Journal of the American Heart Association

BRIEF COMMUNICATION

Bias Implications of Outcome Misclassification in Observational Studies Evaluating Association Between Treatments and All-Cause or Cardiovascular Mortality Using Administrative Claims

Rishi J. Desai, MS, PhD; Raisa Levin, MS; Kueiyu Joshua Lin, MD, ScD; Elisabetta Patorno, MD, DrPH

BACKGROUND: The bias implications of outcome misclassification arising from imperfect capture of mortality in claims-based studies are not well understood.

METHODS AND RESULTS: We identified 2 cohorts of patients: (1) type 2 diabetes mellitus (n=8.6 million), and (2) heart failure (n=3.1 million), from Medicare claims (2012–2016). Within the 2 cohorts, mortality was identified from claims using the following approaches: (1) all-place all-cause mortality, (2) in-hospital all-cause mortality, (3) all-place cardiovascular mortality (based on diagnosis codes for a major cardiovascular event within 30 days of death date), or (4) in-hospital cardiovascular mortality, and compared against National Death Index identified mortality. Empirically identified sensitivity and specificity based on observed values in the 2 cohorts were used to conduct Monte Carlo simulations for treatment effect estimation under differential and nondifferential misclassification scenarios. From National Death Index, 1 544 805 deaths (549 996 [35.6%] cardiovascular deaths) in the type 2 diabetes mellitus cohort and 1 175 202 deaths (523 430 [44.5%] cardiovascular deaths) in the heart failure cohort were included. Sensitivity was 99.997% and 99.207% for the all-place all-cause mortality approach, whereas it was 27.71% and 33.71% for the in-hospital all-cause mortality approach in the type 2 diabetes mellitus and heart failure cohorts, respectively, with perfect positive predicted values. For all-place cardiovascular mortality, sensitivity was 52.01% in the type 2 diabetes mellitus cohort and 53.83% in the heart failure cohort with positive predicted values of 49.98% and 54.45%, respectively. Simulations suggested a possibility for substantial bias in treatment effects.

CONCLUSIONS: Approaches to identify mortality from claims had variable performance compared with the National Death Index. Investigators should anticipate the potential for bias from outcome misclassification when using administrative claims to capture mortality.

Key Words: bias ■ mortality ■ observational studies ■ outcome misclassification

Administrative claims data sources have served as an important resource to monitor postmarketing safety of medications, especially among patients under-represented in randomized controlled trials (RCTs).[1] More recently, there has been a substantial interest in exploring the utility of such sources for regulatory decision making, with the US Food and Drug Administration supporting demonstration projects aiming to replicate findings from completed RCTs or predict findings of ongoing RCTs using administrative claims.[2,3] Researchers have also explored the feasibility of using administrative claims for outcome

Correspondence to: Rishi J. Desai, MS, PhD, Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, 1620 Tremont St, Suite 3030-R, Boston, MA 02120. E-mail: rdesai@bwh.harvard.edu

Supplementary Materials for this article are available at https://www.ahajournals.org/doi/suppl/10.1161/JAHA.120.016906

For Sources of Funding and Disclosures, see page 6.

© 2020 The Authors. Published on behalf of the American Heart Association, Inc., by Wiley. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

JAHA is available at: www.ahajournals.org/journal/jaha

ascertainment after initial randomization to reduce costs and minimize loss to follow-up that is frequent in prospective RCTs.[4]

All-cause and cause-specific mortality are important outcomes for many clinical research studies of medications. Assessment of mortality in patient cohorts identified from administrative claims can be achieved through linkage with external data sources such as the National Death Index (NDI) or the Death Master File (DMF). NDI is a central database based on death certificate data and maintained by Centers for Disease Control and Prevention for the purpose of providing researchers with passive follow-up for mortality outcomes. DMF is a database maintained by the Social Security Administration based on mortality information from various sources such as family members, funeral homes, hospitals, States, Federal agencies, postal authorities, and financial institutions with a goal to prevent fraud and abuse of federally funded benefits. Mortality captured in NDI is generally complete; however, the linkage is expensive and recording lags of up to 2 years make it less useful for contemporary cohorts. DMF linkage is less expensive and more contemporary information is available, but the recording is incomplete because after 2011 deaths reported by state agencies are not captured for certain states.[5]

It may also be possible to assess mortality information directly from administrative claims data sources. Some sources such as Medicare claims administratively capture all-cause all-place mortality, while some commercial insurance claims databases only contain information on deaths occurring during hospital admissions.[3] Further, information on the cause of death is not available in administrative claims, which has led researchers to use ad hoc approaches such as attributing the cause as cardiovascular for deaths occurring within 30 days of a healthcare encounter where a major cardiovascular condition such as myocardial infarction is recorded.[6,7] Direct assessment of mortality from administrative claims is appealing because it avoids challenges associated with linking to alternate sources such as NDI or DMF. However, the bias implications of outcome misclassification arising from imperfect capture of mortality or of its cause in studies conducted using administrative claims have not been systematically evaluated. To this end, the key objectives of the current investigation were to (1) report performance characteristics of various approaches for ascertaining mortality in administrative claims using NDI-recorded mortality as the criterion standard, and (2) investigate the impact of outcome misclassification on treatment effect estimates using simulated scenarios.

METHODS

Data Sources and Study Cohorts

We used 2012 to 2016 Medicare claims data to create 2 separate cohorts of patients older than 65 years of age with (1) type 2 diabetes mellitus (T2D), and (2) heart failure (HF), to evaluate potential variation in performance of various administrative claims–based approaches for identifying mortality across disease populations. Patients were included based on recorded *International Classification of Diseases, Ninth Revision* or *Tenth Revision* (*ICD-9* or *ICD-10*) diagnosis codes for T2D or HF in inpatient or outpatient claims after 6 months of continuous enrollment in Medicare (Parts A [inpatient coverage], B [outpatient coverage], and D [prescription benefits]). To increase the specificity of diagnosis, we excluded patients with *ICD* codes of secondary diabetes mellitus or type 1 diabetes mellitus in the 6-month continuous enrollment period from the T2D cohort. For the HF cohort, we excluded patients without echocardiogram or cardiac catheterization *Current Procedural Terminology* codes in 30 days before the HF diagnosis. Linkage of the NDI to Medicare claims was established at the patient level for the study years. Mortality information recorded in the NDI, along with the cause, was considered criterion standard for comparison against claims-based mortality assessment approaches. A signed data use agreement with the Centers for Medicare and Medicaid Services was available and the Brigham and Women's Hospital's Institutional Review Board approved this study. Patient consent was waived because of use of deidentified data. Because of the data use agreement with the Centers for Medicare and Medicaid Services, administrative claims data will not be made available to other researchers for purposes of reproducing the results or replicating the procedure to protect patient privacy.

Mortality Assessment From Administrative Claims

We implemented the following approaches to identify mortality from Medicare claims: (1) *all-place all-cause mortality*: defined based on mortality recorded in Master Beneficiary Summary File or hospitalization claims with discharge status of death, (2) *in-hospital all-cause mortality*: defined based on hospitalization claims with discharge status of death, (3) *all-place cardiovascular mortality*: identified based on the presence of diagnosis codes for myocardial infarction, ischemic stroke, intracranial hemorrhage, sudden cardiac death, or hospitalization for HF within 30 days of death date recorded in Master Beneficiary Summary File or hospitalization claims with discharge status of death,[6,7] (4) *in-hospital cardiovascular mortality*: identified based on the presence of diagnosis codes for

myocardial infarction, ischemic stroke, intracranial hemorrhage, sudden cardiac death, or hospitalization for heart failure within 30 days of death date recorded hospitalization claims with discharge status of death.[6,7] The purpose for implementing approaches that only use hospitalization claims to identify mortality was to report performance in circumstances where this is the only source of mortality information, for instance, when using certain commercial insurance claims data sources.[3]

Statistical Analysis

For claims-based approaches to identify all-cause and cardiovascular mortality, we reported positive predictive value (PPV), sensitivity, and false positive rate (FPR or 1-specificity) in the 2 cohorts using NDI-based death (and cause) as the criterion standard.

Next, we designed a series of Monte Carlo simulations to investigate the impact of misclassification of mortality on treatment effect estimates in observational studies conducted using administrative claims data. In order to base our simulation parameters on realistic values of treatment-mortality association and event rates, we extracted summary-level data from 2 large clinical trials that had cardiovascular mortality as a component of their primary endpoints: (1) the BI 10773 (Empagliflozin) Cardiovascular Outcome Event Trial in Type 2 Diabetes Mellitus Patients (EMPA-REG OUTCOME) trial comparing empagliflozin with placebo in patients with T2D,[8] and (2) the prospective comparison of ARNI (angiotensin receptor–neprilysin inhibitor) with ACEI (angiotensin-converting–enzyme inhibitor) to determine impact on global mortality and morbidity in heart failure (PARADIGM-HF) trial comparing sacubitril-valsartan with enalapril.[9] We used ranges for sensitivity and FPR (or 1-specificity) based on observed values in our 2 empirical cohorts as variable input parameters that were drawn from a uniform distribution in 1000 simulations. We simulated nondifferential misclassification of mortality between treatment and reference groups for all 4 approaches. Further, we noted that approaches that only used hospitalization claims for mortality assessment and the approach that used recoding of cardiovascular-specific *ICD* coding before mortality to identify cardiovascular deaths (approaches 2, 3, and 4 described above) had the potential for differential misclassification because these approaches relied on contact with the healthcare system for capture of the outcome of mortality. To wit, in observational studies, if patients included in the treatment group (for instance, users of a newly approved medication such as empagliflozin or sacubitril-valsartan) are better connected to the healthcare system because of higher socioeconomic status or geographic location of their residence than patients in the reference group (for instance, users of sulfonylurea or enalapril), then it is plausible to hypothesize that the probability of capturing mortality through administrative claims, particularly in-hospital mortality, would be higher for the treatment group. Therefore, we also simulated differential misclassification for these approaches by using a higher range of sensitivity values in the treatment group compared with the reference group, while holding the FPR constant. Tables S1 and S2 contains the input parameters for our Monte Carlo simulations. Results from the simulations were presented as distribution (median, 2.5th, and 97.5th percentile) of the treatment effect estimates—risk ratios (RRs) and risk differences (RDs)—for all scenarios considered.

RESULTS

Study Cohorts

We identified a total of 8.6 million Medicare beneficiaries with T2D with an average (SD) age of 74 (8) years and 3.1 million with HF with an average age of 79 (8) years. Table 1 summarizes the key demographic characteristics of the populations included. From NDI, we observed a total of 1 544 805 deaths (549 996 [35.6%] cardiovascular deaths) over a mean (SD) follow-up period of 2.4 (1.5) years in the T2D cohort and 1 175 202 deaths (523 430 [44.5%] cardiovascular deaths) over a mean follow-up period of 1.5 (1.3) years in the HF cohort (Table 2).

Identification of Mortality From Administrative Claims

For all-cause mortality, we observed a near complete capture with sensitivities in the range of 99.207% to

Table 1. Characteristics of the Study Cohorts

	Diabetes Mellitus Cohort	Heart Failure Cohort
Total sample	8 644 401	3 134 414
Age (mean [SD] y)	74 (8)	79 (8)
Male sex, %	43.1	43.2
Race, %		
White	77.9	82.8
Black	12	10.8
Others	10.1	6.4
Index year, %		
2012	41.7	16
2013	21.6	25.5
2014	13.8	21.9
2015	12.3	19.9
2016	10.5	16.7
Follow-up (mean [SD] y)	2.3 (1.5)	1.5 (1.3)

Table 2. Performance Characteristics of Administrative Claims–Based Approaches to Identify Mortality Compared With the National Death Index

Outcome Assessment Approach	Diabetes Mellitus Cohort (n=8 644 401)					Heart Failure Cohort (n=3 134 414)				
	Total Number of Events	Events Matching Accurately With NDI	Sensitivity*	PPV*	FPR*	Total Number of Events	Events Matching Accurately With NDI	Sensitivity*	PPV*	FPR*
NDI all-cause mortality	1 544 805	…	…	…		1 175 202	…	…	…	
NDI cardiovascular mortality	549 996	…	…	…		523 430	…	…	…	
Mortality identified from claims										
All-place all-cause mortality	1 544 757	1 544 757	99.997%	100.00%	0	1 165 885	1 165 885	99.207%	100.00%	0
In-hospital all-cause mortality	428 003	428 003	27.71%	100.00%	0	396 130	396 130	33.71%	100.00%	0
All-place cardiovascular mortality	572 341	286 041	52.01%	49.98%	28.78%	517 456	281 752	53.83%	54.45%	36.16%
In-hospital cardiovascular mortality	248 284	111 461	20.27%	44.89%	13.75%	240 858	116 950	22.34%	48.56%	19.01%

FPR indicates false positive rate; NDI, National Death Index; and PPV, positive predicted value.

*For all-cause mortality, PPVs can be interpreted as the probability of being a true case of death, given identification as dead from administrative claims; sensitivity can be interpreted as the probability of being identified as dead from administrative claims for a true death; and FPR can be interpreted as proportion of alive patients incorrectly identified as dead from administrative claims. For cardiovascular mortality, PPVs can be interpreted as the probability of being a true case of cardiovascular death, given identification as cardiovascular death from administrative claims; sensitivity can be interpreted as the probability of being identified as cardiovascular death from administrative claims for a true cardiovascular death; and FPR can be interpreted as proportion of patients with noncardiovascular deaths incorrectly identified as cardiovascular deaths from administrative claims.

99.997% with perfect PPVs in both cohorts when using mortality recording from the Master Beneficiary Summary File or hospitalization claims with discharge status of death (Table 2). The approach that only used hospitalization claims to capture all-cause mortality had sensitivity of 27.71% in the T2D cohort and 33.71% in the HF cohort, again with perfect PPVs.

The approach identifying all-place cardiovascular mortality had sensitivity of 52.01% in the T2D cohort and 53.83% in the HF cohort and PPVs of 49.98% and 54.45%, respectively. Importantly, the FPR (proportion of patients with noncardiovascular deaths incorrectly identified as cardiovascular deaths from administrative claims) was 28.78% in the T2D cohort and 36.16% in the HF cohort. When using only hospitalization claims for identifying cardiovascular mortality, the sensitivities and PPVs were compromised, but FPRs improved (Table 2).

Results From Monte Carlo Simulations

Table 3 summarizes results from the simulations. For all-cause mortality, we simulated RRs of 0.68 and 0.84 and RDs/100 person-years of −0.92 and −1.38 for EMPA-REG and PARADIGM scenarios, respectively. Because of reliable capture of all-place all-cause mortality from Medicare administrative claims, no bias caused by outcome misclassification was observed in estimating RRs or RDs for all-cause mortality in both EMPA-REG (median [2.5th, 97.5th percentile] RR 0.68 [0.68, 0.68]; RD −0.92 [−0.93, −0.91]) and PARADIGM scenarios (RR 0.84 [0.84, 0.84]; RD −1.38 [−1.38, −1.37]). When only using hospitalization claims for ascertaining all-cause mortality, RRs remained unbiased under nondifferential misclassification (median [2.5th, 97.5th percentile] RR 0.68 [0.67, 0.68] for EMPA-REG inputs; 0.84 [0.84, 0.85] for PARADIGM inputs). However, under plausible ranges of differential misclassification, RRs were substantially biased towards the null when only using hospitalization claims for ascertaining all-cause mortality for both scenarios (median [2.5th, 97.5th percentile] RR 0.81 [0.7, 0.92] for EMPA-REG inputs; 1.00 [0.87, 1.14] for PARADIGM inputs). RDs were severely biased under differential and nondifferential misclassification because of compromised sensitivity in capturing mortality only with hospitalization claims (Table 3).

For cardiovascular mortality, we simulated RRs of 0.62 and 0.80 and RDs/100 person-years of −0.78 and −1.49 for EMPA-REG and PARADIGM scenarios, respectively. For all-place cardiovascular mortality, a bias towards the null was noted in RRs under differential as well as nondifferential misclassification (median [2.5th, 97.5th percentile] RRs 0.66 [0.65, 0.66] and 0.70 [0.67, 0.74], respectively, for EMPA-REG inputs;

Table 3. Results From Monte Carlo Simulations for Treatment Effect Estimation When Ascertaining Mortality From Administrative Claims–Based Sources

Outcome	Monte Carlo Simulation Input Parameters*				Monte Carlo Simulation Results: Distribution of Point Estimates (Median, 2.5th, 97.5th Percentile)			
	Scenario	True (Simulated) Measures of Effect		Misclassification Type	All-Place Mortality		In-Hospital Mortality	
		RR	RD/100 Person-Years		RR	RD/100 Person-Years	RR	RD/100 Person-Years
All-cause mortality	EMPA-REG inputs	0.68	−0.92	Nondifferential	0.68 (0.68, 0.68)	−0.92 (−0.93, −0.91)	0.68 (0.67, 0.68)	−0.29 (−0.31, −0.26)
				Differential	…	…	0.81 (0.7, 0.92)	−0.15 (−0.24, −0.07)
	PARADIGM inputs	0.84	−1.38	Nondifferential	0.84 (0.84, 0.84)	−1.38 (−1.38, −1.37)	0.84 (0.84, 0.85)	−0.42 (−0.47, −0.38)
				Differential	…	…	1.00 (0.87, 1.14)	−0.04 (−0.34, 0.31)
Cardiovascular mortality	EMPA-REG inputs	0.62	−0.78	Nondifferential	0.66 (0.65, 0.66)	−0.45 (−0.47, −0.45)	0.67 (0.65, 0.68)	−0.19 (−0.20, −0.18)
				Differential	0.70 (0.67, 0.74)	−0.39 (−0.46, −0.32)	0.83 (0.74, 0.93)	−0.08 (−0.14, −0.03)
	PARADIGM inputs	0.80	−1.49	Nondifferential	0.83 (0.82, 0.83)	−0.76 (−0.77, −0.73)	0.83 (0.83, 0.84)	−0.30 (−0.31, −0.28)
				Differential	0.84 (0.83, 0.85)	−0.69 (−0.74, −0.65)	1.04 (0.88, 1.22)	0.05 (−0.20, 0.30)

EMPA-REG OUTCOME indicates BI 10773 (Empagliflozin) Cardiovascular Outcome Event Trial in Type 2 Diabetes Mellitus Patients; PARADIGM, prospective comparison of ARNI (angiotensin receptor–neprilysin inhibitor) with ACEI (angiotensin-converting–enzyme inhibitor) to determine impact on global mortality and Morbidity; RD, risk difference; and RR, risk ratio.

*See Tables S1 and S2 for full details on input parameters.

0.83 [0.82, 0.83] and 0.84 [0.83, 0.85], respectively, for PARADIGM inputs). Substantial bias in RDs was also noted (median [2.5th, 97.5th percentile] RDs −0.45 [−0.47, −0.45] and −0.39 [−0.46, −0.32], respectively for EMPA-REG inputs; −0.76 [−0.77, −0.73] and −0.69 [−0.74, −0.65], respectively for PARADIGM inputs). The bias in both measures was more pronounced, when using only hospitalization claims to ascertain cardiovascular mortality, especially under differential misclassification (Table 3).

DISCUSSION

In this large study of 2 cohorts of Medicare-enrolled patients with chronic conditions, we noted that approaches to identify mortality from administrative claims had variable success compared with NDI. Administrative recording of all-cause mortality in Medicare claims was near-complete; however, sources only capturing in-hospital mortality may be unable to identify a substantial proportion of all-cause deaths that occur in nonhospital settings. Further, cardiovascular-specific mortality defined from administrative claims[6,7] had poor sensitivity and PPVs. Collectively, these results suggest that investigators should anticipate the potential for bias from outcome misclassification when using claims to capture mortality.

Given an increasing emphasis on utilizing routinely collected healthcare data such as administrative claims for conducting investigations of treatment effects or for ascertaining outcomes in prospective RCTs to support regulatory decisions,[2,4] it is important to understand the implications of outcome misclassification that is common in such sources. Our findings have direct implications on comparative effectiveness research based on administrative claims data with mortality as the outcome of interest. First and foremost, we advise caution when using ad hoc approaches to determine cardiovascular-specific mortality,[6,7] because of potential misclassification. In this study, we noted that 28% and 36% of noncardiovascular deaths were incorrectly identified as cardiovascular deaths using such an approach among patients with T2D and HF, respectively. Misclassification of a substantial proportion of noncardiovascular deaths as cardiovascular could lead to bias towards the null in studies of cardiovascular treatments, as observed in our simulations, and threaten the validity of conclusions. The misclassification of cardiovascular mortality we observed in this study also does not appear to be population or data source specific; a previous study has reported sensitivity of 36.8% and PPV of 36.4% for enrollees in Medicare Advantage plans.[10] Second, when

mortality information is only available through hospitalization claims, risk difference estimates could be severely biased and RR estimates are only unbiased under the assumption of nondifferential misclassification. We recommend conducting rigorous sensitivity analyses[11] to evaluate the impact of outcome misclassification on observed treatment effects in such circumstances. Overall, to address outcome misclassification in high-stakes investigations of medication effects on mortality with potential regulatory implications, it would be important to make the process of linking various claims-based sources with national sources containing detailed mortality information such as the NDI more economical and efficient.

Our observation that Medicare claims-based capture of all-cause mortality appears to be comprehensive is encouraging. In a recent study, Strom et al were able to identify 100% of all-cause mortality events from administrative claims in patients enrolled in CoreValve HiR (US CoreValve Pivotal High Risk) SURTAVI (Surgical or Transcatheter Aortic Valve Replacement in Intermediate-Risk Patients) trials and linked to Medicare claims.[4] Further, the near-complete capture of all-cause mortality based on an administrative process also suggests that differential misclassification, which is a major threat with respect to bias, is unlikely. Therefore, use of Medicare claims to study all-cause mortality may be appropriate. However, it must be noted that bias from outcome misclassification only represents 1 source of bias in observational studies based on administrative claims, and other sources such as confounding or selection bias must be carefully considered and addressed. Another important consideration is that use of all-cause mortality as a proxy for cause-specific mortality[12] may result in an underestimation of the treatment effect for medications that are not expected to influence risk of mortality caused by other causes.

There are some important limitations of this study. First, we used NDI-recorded cause of death as the criterion standard, which may not perfectly capture the cause. For instance, in a recent validation study, the agreement between cardiovascular causes of death between NDI and an endpoints committee was noted to be 77.6%.[13] Inaccuracies in NDI-reported cause of death may lead to inaccurate estimates for performance characteristics of claims-based approaches in determining cause-specific mortality in our study. Second, we defined only 2 populations to inform ranges of performance characteristics to be used as inputs for our simulations. While inclusion of nearly 3 million deaths is a strength of our study, it is possible that performance of claims-based approaches for assessment of mortality in other disease conditions may differ. Finally, we did not evaluate the accuracy of mortality capture through alternate linkable sources such as the DMF, although we note that previous research has noted the potential for inadequate capture of mortality through these sources.[5]

In conclusion, we observed that approaches that only use hospitalization claims to identify mortality and that rely on *ICD* coding before mortality to ascertain cardiovascular cause of death have suboptimal performance characteristics, which could lead to substantial bias in treatment effect estimation. We also noted near-complete capture of all-cause mortality in Medicare claims, which may facilitate clinical investigations focused on all-cause mortality.

ARTICLE INFORMATION

Received April 4, 2020; accepted July 9, 2020.

Affiliations

From the Division of Pharmacoepidemiology and Pharmacoeconomics, Brigham and Women's Hospital & Harvard Medical School, Boston, MA.

Acknowledgments

Dr Desai had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Sources of Funding

This study was funded by Division Pharmacoepidemiology and Pharmacoeconomics, Brigham and Women's Hospital.

Disclosures

Dr Desai reports serving as Principal investigator on research grants from Vertex, Novartis, and Bayer to the Brigham and Women's Hospital for unrelated research projects. Dr Patorno is investigator of investigator-initiated grants to the Brigham and Women's Hospital from Boehringer Ingelheim and Glaxo Smith Kline, not related to the topic of the submitted work. The remaining authors have no disclosures to report. Dr Patorno was supported by a career development grant K08AG055670 from the National Institute on Aging.

Supplementary Materials

Tables S1–S2

REFERENCES

1. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J Clin Epidemiol*. 2005;58:323–337.
2. Franklin JM, Pawar A, Martin D, Glynn RJ, Levenson M, Temple R, Schneeweiss S. Nonrandomized real-world evidence to support regulatory decision-making: process for a randomized trial replication project. *Clin Pharmacol Ther*. 2020;107:817–826.
3. Patorno E, Schneeweiss S, Gopalakrishnan C, Martin D, Franklin JM. Using real-world data to predict findings of an ongoing phase iv cardiovascular outcome trial: cardiovascular safety of linagliptin versus glimepiride. *Diabetes Care*. 2019;42:2204–2210.
4. Strom JB, Faridi KF, Butala NM, Zhao Y, Tamez H, Valsdottir LR, Brennan JM, Shen C, Popma JJ, Kazi DS, et al. Use of administrative claims to assess outcomes and treatment effect in randomized clinical trials for transcatheter aortic valve replacement: findings from the extending trial-based evaluations of medical therapies using novel sources of data (EXTEND) study. *Circulation*. 2020;142:203–213.
5. Navar AM, Peterson ED, Steen DL, Wojdyla DM, Sanchez RJ, Khan I, Song X, Gold ME, Pencina MJ. Evaluation of mortality data from the social security administration death master file for clinical research. *JAMA Cardiol*. 2019;4:375–379.

6. Singh S, Fouayzi H, Anzuoni K, Goldman L, Min JY, Griffin M, Grijalva CG, Morrow JA, Whitmore CC, Leonard CE. Diagnostic algorithms for cardiovascular death in administrative claims databases: a systematic review. *Drug Saf*. 2019;42:515–527.
7. Suchard MA, Schuemie MJ, Krumholz HM, You SC, Chen R, Pratt N, Reich CG, Duke J, Madigan D, Hripcsak G. Comprehensive comparative effectiveness and safety of first-line antihypertensive drug classes: a systematic, multinational, large-scale analysis. *Lancet*. 2019;394:1816–1826.
8. Zinman B, Wanner C, Lachin JM, Fitchett D, Bluhmki E, Hantel S, Mattheus M, Devins T, Johansen OE, Woerle HJ. Empagliflozin, cardiovascular outcomes, and mortality in type 2 diabetes. *N Engl J Med*. 2015;373:2117–2128.
9. McMurray JJ, Packer M, Desai AS, Gong J, Lefkowitz MP, Rizkala AR, Rouleau JL, Shi VC, Solomon SD, Swedberg K. Angiotensin–neprilysin inhibition versus enalapril in heart failure. *N Engl J Med*. 2014;371:993–1004.
10. Suehs B, Bowe A, Davis C, Bunniran S, Uribe C. Mortality ascertainment in a medicare advantage population using health plan enrollment data. *Pharmacoepidemiol Drug Saf*. 2017;26(suppl 2): 306–307.
11. Fox MP, Lash TL, Greenland S. A method to automate probabilistic sensitivity analyses of misclassified binary variables. *Int J Epidemiol*. 2005;34:1370–1376.
12. 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*. 2010;304:411–418.
13. Quin JA, Hattler B, Shroyer ALW, Kemp D, Almassi GH, Bakaeen FG, Carr BM, Bishawi M, Collins JF, Grover FL. Concordance between administrative data and clinical review for mortality in the randomized on/off bypass follow-up study (ROOBY-FS). *J Card Surg*. 2017;32: 751–756.

SUPPLEMENTAL MATERIAL

Table S1. Summary level data from trials used as inputs for Monte-Carlo simulations.

	EMPAREG-OUTCOME trial			PARADIGM trial	
	Empagliflozin	**Placebo**		**Sacubitril-Valsartan**	**Enalapril**
Sample size	**4,687**	**2,333**		**4,187**	**4,212**
Person-years follow up	13,866	6,783		9,491	9,407
All-cause deaths	269	194		711	835
CV deaths	172	137		558	693
Incidence rate/100 py all-cause deaths	1.94	2.86		7.49	8.88
Incidence rate/100 py CV deaths	1.24	2.02		5.88	7.37

Table S2. Input parameter ranges used in the Monte Carlo simulations.

Outcome assessment appraoch	False positive rate	Sensitivity		
		Non differential misclassification	Differential misclassification	
			Treatment	Reference
1. All-place all-cause mortality: Master Beneficiary Summary File or hospitalization claims with discharge status of death	0	0.99207-0.99997	-	-
2. In-hospital all cause mortality: Hospitalization claims with discharge status of death	0	0.2771-0.3371	0.30-0.35	0.25-0.30
3. All-place cardiovascular mortality:	0.2878-0.3616	0.5201-0.5383	0.55-0.60	0.50-0.55
4. In-hospital cardiovascular mortality:	0.1375-0.1901	0.2027-0.2234	0.20-0.25	0.15-0.20