined by the presence of non-diagnostic enzyme elevation and positive but non-diagnostic changes on an electrocardiogram. If enzymes were missing, events were adjudicated as a probable MI if an electrocardiogram has positive but non-diagnostic changes in addition to clinical signs or symptoms of ischemia.
CHD death	Was defined following the 2003 “Case definitions for acute CHD in epidemiology and clinical research studies” [12]. Specifically, CHD death was defined as a death from MI not meeting the criteria for definite or probable MI (e.g., death within six hours of hospital admission with cardiac symptoms and/or signs, but with biomarkers or electrocardiogram absent or not diagnostic), postmortem findings consistent with the occurrence of coronary occlusion within 28 days, or sudden death preceded by cardiac symptoms or signs without evidence of non-coronary causes).

CHD: coronary heart disease; CVD: cardiovascular disease; MI: myocardial infarction; REGARDS: Reasons for Geographic and Racial Differences in Stroke.

Table 2:

Distribution of pre-specified candidate predictors for participants with and without an adjudicated fatal CVD, CHD and stroke.

	All	Fatal CVD		Fatal CHD		Fatal stroke	
	N=2,685	No N=2,077	Yes N=608	No N=2,206	Yes N=479	No N=2,543	Yes N=142
Age at death date in years, Mean (STD)	79.6 (7.6)	79.7 (7.6)	79.1 (7.4)	79.7 (7.6)	78.9 (7.6)	79.6 (7.6)	80.1 (6.9)
Female	41.3%	41.3%	40.5%	41.5%	39.5%	40.9%	45.8%
Black race	32.8%	32.8%	34.5%	33.0%	34.0%	33.0%	35.9%
Hospital discharge diagnosis code (any position), within 28 days of death							
Myocardial infarction	3.5%	0.4%	14.0%	0.4%	17.85%	3.5%	2.1%
Stroke	4.1%	1.0%	15.0%	4.4%	2.9%	1.0%	59.9%
Physician visit diagnosis code, within 28 days of death							
Myocardial infarction	4.5%	1.5%	14.6%	1.5%	18.6%	4.6%	2.1%
Stroke	9.7%	6.4%	21.1%	10.3%	6.9%	6.1%	73.2%
Hospital discharge diagnosis code (any position), beyond 28 days of death							
Myocardial infarction	13.6%	12.0%	19.2%	11.7%	22.3%	14.0%	7.0%
Stroke	13.1%	12.3%	15.8%	13.5%	11.3%	12.1%	31.7%
Physician visit diagnosis code, beyond 28 days of death							
Myocardial infarction	15.6%	14.0%	21.2%	13.6%	24.6%	16.0%	7.8%
Stroke	29.5%	28.9%	31.4%	30.2%	25.9%	28.2%	51.4%
Hospital discharge diagnosis code (any position), any time prior to death							
Abdominal aortic aneurism or Peripheral arterial disease	11.7%	11.2%	13.3%	11.4%	12.9%	11.6%	13.4%
Malignancy	31.9%	37.8%	11.7%	36.4%	11.3%	33.0%	12.7%
Other heart disease	57.5%	56.5%	60.7%	56.6%	61.6%	57.5%	57.0%
Atrial fibrillation	34.8%	34.0%	37.5%	34.5%	36.3%	34.4%	43.0%
Sudden death or ventricular arrhythmias	14.6%	12.6%	21.6%	12.6%	23.8%	14.7%	13.4%
Other and unspecified intracranial hemorrhage	1.3%	1.0%	2.3%	1.5%	0.4%	0.9%	8.5%
Transient cerebral ischemia	6.6%	6.6%	6.4%	6.9%	5.0%	6.3%	11.3%
Viral or opportunistic infection	8.9%	9.3%	7.6%	9.3%	7.1%	9.0%	8.5%
End stage renal disease	7.4%	6.8%	9.2%	6.8%	10.2%	7.4%	6.3%
Chronic obstructive pulmonary disease	44.4%	45.3%	41.5%	44.8%	42.6%	44.9%	36.6%
Organ transplant	0.8%	0.8%	1.0%	0.8%	1.0%	0.8%	0.7%
Gastrointestinal perforation	1.0%	1.0%	0.8%	1.0%	1.0%	1.0%	0.0%
Interstitial lung disease	6.3%	6.6%	5.4%	6.5%	5.2%	6.3%	6.3%
Hernia, Intestinal obstruction, Crohn's disease	28.8%	30.2%	23.9%	29.7%	24.6%	29.2%	21.1%
Liver, gall and pancreas disease	15.9%	17.1%	11.8%	16.8%	12.1%	16.2%	10.6%
General symptoms	38.6%	39.3%	36.0%	39.7%	33.4%	38.2%	45.1%
Symptom involving respiratory and other chest symptoms	23.5%	23.8%	22.2%	23.7%	22.6%	23.7%	19.7%

	All	Fatal CVD		Fatal CHD		Fatal stroke	
	N=2,685	No N=2,077	Yes N=608	No N=2,206	Yes N=479	No N=2,543	Yes N=142
Other symptoms involving abdomen and pelvis	9.2%	10.2%	5.6%	9.8%	6.1%	9.5%	3.5%
Diagnosis code from primary hospital discharge or emergency room visit, any time prior to death							
Sudden death	5.3%	2.9%	13.5%	3.0%	16.3%	5.5%	2.8%
Physician visit diagnosis code, any time prior to death							
Abdominal aortic aneurism or peripheral arterial disease	26.2%	25.7%	28.1%	25.9%	27.8%	26.1%	28.9%
Malignancy	45.9%	51.4%	27.3%	50.1%	26.7%	46.9%	28.2%
Other heart disease	73.6%	72.2%	78.5%	72.6%	78.3%	73.4%	78.2%
Atrial fibrillation	40.3%	39.6%	42.8%	39.8%	42.6%	40.0%	45.1%
Physician visit diagnosis code, any time prior to death							
Sudden death or ventricular arrhythmias	21.3%	17.1%	35.7%	17.0%	41.3%	21.6%	15.5%
Other and unspecified intracranial hemorrhage	3.3%	2.7%	5.1%	3.7%	1.5%	2.5%	16.9%
Transient cerebral ischemia	21.0%	20.9%	21.2%	21.8%	17.1%	20.3%	33.8%
Viral or opportunistic infection	32.4%	33.1%	30.1%	33.1%	29.4%	32.5%	31.7%
End stage renal disease	20.9%	20.7%	21.9%	20.0%	25.1%	21.5%	10.6%
Chronic obstructive pulmonary disease	59.0%	59.5%	57.6%	59.2%	58.3%	59.2%	56.3%
Organ transplant	0.9%	0.9%	1.2%	0.9%	1.3%	0.9%	0.7%
Gastrointestinal perforation	1.5%	1.7%	0.7%	1.6%	0.8%	1.5%	0.0%
Interstitial lung disease	7.9%	8.3%	6.7%	8.1%	7.1%	8.1%	5.6%
Hernia, Intestinal obstruction, Crohn's disease	39.3%	40.3%	35.9%	40.3%	34.7%	39.2%	41.6%
Liver, gall and pancreas disease	21.5%	22.9%	16.5%	22.5%	16.5%	21.8%	16.2%
General symptoms	88.5%	89.2%	86.2%	89.5%	83.9%	88.1%	95.1%
Symptom involving respiratory and other chest symptoms	85.0%	85.3%	83.9%	85.1%	84.6%	85.2%	81.7%
Other symptoms involving abdomen and pelvis	57.8%	60.1%	50.2%	59.7%	49.5%	58.2%	52.1%
Hospital discharge (any position) or physician visit diagnosis code, any time prior to death							
Diabetes	53.5%	51.3%	61.2%	51.8%	61.4%	53.1%	62.0%
Obesity	18.6%	18.6%	18.6%	18.3%	20.3%	18.9%	13.4%
Hyperlipidemia	79.0%	78.3%	81.4%	78.5%	81.4%	78.9%	81.7%
Hypertension	93.1%	92.6%	94.6%	92.9%	93.7%	92.8%	97.9%
Smoking	37.5%	39.0%	32.4%	38.5%	32.8%	37.9%	29.6%
ICD-9-CM procedure code or CPT code, any time prior to death							
Coronary artery bypass grafting or percutaneous coronary intervention	14.7%	14.7%	23.0%	14.6%	25.5%	16.7%	14.8%
Any time prior to death or otherwise specified							
Died in hospital	32.3%	30.1%	39.5%	31.4%	36.1%	31.0%	52.1%
Discharge to skilled nursing facility	42.6%	45.2%	33.7%	45.3%	30.3%	42.4%	46.5%
Discharged to hospice	18.4%	20.4%	11.5%	20.4%	9.4%	18.3%	19.7%
28 days or more for last hospitalization	3.7%	4.0%	2.6%	4.0%	2.3%	3.7%	4.2%

	All	Fatal CVD		Fatal CHD		Fatal stroke	
	N=2,685	No N=2,077	Yes N=608	No N=2,206	Yes N=479	No N=2,543	Yes N=142
Number of cardiologist visit within 6 months prior to death	4.9 (8.0)	4.6 (7.8)	6.1 (8.5)	4.6 (7.7)	6.4 (9.0)	4.9 (8.1)	4.7 (5.9)
Number of oncologist visit within 6 months prior to death	2.5 (7.7)	3.1 (8.6)	0.5 (2.1)	3.0 (8.4)	0.6 (2.3)	2.7 (7.9)	0.2 (0.9)
Number of neurologist visit within 6 months prior to death	1.1 (3.6)	1.0 (3.5)	1.5 (3.9)	1.1 (3.7)	1.0 (3.0)	1.0 (3.4)	3.3 (5.8)

Numbers are mean (standard deviation) or percentage.

CHD: Coronary heart disease; CVD: Cardiovascular disease.

ICD-9-CM: International classification of diseases, 9th version, clinical modification.

CPT: Current procedure terminology code.

STD: Standard deviation

Table 3:

Performance of the direct and indirect approaches for discriminating fatal cardiovascular disease and coronary heart disease.

Outcome	Number predictors	C-index (95% CI)	Cut-points [‡]	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
Fatal CVD							
Direct approach [‡]	27	0.86 (0.84, 0.88)	0.32	0.61 (0.57, 0.65)	0.88 (0.87, 0.89)	0.60 (0.56, 0.63)	0.88 (0.87, 0.90)
Indirect approach 1 [§]	NA	NA	NA	0.62 (0.58, 0.65)	0.89 (0.88, 0.90)	0.62 (0.58, 0.66)	0.89 (0.87, 0.90)
Indirect approach 2 [¶]	#	0.87 (0.85, 0.89)	0.27	0.64 (0.61, 0.68)	0.90 (0.88, 0.91)	0.65 (0.61, 0.69)	0.90 (0.88, 0.91)
Fatal CHD							
Direct approach [‡]	26	0.86 (0.84, 0.88)	0.30	0.58 (0.54, 0.63)	0.91 (0.89, 0.92)	0.57 (0.53, 0.62)	0.91 (0.90, 0.92)
Indirect approach 1 [§]	NA	NA	NA	0.56 (0.51, 0.60)	0.91 (0.90, 0.92)	0.56 (0.52, 0.61)	0.90 (0.89, 0.92)
Indirect approach 2 [¶]	#	0.86 (0.84, 0.88)	0.24	0.58 (0.54, 0.62)	0.91 (0.90, 0.92)	0.59 (0.54, 0.63)	0.91 (0.90, 0.92)
Components used in the indirect approaches							
Fatal MI	20	0.91 (0.89, 0.94)	0.22	0.59 (0.52, 0.66)	0.97 (0.96, 0.98)	0.59 (0.52, 0.66)	0.97 (0.96, 0.98)
CHD death	25	0.87 (0.85, 0.89)	0.29	0.52 (0.46, 0.57)	0.92 (0.91, 0.93)	0.51 (0.46, 0.56)	0.93 (0.91, 0.94)
Fatal stroke	13	0.95 (0.92, 0.97)	0.27	0.70 (0.63, 0.78)	0.98 (0.98, 0.99)	0.72 (0.64, 0.79)	0.98 (0.98, 0.99)

CHD: Coronary heart disease; CI: confidence interval; CVD: Cardiovascular disease; MI: myocardial infarction; NA: not applicable, fatal event here was identified based on the identification of each component; NPV: Negative predictive value; PPV: Positive predictive value.

[‡]: Cut-points: Cut-points were chosen to provide the closest observed-to-predicted events. Participants with a predicted probability equal to or greater than the cut-points were considered to have a positive test result.

[§]: Direct approach: Model probability of event (e.g., fatal cardiovascular disease) on candidate predictors.

[¶]: First indirect approach: conducted separate logistic regression models for fatal MI, CHD death and fatal stroke, using stepwise selection for each outcome, as described above. For each of these outcomes, we identified the cut-point of predicted probability result in a closest observed-to-predicted event and had a predicted probability above the cut-point (first indirect approach). Participants with a predicted probability higher than the specific cut-point for any outcome were considered to have a high probability of fatal CVD (i.e., a positive test result).

[#]: Second indirect approach: used the predicted probabilities for fatal MI, CHD death and fatal stroke from the three separate models described above as the independent variables in a logistic regression model with fatal CVD as the outcome (second indirect approach).

^{*}: Predicted probability for fatal CVD was derived using a model with three components (predicted probability of fatal MI, fatal stroke, and CHD death). Predicted probability of each component was determined in separate models (see Components used in the indirect approaches in this table and Table 3). Predicted probability for the 2 components (fatal MI, and CHD death).

Table 4:

Performance of algorithms for discriminating cause of death under different cut-points of predicted probability compared with using hospital discharge diagnosis codes.

Outcome (Observed proportion)	Approach for defining high probability of the outcome	Cut- points [†]	Predicted Proportion %	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	Sum of sensitivity and specificity	NRI (95% CI)
Fatal CVD (22.7%)	Discharge diagnosis codes [‡]	NA	7.5	0.28 (0.25, 0.32)	0.99 (0.98, 0.99)	0.86 (0.81, 0.91)	0.82 (0.81, 0.84)	1.27	Reference
	Closest observed-to-predicted	0.27	22.6	0.64 (0.61, 0.68)	0.90 (0.88, 0.91)	0.65 (0.61, 0.69)	0.90 (0.89, 0.91)	1.56	0.285 (0.244, 0.415)
	0.80 sensitivity	0.15	38.5	0.82 (0.78, 0.85)	0.74 (0.72, 0.76)	0.48 (0.45, 0.51)	0.93 (0.92, 0.94)	1.56	0.286 (0.242, 0.330)
	0.80 PPV	0.54	13.0	0.46 (0.42, 0.50)	0.97 (0.96, 0.98)	0.81 (0.77, 0.85)	0.86 (0.85, 0.87)	1.43	0.160 (0.130, 0.194)
	Maximize the sum of sensitivity and specificity	0.17	34.5	0.78 (0.75, 0.82)	0.78 (0.77, 0.80)	0.51 (0.48, 0.55)	0.93 (0.91, 0.94)	1.57	0.296 (0.251, 0.340)
Fatal CHD (17.8%)	Discharge diagnosis codes [§]	NA	3.5	0.18 (0.14, 0.21)	1.00 (0.99, 1.00)	0.91 (0.86, 0.97)	0.85 (0.83, 0.86)	1.17	Reference
	Closest observed-to-predicted	0.24	17.7	0.58 (0.54, 0.62)	0.91 (0.90, 0.92)	0.59 (0.54, 0.63)	0.91 (0.90, 0.92)	1.49	0.317 (0.272, 0.364)
	0.80 sensitivity	0.12	35.3	0.81 (0.78, 0.85)	0.75 (0.73, 0.77)	0.41 (0.38, 0.44)	0.95 (0.94, 0.96)	1.56	0.385 (0.337, 0.432)
	0.80 PPV	0.57	8.2	0.36 (0.32, 0.41)	0.98 (0.97, 0.99)	0.80 (0.75, 0.85)	0.88 (0.86, 0.89)	1.35	0.174 (0.140, 0.212)
	Maximize the sum of sensitivity and specificity	0.13	32.7	0.79 (0.75, 0.83)	0.77 (0.76, 0.79)	0.43 (0.40, 0.46)	0.95 (0.94, 0.96)	1.56	0.388 (0.339, 0.435)
Fatal stroke (5.3%)	Discharge diagnosis codes [¶]	NA	4.1	0.60 (0.52, 0.68)	0.99 (0.99, 0.99)	0.77 (0.69, 0.84)	0.98 (0.97, 0.98)	1.59	Reference
	Closest observed-to-predicted	0.27	5.2	0.70 (0.63, 0.78)	0.98 (0.98, 0.99)	0.72 (0.64, 0.79)	0.98 (0.98, 0.99)	1.69	0.114 (0.103, 0.182)
	0.80 sensitivity	0.13	7.5	0.80 (0.74, 0.87)	0.97 (0.96, 0.97)	0.57 (0.50, 0.64)	0.99 (0.98, 0.99)	1.77	0.180 (0.119, 0.258)
	0.80 PPV	0.44	4.1	0.63 (0.55, 0.71)	0.99 (0.99, 1.00)	0.81 (0.74, 0.88)	0.98 (0.97, 0.98)	1.62	0.030 (-0.031, 0.090)
	Maximize the sum of sensitivity and specificity	0.06	12.6	0.87 (0.81, 0.92)	0.92 (0.91, 0.93)	0.36 (0.31, 0.42)	0.99 (0.99, 1.00)	1.78	0.194 (0.124, 0.276)

CHD: Coronary heart disease; CI: Confidence interval; CVD: Cardiovascular disease; NA: Not applicable, dichotomous variable; NPV: Negative predictive value; NRI: Net reclassification improvement; PPV: Positive predictive value.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

⁷; Cut-point: Participants with a predicted probability above this cut-point are identified as having a fatal event by algorithms.
⁸; Discharge diagnosis code for cardiovascular disease included International Classification of Diseases, 9th version, clinical modification (ICD-9-CM) 410.xx, 430.xx, 431.xx, 433.x1, 434.x1, 436.xx.
⁸; Discharge diagnosis code for coronary heart disease included ICD-9-CM 410.xx.
⁹; Discharge diagnosis codes for stroke included ICD-9-CM 430.xx, 431.xx, 433.x1, 434.x1, 436.xx