Validation of an algorithm to identify heart failure hospitalisations in patients with diabetes within the veterans health administration

Caroline A Presley,1,2 Jea Young Min,1,3 Jonathan Chipman,4 Robert A Greevy,4 Carlos G Grijalva,1,3 Marie R Griffin,1,2,3 Christianne L Roumie1,2

ABSTRACT
Objectives We aimed to validate an algorithm using both primary discharge diagnosis (International Classification of Diseases Ninth Revision (ICD-9)) and diagnosis-related group (DRG) codes to identify hospitalisations due to decompensated heart failure (HF) in a population of patients with diabetes within the Veterans Health Administration (VHA) system.

Design Validation study.

Setting Veterans Health Administration—Tennessee Valley Healthcare System

Participants We identified and reviewed a stratified, random sample of hospitalisations between 2001 and 2012 within a single VHA healthcare system of adults who received regular VHA care and were initiated on an antidiabetic medication between 2001 and 2008. We sampled 500 hospitalisations; 400 hospitalisations that fulfilled algorithm criteria, 100 that did not. Of these, 497 had adequate information for inclusion. The mean patient age was 66.1 years (SD 11.4). Majority of patients were male (98.8%); 75% were white and 20% were black.

Primary and secondary outcome measures To determine if a hospitalisation was due to HF, we performed chart abstraction using Framingham criteria as the referent standard. We calculated the positive predictive value (PPV), negative predictive value (NPV), sensitivity and specificity for the overall algorithm and each component (primary diagnosis code (ICD-9), DRG code or both).

Results The algorithm had a PPV of 89.7% (95% CI 86.8 to 92.7), NPV of 93.9% (89.1 to 98.6), sensitivity of 45.1% (25.1 to 65.1) and specificity of 99.4% (99.2 to 99.6). The PPV was highest for hospitalisations that fulfilled both the ICD-9 and DRG algorithm criteria (92.1% (89.1 to 95.1)) and lowest for hospitalisations that fulfilled only DRG algorithm criteria (62.5% (28.4 to 96.6)).

Conclusions Our algorithm, which included primary discharge diagnosis and DRG codes, demonstrated excellent PPV for identification of hospitalisations due to decompensated HF among patients with diabetes in the VHA system.

INTRODUCTION
Patients with diabetes are up to two and a half times more likely to develop heart failure (HF) than those without diabetes.1 Several mechanisms may play a role in this increased risk of HF including diabetic cardiomyopathy, as well as comorbid hypertension and atherosclerotic cardiovascular disease.2 Thiazolidinediones have been shown to increase HF risk in patients with type 2 diabetes.3 Little evidence exists on the risk of HF outcomes associated with use of common first and second-line antidiabetic medications (ie, metformin, sulfonylurea, insulin), as HF has been an infrequent primary outcome in clinical trials.4

Observational studies using administrative data are an important alternative to randomised clinical trials to evaluate the risk of HF, including hospitalisations due to decompensated HF, associated with commonly used antidiabetic treatment regimens. These studies may be limited if they identify outcomes using...
algorithms with poor diagnostic performance. To address
this limitation and minimise misclassification of outcomes,
it is necessary to validate algorithms that identify decomp-
ensated HF as the primary reason for hospital admission,
not as a pre-existing comorbidity or a complication that
developed during the course of hospitalisation.

Although algorithms to identify HF events have been vali-
dated in the Veterans Health Administration (VHA) system,
these included both inpatient and outpatient encounters
did not specifically focus on events resulting from
decompensated HF.5–7 Additionally, these algorithms only
relied on International Classification of Diseases Ninth Revi-
sion (ICD-9) codes, and few studies have examined their
performance in a high-risk population, including patients
with diabetes. An algorithm including both ICD-9 code and
disease-related group (DRG) code criteria to identify hospi-
talisations due to decompensated HF has not been tested
within VHA.8 9 Such algorithms have performed well in
academic and community health systems (positive predic-
tive value (PPV) 83%–96%).9–11 We aimed to validate an
algorithm using both primary discharge diagnosis (ICD-9)
and DRG codes to identify hospitalisations due to decom-
ensated HF in a population of patients with diabetes
within the VHA system.

METHODS
Study design
This was a validation study of an algorithm to identify
HF hospitalisations that occurred between 2001 and 2012
in the VHA’s Tennessee Valley Healthcare System (TVHS),
which includes two hospitals. This study was approved by
the TVHS Institutional Review Board. We used existing
data; a waiver of informed consent was allowed.

Study population
The underlying study population was a national obser-
vational cohort of veterans who were initiated on an
oral hypoglycaemic medication between 2001 and 2008
(n=411055); follow-up data for these veterans was available
through 2012.12 From this cohort, veterans were eligible for
inclusion if they met the following criteria: aged 18 years
or older, received regular VHA care (presence of an outpa-
tient encounter, emergency department visit, hospitalisa-
tion or medication refill at least once every 180 days) were
diagnosed with diabetes (at least one prescription filled for
an oral hypoglycaemic medication) between 2001 and 2008
and were hospitalised in TVHS between 2001 and 2012.
For this study, a patient’s diagnosis of diabetes could have
occurred before or after the included study hospitalisation
to allow adequate sampling of hospitalisations meeting
HF algorithm criteria.

Study events
The algorithm identified hospitalisations with a primary
discharge diagnosis code (ICD-9) of HF or cardiomyopathy
(425.x; 428.x; 404.01, 404.03, 404.11, 404.13, 404.91, 404.93, 398.91, 402.01, 402.11, 402.91; online supplementary appendix table A1) and/or a
DRG code for HF (127, used prior to fiscal year 2008; 291–293, used after fiscal year 2008). We sampled 500
hospitalisations from the underlying study population;
400 that met algorithm criteria (algorithm positive) and
100 that did not (algorithm negative). The 500 patients
were sampled with a 4:1 algorithm positive:negative ratio
to allow measuring PPV with greater precision. Stratified
random sampling was used to select hospitalisations from
the following strata: hospitalisations fulfilling both ICD-9
d and DRG code criteria, only ICD-9 code criteria and only
DRG code criteria, as well as, algorithm-negative hospi-
talisations. The probability of selection within strata was
used to calculate sampling weights in each stratum (ie, weights=(# of hospitalisations in the sampling strata)/
(# of hospitalisations sampled from that strata)). We
weighted observations so the stratified sample accurately
reflected the underlying study population of hospitalisa-
tions. An individual could be included in the study more
than once if they had multiple hospitalisations sampled.
The HF algorithm operates on each hospitalisation inde-
pendently, thus a random sample hospitalisations (as
opposed to patients who may have a mix of algorithm posi-
tive and negative hospitalisations over time) was needed
for unbiased estimates of the algorithm’s performance on
identifying HF in hospitalisations for this population.

Data collection
Data were abstracted from the VHA’s electronic medical
record using standardised forms by an internal medi-
cine physician, blinded to HF algorithm status. We used
the standardised Framingham criteria to classify hospi-
talisations as decompensated HF.13 The presence or
absence of symptoms, signs and radiological features of
HF were abstracted from the electronic medical record
from within the first 24 hours of the admission date to
avoid capturing signs or symptoms of HF not present on
admission. A hospitalisation met criteria for HF if it had
a minimum of two major or one major and two minor
Framingham criteria not attributable to another medical
condition (table 1).14

Additionally, we used ejection fraction (EF) data to clas-
sify HF hospitalisations as HF with reduced EF (HFrEF,
EF≤40%), HF with preserved EF (HFpEF, EF≥50%) or
borderline HFpEF (EF 41%–49%) according to Amer-
ican College of Cardiology/Cardiology Foundation/American Heart
Association guidelines.15 The EF measurement collected
during or in closest proximity (up to 1 year prior) to the
study hospitalisation was used. If multiple assessments
were present, the EF measurement from an echocardi-
ogram was used if available, followed by measurements
from cardiac catheterisation or a nuclear medicine study,
respectively. Furthermore, HF hospitalisations were clas-
sified as incident (new-onset HF) or prevalent (exacerba-
tion of chronic HF). For this, the investigator examined
the electronic medical record for the 2 years preceding
the study hospitalisation to determine if the patient had a
prior diagnosis of or hospitalisation for HF.16
et al for hospitalisations that met both ICD-9 and DRG code for each distinct ICD-9 code included in the algorithm. We calculated SEs with sampling weights. We calculated PPVs by creating 95% CIs, a Taylor series linearisation was used to analyse the performance of the algorithm in the underlying study population of TVHS hospitalisations. To assess algorithm performance when not restricted to primary discharge diagnoses, we examined algorithm-negative hospitalisations containing an HF or cardiomyopathy code in any of the four non-primary discharge diagnosis code positions. For this sensitivity analysis, we reclassified these algorithm-negative hospitalisations as algorithm-positive hospitalisations, and using weighted analysis, calculated the PPV, NPV, sensitivity and specificity for this alternate algorithm.

Statistical analyses were performed using Stata Statistical Software V.14 (StataCorp LP).

RESULTS

Of 10 766 eligible hospitalisations in TVHS between 2001 and 2012, a total of 500 hospitalisations were sampled. Of the 500 sampled hospitalisations, 324 unique patients were represented only once (ie, contributed only one hospitalisation for review); the remaining 176 hospitalisations were from patients who contributed more than 1 hospitalisation (range 2–9). Of the algorithm-positive hospitalisations, 83% fulfilled both ICD-9 and DRG code criteria, 15% met ICD-9 code criteria only and 1% met DRG code criteria only. Of sampled hospitalisations, three had insufficient documentation to assess Framingham criteria for decompensated HF. Of these, 497 hospitalisations were included.

The patients were on average 66.1 years old (SD 11.4) with a median age of 65 (IQR 58, 75) (table 2). Patients were overwhelmingly male (98.8%); 75% were white and 20% were black. There was a high prevalence of hypertension (83.7%), hyperlipidaemia (58.8%), atherosclerotic cardiovascular disease (61.8%) and chronic kidney disease (stage 3 and higher, 41.5%). In this sample, 430 of 497 patients (86.5%) of patients had a diagnosis of type 2 diabetes at the time of study hospitalisation. Mean haemoglobin A1c was 6.96% (SD 1.6).

Of 497 hospitalisations reviewed, 360 (72.4%) fulfilled Framingham criteria for decompensated HF. Of these 360, 127 (35.3%) were incident HF events, 229 (63.6%) were prevalent events and four (1.1%) had insufficient documentation for this determination. Additionally, 186 of the 360 HF hospitalisations (51.7%) were classified as HFpEF, 86 (23.9%) were HFrEF, 36 (10.0%) were HfPEF borderline and 52 (14.4%) did not have EF data.

Covariates

Data on multiple covariate measures were collected from VHA data for the 730 days preceding the study hospitalisation. For Medicare or Medicaid enrollees, we obtained enrolment, claims files and prescription (part D) data. Covariate measures included age, sex, race, presence of medical comorbidities, body mass index and laboratory values (haemoglobin A1c, estimated glomerular filtration rate).

Table 1  Framingham criteria for HF, the reference standard for classification of hospitalisations

| Major criteria                                           | Minor criteria                  |
|----------------------------------------------------------|---------------------------------|
| Paroxysmal nocturnal dyspnoea or orthopnoea             | Night cough                     |
| Elevated jugular venous pressure                         | Dyspnoea with exertion          |
| HF treatment-related 10 lbs weight loss in preceding 5 days | Non-HF treatment-related 10 lbs weight loss in preceding 5 days |
| S3 gallop                                                | Hepatomegaly                    |
| Hepatojugular reflex                                    | Bilateral ankle oedema          |
| Rales, crackles                                          | Pleural effusion (on imaging)   |
| Cardiomegaly (on imaging)                               | Pulmonary vascular engorgement (on imaging) |
| Pulmonary oedema (on imaging)                            | Tachycardia (heart rate > 120 bpm) |

*A hospitalisation was classified as HF if it met a minimum of two major or one major and two minor criteria. HF, heart failure.

Statistical analysis

Descriptive statistics were used to characterise the study sample and hospitalisations including type of HF and incident or prevalent classification for confirmed HF hospitalisations.

Using the chart review classification based on Framingham criteria as the reference standard, we calculated the PPV (proportion of algorithm-positive cases confirmed as HF) for the overall algorithm and each component (primary diagnosis code (ICD-9), DRG code or both). Chart review classifications for each hospitalisation were treated as statistically independent, as they were determined using only data collected from each discrete hospitalisation. We also calculated the NPV (proportion of algorithm-negative cases confirmed as non-HF), sensitivity (proportion of HF hospitalisations correctly identified by the algorithm) and specificity (proportion of non-HF hospitalisations correctly identified by the algorithm). We included sampling weights in the analysis to reflect the performance of the algorithm in the underlying study population of TVHS hospitalisations. To create 95% CIs, a Taylor series linearisation was used to calculate SEs with sampling weights. We calculated PPVs for each distinct ICD-9 code included in the algorithm for hospitalisations that met both ICD-9 and DRG code criteria, as well as for hospitalisations that fulfilled only ICD-9 code criteria. Each of these was done within a given sampling stratum; sampling weights were not needed. Wilson’s formula for proportions was used to calculate 95% CIs due to smaller sample sizes.

We performed subgroup analyses to determine the performance of the algorithm in subsets of the sample including hospitalisations in which the patient had a diagnosis of diabetes prior to or at the time of hospitalisation, as well as comparing hospitalisations prior to fiscal year 2008 and after 2008 when the DRG codes for HF changed. Additionally, up to five discharge diagnosis codes (ICD-9 codes) were available for each hospitalisation. To assess algorithm performance when not restricted to primary discharge diagnoses, we examined algorithm-negative hospitalisations containing an HF or cardiomyopathy code in any of the four non-primary discharge diagnosis code positions. For this sensitivity analysis, we reclassified these algorithm-negative hospitalisations as algorithm-positive hospitalisations, and using weighted analysis, calculated the PPV, NPV, sensitivity and specificity for this alternate algorithm.

Statistical analyses were performed using Stata Statistical Software V.14 (StataCorp LP).
Open Access

available. Of patients who had a confirmed HF hospitalisation and available EF data, 172 of 308 (55.8%) patients had their EF assessed during the study hospitalisation; the remainder had an assessment of EF during the year prior to the study hospitalisation.

Overall, we found 354 true positive hospitalisations due to HF, 45 false positives, 6 false negatives and 92 true negatives. Of the six HF algorithm-negative hospitalisations that fulfilled Framingham criteria, four had an HF or cardiomyopathy ICD-9 code listed among their four non-primary discharge diagnosis codes, but not in the algorithm-targeted primary discharge diagnosis position. Primary discharge diagnosis codes in these four hospitalisations included: subendocardial infarction, initial episode of care; diabetes with ophthalmic manifestations, type 2 or unspecified type, uncontrolled; anxiety state, unspecified and atrioventricular block, complete. Primary discharge diagnosis codes for the two hospitalisations that did not include a HF or cardiomyopathy ICD-9 code among their discharge diagnosis codes were atherosclerotic heart disease of native coronary artery without angina pectoris and chest pain unspecified, respectively.

In weighted analysis reflecting algorithm performance in the underlying study population, the overall algorithm had a PPV of 89.7% (95% CI, 86.8 to 92.7) and NPV of 93.9% (89.1, 98.6) (table 3). The sensitivity was 45.1% (25.1, 65.1) and specificity was 99.4% (99.2, 99.6). For hospitalisations that fulfilled both ICD-9 and DRG criteria, the algorithm had a PPV of 92.1% (89.1, 95.1) with a sensitivity of 41.3% (21.6, 61.0) (table 4). For hospitalisations that fulfilled only ICD-9 or DRG criteria, the algorithm had a PPV of 79.3% (70.7, 87.9) and 62.5% (28.4, 96.6), respectively.

To evaluate the performance of specific ICD-9 codes, we calculated the PPV for hospitalisations with different ICD-9 primary discharge diagnosis codes. The PPV of the algorithm limited to hospitalisations with 428.x codes (HF) that fulfilled both ICD-9 and DRG code criteria was highest, 92.8% (89.3, 95.3) (see online supplementary appendix table A1). For hospitalisations with 428.x codes that only fulfilled ICD-9 code criteria, PPV was

| Table 3 | PPV and NPV, sensitivity, specificity for overall HF hospitalisation identification algorithm,* weighted analysis |
|---------|-------------------------------------------------------------------------------------------------------------|
|         | Confirmed HF hospitalisation, sum weight† (n)‡ | Confirmed non-HF hospitalisation, sum weight (n) | Total hospitalisations, sum weight (n) | Performance metric (95% CI)§ |
| HF algorithm positive | 513 (354) | 59 (45) | 572 (399) | PPV 89.7 (86.8 to 92.7) |
| HF algorithm negative | 624 (6) | 9570 (92) | 10194 (98) | NPV 93.9 (89.1 to 98.6) |
| Total | 1138 (360) | 9628 (137) | 10766 (497) |
| Validity measure | Sensitivity (95% CI) | Specificity (95% CI) | 45.1 (25.1 to 65.1) | 99.4 (99.2 to 99.6) |

*The HF algorithm consisted of a primary discharge diagnosis ICD-9 code 425.X, 428.X, 402.01, 402.11, 402.91, 404.01, 404.03, 404.11, 404.13, 404.91, 404.93 or 398.91 and/or a DRG code 127 or 291–293.
†Sum weight represents the number of hospitalisations from the overall study population that would have fallen into each category when inverse probability of sampling weights were applied to the study sample.
‡n represents the actual number of charts reviewed that were true positives, false positives, false negatives or true negatives in each category.
§To create 95% CIs, we used a Taylor series linearisation to calculate SEs with sampling weights.
85.3% (75.0, 91.8). For hospitalisations with ICD-9 code of 402.x (hypertensive heart disease with HF), the PPV of the algorithm was 83.3% (43.6, 97.0) for both hospitalisations that met both ICD-9 and DRG code criteria and for those that only fulfilled ICD-9 code criteria. The algorithm had the poorest performance for hospitalisations with a primary discharge diagnosis code of 404.x (hypertensive heart disease and chronic kidney disease with HF) or 425.x (cardiomyopathy). The PPV was 50.0% (15.0, 85.0) for hospitalisations with a 404.x code that met both ICD-9 and DRG code criteria and 0% (0, 79.3) for hospitalisations with 404.x code that met only ICD-9 criteria. In our sample, no hospitalisations with an ICD-9 code of 425.x met both ICD-9 and DRG code criteria. The PPV for hospitalisations with a 425.x code that met only ICD-9 code criteria was 50.0% (25.4, 74.6).

**Subgroup analyses**

Performance of the algorithm was similar when restricted to patients (n=430) who had a diagnosis of diabetes at the time of their study hospitalisation, PPV 90.2% (87.2, 93.3). Additionally, the PPVs were comparable for the periods when different DRG codes were used; PPV was 90.4% (86.6, 94.2) for DRG 127 (prior to fiscal year 2008) and 88.9% (84.3, 93.6) for DRG 291–293 (after fiscal year 2008).

**Sensitivity analyses**

To determine the performance of an algorithm with broader discharge diagnosis code criteria, we calculated the PPV, NPV, sensitivity and specificity of an alternate algorithm that allowed ICD-9 criteria to be present in any of the first five discharge diagnosis code positions. In total, 16 hospitalisations were reclassified as algorithm-positive hospitalisations using this alternate algorithm. Of these, 4 hospitalisations were confirmed HF hospitalisations by chart review (events discussed above) and 12 hospitalisations were confirmed non-HF hospitalisations. This alternate algorithm had higher sensitivity, 81.7% (59.9, 100.0) versus 45.1% (25.1, 65.1), but had poor PPV, 41.6% (24.5, 58.6) versus 89.7% (86.8, 92.7) and lower specificity, 86.4% (79.6, 93.3) versus 99.4% (99.2, 99.6) compared with the original HF hospitalisation study algorithm (see online supplementary appendix table A2).

**DISCUSSION**

Our algorithm to identify hospitalisations due to decompensated HF in a sample of veterans with diabetes used both primary discharge diagnosis and DRG codes and demonstrated high PPV (89.7%), NPV (93.9%), specificity (99.4%), though the sensitivity was only 45.1%. This algorithm has comparable PPV to prior studies conducted in non-VHA populations that validated algorithms based on both ICD-9 and DRG code criteria (PPV 83%–96%). Our algorithm has slightly lower PPV compared with the study in non-VHA patients with diabetes receiving care in an integrated managed care system (PPV 97%), likely because the study by Iribarren et al included only the codes 428.x and 402.x ICD-9 codes which were highly specific in our study. Our study complements findings from previous studies, as we applied a weighting strategy which provides information about the performance of the algorithm in the underlying study population and calculated sensitivity, specificity and NPV for the algorithm due to the inclusion of algorithm-negative hospitalisations.

Our algorithm, which focused on primary diagnoses, has a good PPV (89.7%), is highly specific (99.4%) but has poor sensitivity (45.1%). Another study conducted within VHA by Floyd et al reported a 90% sensitivity for their algorithm in identifying chronic (prevalent) HF based on the presence of an ICD-9 code for HF recorded in the inpatient or outpatient setting in the preceding 12 to 24 months. We believe the lower sensitivity in our study is due to the stringent criteria for our HF algorithm, namely presence of an ICD-9 code for HF as the primary diagnosis code and/or a DRG code for HF and rigorous use of the Framingham criteria to adjudicate potential HF events. We found that an alternate, expanded
algorithm that included all available diagnoses, was more sensitive (81.7%) but had lower PPV (41.6%) and specificity (86.4%). The more specific algorithm may be more appropriate in comparative effectiveness studies of HF as an outcome for antidiabetic medications. In these studies, high-specificity outcome definitions help minimise the impact of outcome misclassification when the relative risks of events are calculated among different medication exposures. Our study algorithm has good discriminatory ability in that hospitalisations selected as algorithm positive are very likely due to a true HF hospitalisation. An algorithm with higher sensitivity may be more appropriate if one is seeking to capture HF as a comorbidity and adequately account for potential confounding between exposure groups. Broader discharge diagnosis code criteria may be more appropriate when the objective is to identify as many potential events as possible.

Our study adds to the evidence from prior studies because we validated an algorithm that included both ICD-9 and/or DRG criteria and assessed the performance of individual components of the algorithm. Our algorithm demonstrated higher PPV when limited to hospitalisations that fulfilled both the primary discharge diagnosis code and DRG code criteria and had the lowest PPV for hospitalisations fulfilling only DRG code criteria. The algorithm has the lowest risk for misclassification of outcomes when primary discharge diagnosis and DRG codes are aligned and the highest risk when these are not aligned. Additionally, given that DRG only cases are rare and have poor PPV, it may not be necessary or appropriate to include this component in an algorithm to identify HF hospitalisations.

 Previously validated algorithms have most commonly included criteria of ICD-9 code 428.x in the primary discharge diagnosis position without DRG code criteria and have demonstrated PPV of 84% to 100%.[13-19,21] Algorithms including additional ICD-9 codes have shown varying performance with PPV ranging from 77% to 99%.[20,21-24] By including multiple ICD-9 codes in our algorithm, we were able to compare PPVs for individual ICD-9 codes. The algorithm performed best for hospitalisations with ICD-9 code 428.x and had lowest PPV for ICD-9 codes 404.x and 425.x, although the number of hospitalisations with the latter two codes was limited. While we did not evaluate an algorithm that included ICD-10 codes, our data suggest that I50.x (HF) and I11.0 (HF due to hypertension), which correspond to the 428.x and 402.x ICD-9 codes, will perform best to identify HF hospitalisations.

**Strengths**

Our study has important strengths. We applied a sampling strategy that allowed weighted estimations to extrapolate findings to our underlying study population, and unlike some studies that have only reported PPVs, we performed a complete validation assessment. We also used standardised Framingham HF criteria for our adjudications and complemented those data with HF classifications based on EF and disease onset information.

**Limitations**

Our study has some limitations. Data abstraction by chart review may be subject to error due to low quality or missing information. We tried to minimise this potential issue by using a standardised abstraction process. However, we did not calculate the reliability of our reviews. This study was limited to a sample of hospitalisations within VHA healthcare system and the sample was predominantly older men, which may limit the generalisability of the study findings to other settings. Additionally, misclassification of HF hospitalisations by EF may exist as we used EF assessments from up to 1 year prior to the study hospitalisation, though 55.8% of assessments were completed during the study hospitalisation.

**Implications**

The validation of this algorithm will facilitate future study of the risk of HF hospitalisations in VHA patients with diabetes, especially in comparative effectiveness studies. Our algorithm demonstrated a very good PPV and specificity and can be used to identify important HF outcomes in the study of antidiabetic medications in the VHA population.

**Contributors**

All authors listed have contributed sufficiently to the project to be included as authors, and all those who are qualified to be authors are listed in the author byline. CAP, CLR, JYM, CGG and MRG: design of the study. CAP and RAG: collection of data. CAP, CLR, JC and RAG: analysis or interpretation of data. CAP and CLR: drafting of the manuscript. JC, RAG, JYM, CGG and MRG: critical revision of the manuscript. All authors contributed to the final approval of the manuscript for submission.

**Funding**

This project was funded by the by VA Clinical Science research and development investigator initiated grant CX000570-06 (Roumie). CLR was supported in part by Center for Diabetes Translation Research P30DK092986. JYM was supported by the Clinical and Translational Science Award (CTSA) No. TL1TR000447-09 from the National Center for Advancing Translational Sciences. CAP and JYM were supported by the Office of Academic Affiliations VA Quality Scholars Program.

**Competing interests**

None declared.

**Ethics approval**

VA—Valley Healthcare System Institutional Review Board.

**Provenance and peer review**

Not commissioned; externally peer reviewed.

**Data sharing statement**

No additional data available.

**Open Access**

This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2018. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

**REFERENCES**

1. Nichols GA, Gullion CM, Koro CE, et al. The incidence of congestive heart failure in type 2 diabetes: an update. *Diabetes Care* 2004;27:1879–84.
2. Iribarren C, Karter AJ, Go AS, et al. Glycemic control and heart failure among adult patients with diabetes. Circulation 2001;103:2668–73.
3. Singh S, Loke YK, Furberg CD. Thiazolidinediones and heart failure: a teleo-analysis. Diabetes Care 2007;30:2148–53.
4. Palmer SC, Mavridis D, Nicolucci A, et al. Comparison of clinical outcomes and adverse events associated with glucose-lowering drugs in patients with type 2 diabetes: a meta-analysis. JAMA 2016;316:313–24.
5. Floyd JS, Blvdon M, Moore KP, et al. Validation of methods for assessing cardiovascular disease using electronic health data in a cohort of Veterans with diabetes. Pharmacoepidemiol Drug Saf 2016;25:467–71.
6. 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;19:201–8.
7. Brophy MT, Snyder KE, Gaehde S, et al. Anticoagulant use for atrial fibrillation in the elderly. J Am Geriatr Soc 2004;52:1151–6.
8. Floyd JS, Wellman R, Fuller S, et al. Use of electronic health data to estimate heart failure events in a population-based cohort with CKD. Curr J Am Soc Nephrol 2016;11:1954–61.
9. McCullough PA, Philbin EF, Spertus JA, et al. Confirmation of a heart failure epidemic: findings from the Resource Utilization Among Congestive Heart Failure (REACH) study. J Am Coll Cardiol 2002;39:60–9.
10. Owan TE, Hodge DO, Herges RM, et al. Trends in prevalence and outcome of heart failure with preserved ejection fraction. N Engl J Med 2006;355:251–9.
11. Philbin EF, Rocco TA, Lynch JJ, et al. Predictors and determinants of hospital length of stay in congestive heart failure in ten community hospitals. J Heart Lung Transplant 1997;16:548–55.
12. Roumie CL, Hung AM, Greely RA, et al. Comparative effectiveness of sulfonylurea and metformin monotherapy on cardiovascular events in type 2 diabetes mellitus: a cohort study. Ann Intern Med 2012;157:601–10.
13. Saczynski JS, Andrade SE, Harrold LR, et al. A systematic review of validated methods for identifying heart failure using administrative data. Pharmacoepidemiol Drug Saf 2012;21 Suppl 1:129–40.
14. McKee PA, Castelli WP, McNamara PM, et al. The natural history of congestive heart failure: the Framingham study. N Engl J Med 1971;285:1441–6.
15. Yancy CW, Jessup M, Bozkurt B, et al. 2013 ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. J Am Coll Cardiol 2013;62:e147–239.
16. Camplain R, Kucharska-Newton A, Cuthbertson CC, et al. Misclassification of incident hospitalized and outpatient heart failure in administrative claims data: the Atherosclerosis Risk in Communities (ARIC) study. Pharmacoepidemiol Drug Saf 2017;26:421–8.
17. Lumley T. Analysis of complex survey samples. J Stat Softw 2004;9:18.
18. Brown LD, Cai TT, DasGupta A. Interval estimation for a binomial proportion. Statistical Science 2001;16:101–17.
19. Grijalva CG, Chung CP, Stein CM, et al. Computerized definitions showed high positive predictive values for identifying hospitalizations for congestive heart failure and selected infections in Medicaid enrollees with rheumatoid arthritis. Pharmacoepidemiol Drug Saf 2008;17:890–5.
20. Goff DC, Pandey DK, Chan FA, et al. Congestive heart failure in the United States: is there more than meets the I(CD code)? The Corpus Christi Heart Project. Arch Intern Med 2000;160:197–202.
21. Lee DS, Donovan L, Austin PC, et al. Comparison of coding of heart failure and comorbidities in administrative and clinical data for use in outcomes research. Med Care 2005;43:182–8.
22. Havranek EP, Masoudi FA, Westfall KA, et al. Spectrum of heart failure in older patients: results from the National Heart Failure project. Am Heart J 2002;143:412–7.
23. Go AS, Lee WY, Yang J, et al. Statin therapy and risks for death and hospitalization in chronic heart failure. JAMA 2006;296:2105–11.
24. Philbin EF, DISalvo TG. Influence of race and gender on care process, resource use, and hospital-based outcomes in congestive heart failure. Am J Cardiol 1998;82:76–81.