BMJ Open

Validation of an algorithm to identify heart failure hospitalizations in patients with diabetes within the Veterans Health Administration

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-020455
Article Type:	Research
Date Submitted by the Author:	07-Nov-2017
Complete List of Authors:	Presley, Caroline; VA - Tennessee Valley Healthcare System, Geriatric Research Education and Clinical Center; Vanderbilt University Medical Center, Medicine Min, Jea Young ; VA - Tennessee Valley Healthcare System, Geriatric Research Education and Clinical Center; Vanderbilt University Medical Center, Health Policy Chipman, Jonathan; Vanderbilt University Medical Center, Biostatistics Greevy, Robert; Vanderbilt University Medical Center, Biostatistics Grijalva, Carlos; VA - Tennessee Valley Healthcare System, Geriatric Research Education and Clinical Center; Vanderbilt University Medical Center, Health Policy Griffin, Marie R; VA - Tennessee Valley Healthcare System, Geriatric Research Education and Clinical Center; Vanderbilt University Medical Center, Medicine Roumie, Christianne L.; VA - Tennessee Valley Healthcare System, Geriatric Research Education and Clinical Center; Vanderbilt University Medical Center, Medicine
Primary Subject Heading:	Cardiovascular medicine
Secondary Subject Heading:	Epidemiology
Keywords:	Validation study, Pharmacoepidemiology, General diabetes < DIABETES & ENDOCRINOLOGY, Heart failure < CARDIOLOGY

Validation of an algorithm to identify heart failure hospitalizations in patients with diabetes within the Veterans Health Administration

Caroline Presley, M.D., MPH[1,2]; Jea Young Min, Pharm.D., M.P.H[1]; Jonathan Chipman, M.S.[3]; Robert A. Greevy, Jr, PhD[3]; Carlos G. Grijalva, M.D., M.P.H.[1,4]; Marie R. Griffin, M.D., M.P.H.[1,2,4]; Christianne L. Roumie, M.D., M.P.H.[1,2]

1. Veterans Health Administration-Tennessee Valley Healthcare System Geriatric Research Education Clinical Center (GRECC), HSR&D Center, Nashville, TN
2. Department of Medicine, Vanderbilt University Medical Center, Nashville, TN.
3. Department of Biostatistics, Vanderbilt University Medical Center, Nashville TN
4. Department of Health Policy, Vanderbilt University Medical Center, Nashville, TN.

Corresponding author: Caroline Presley, M.D., M.P.H.

1310 24th Avenue South, GRECC, Nashville, TN 37212

Phone: 615-873-8012; Fax: 615-873-7981

Caroline.a.presley@vanderbilt.edu; caroline.presley@va.gov

Word count: 2995

Abstract

Objectives: We aimed to validate an algorithm using both primary discharge diagnosis (ICD-9) and diagnosis-related group (DRG) codes to identify hospitalizations due to decompensated heart failure in a population of patients with diabetes within the Veterans Health Administration system.

Design: Validation study

Setting: Veterans Health Administration - Tennessee Valley Healthcare System

Participants: We identified and reviewed a stratified, random sample of hospitalizations within Veterans Health Administration. We sampled 500 hospitalizations; 400 hospitalizations that fulfilled algorithm criteria, 100 that did not. Of these, 497 had adequate information for inclusion. The mean patient age was 66.1 years (Standard deviation 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 hospitalization was due to heart failure, we performed chart abstraction using Framingham criteria as the referent standard. We calculated the positive predictive value (PPV), negative predictive value, sensitivity, and specificity for the overall algorithm and each component (primary diagnosis code [ICD-9], DRG code, or both).

Results: The algorithm had a positive predictive value of 89.7% (95% confidence interval: 86.8, 92.7), negative predictive value of 93.9% (89.1, 98.6), sensitivity of 45.1% (25.1, 65.1), and specificity of 99.4% (99.2, 99.6). The PPV was highest for hospitalizations that fulfilled both the ICD-9 and DRG algorithm criteria (92.1% [89.1, 95.1]), and lowest for hospitalizations that fulfilled only DRG algorithm criteria (62.5% [28.4, 96.6]).

Conclusions: Our algorithm, which included primary discharge diagnosis and diagnosis-related group codes, demonstrated excellent positive predictive value for identification of hospitalizations due to decompensated heart failure among patients with diabetes in the Veterans Health Administration system.

Strengths and Limitations of this Study

- This is the first study to validate an algorithm using both primary discharge diagnosis (ICD-9) and diagnosis-related group (DRG) codes to identify hospitalizations due to decompensated heart failure within the Veterans Health Administration system.
- We applied a sampling strategy that allowed weighted estimations to extrapolate findings to our underlying study population.
- We used standardized Framingham heart failure criteria for our adjudications; we performed a complete validation assessment, contrasted with other studies that have only reported positive predictive values.
- Study limitations include potentially limited generalizability of findings to other settings, and data abstraction by chart review may be subject to error.
- The validation of this algorithm will facilitate future study of the risk of heart failure hospitalizations associated with antidiabetic medication regimens in Veterans Health Administration patients with diabetes, especially in comparative effectiveness studies.

Introduction

Patients with diabetes are up to two and a half times more likely to develop heart failure than those without diabetes.[1] Several mechanisms may play a role in this increased risk of heart failure including diabetic cardiomyopathy, as well as co-morbid hypertension and atherosclerotic cardiovascular disease.[2] Thiazolidinediones have been shown to increase heart failure risk in patients with type 2 diabetes (T2DM).[3] Little evidence exists on the risk of heart failure outcomes associated with use of common first and second line antidiabetic medications (i.e. metformin, sulfonylurea, insulin), as heart failure has been an infrequent primary outcome in clinical trials.[4]

Observational studies using administrative data are an important alternative to randomized clinical trials to evaluate the risk of heart failure, including hospitalizations due to decompensated heart failure, 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 minimize misclassification of outcomes, it is necessary to validate algorithms that identify decompensated heart failure as the primary reason for hospital admission, not as a preexisting comorbidity or a complication that developed during the course of hospitalization.

Although algorithms to identify heart failure events have been validated in the Veterans Health Administration (VHA) system, these included both inpatient and outpatient encounters and did not specifically focus on events resulting from decompensated heart failure.[5-7] Additionally, these algorithms only relied on International Classification of Diseases, 9th revision [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 hospitalizations due to decompensated heart failure has not been tested within VHA.[2,8] Such algorithms have performed well in academic and community health systems (PPV 83-96%).[9-11] We aimed to validate an algorithm using both

primary discharge diagnosis (ICD-9) and DRG codes to identify hospitalizations due to decompensated heart failure in a population of patients with diabetes within the VHA system.

Methods

Study Design

This was a validation study of an algorithm to identify heart failure hospitalizations 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

A national observational cohort of Veterans with diabetes comprised the underlying study population. 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 a prescription fill or visit at least once every 180 days), were diagnosed with diabetes (at least one prescription filled for an antidiabetic medication) between 2001 and 2008, and were hospitalized in TVHS between 2001 and 2012. For this study, a patient's diagnosis of diabetes could have occurred before or after the included study hospitalization to allow adequate sampling of hospitalizations meeting heart failure algorithm criteria.

Study events

The algorithm identified hospitalizations with a primary discharge diagnosis code (ICD-9) of heart failure 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, Appendix Table A1), and/or a diagnosis-related group (DRG) code for heart failure (127, used prior to fiscal year 2008; 291-293, used after fiscal year 2008). We sampled 500 hospitalizations from the underlying study population; 400 that met algorithm criteria (algorithm-positive) and 100 that did not (algorithm-negative). Stratified random sampling was used to select hospitalizations from the following strata: hospitalizations fulfilling both ICD-9 and DRG code criteria, only ICD-9 code criteria, and only DRG code criteria,

as well as, algorithm-negative hospitalizations. The probability of selection within strata was used to calculate sampling weights in each stratum (i.e. weights = (# of hospitalizations in the sampling strata) / (# of hospitalizations sampled from that strata)). We weighted observations so the stratified sample accurately reflected the underlying study population of hospitalizations. An individual could be included in the study more than once if they had multiple hospitalizations sampled.

Data collection

Data were abstracted from the VHA's electronic medical record using standardized forms by an Internal Medicine physician, blinded to heart failure algorithm status. We used the standardized Framingham criteria, to classify hospitalizations as decompensated heart failure.[12] The presence or absence of symptoms, signs, and radiologic features of heart failure were abstracted from the electronic medical record from within the first 24 hours of the admission date to avoid capturing signs or symptoms of heart failure not present upon admission. A hospitalization met criteria for heart failure if it had a minimum of two major or one major and two minor Framingham criteria, not attributable to another medical condition (Table 1).[13]

Additionally, we used ejection fraction (EF) data to classify heart failure hospitalizations as heart failure with reduced ejection fraction (HFrEF, EF ≤ 40%), heart failure with preserved ejection fraction (HFpEF, EF ≥ 50%), or borderline HFpEF (EF 41-49%) according to American College of Cardiology Foundation/American Heart Association guidelines.[14] The ejection fraction measurement collected during or in closest proximity (up to one year prior) to the study hospitalization was used. If multiple assessments were present, the ejection fraction measurement from an echocardiogram was used if available, followed by measurements from cardiac catheterization or a nuclear medicine study, respectively. Furthermore, heart failure hospitalizations were classified as incident (new-onset heart failure) or prevalent (exacerbation of chronic heart failure). For this, the investigator examined the electronic medical record for the

two years preceding the study hospitalization to determine if the patient had a prior diagnosis of or hospitalization for heart failure.[15]

Covariates

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

Statistical Analysis

Descriptive statistics were used to characterize the study sample and hospitalizations including type of heart failure and incident or prevalent classification for confirmed heart failure hospitalizations.

Using the chart review classification based on Framingham criteria as the reference standard, we calculated the positive predictive value (PPV, proportion of algorithm-positive cases confirmed as heart failure) for the overall algorithm and each component (primary diagnosis code [ICD-9], DRG code, or both). Chart review classifications for each hospitalization were treated as statistically independent, as they were determined using only data collected from each discrete hospitalization. We also calculated the negative predictive value (NPV, proportion of algorithm-negative cases confirmed as non-heart failure), sensitivity (proportion of heart failure hospitalizations correctly identified by the algorithm), and specificity (proportion of non-heart failure hospitalizations 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 hospitalizations. To create 95% confidence intervals, a Taylor Series linearization was used to calculate standard errors with sampling weights.[16] We calculated positive predictive values for each distinct ICD-9 code included in the algorithm for hospitalizations that met both ICD-9 and DRG code criteria, as well as, for hospitalizations 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% confidence intervals due to smaller sample sizes.[17]

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

Statistical analyses were performed using Stata Statistical Software: Release 14, College Station, TX: StataCorp LP.

Results

Of 10,766 eligible hospitalizations in TVHS between 2001 and 2012, a total of 500 hospitalizations were sampled. Of the algorithm-positive hospitalizations, 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 hospitalizations, three had insufficient documentation to assess Framingham criteria (one algorithm-positive, two algorithm-negative); thus, 497 hospitalizations were included.

The majority of the patients were aged 65 years or older with a mean age of 66.1 years (Standard deviation [SD] 11.4), Table 2. Patients were overwhelmingly male (98.8%); 75% were white and 20% were black. There was a high prevalence of hypertension (83.7%), hyperlipidemia (58.8%), atherosclerotic cardiovascular disease (61.8%), and chronic kidney

disease (stage 3 and higher, 41.5%). In this sample, 87% of patients had a diagnosis of type 2 diabetes at the time of study hospitalization. Mean hemoglobin A1c was 6.96% (SD 1.6).

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

Overall, we found 354 true positive hospitalizations due to heart failure, 45 false positives, six false negatives, and 92 true negatives. Of the six heart failure algorithm-negative hospitalizations that fulfilled Framingham criteria, four had a heart failure 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 hospitalizations included: subendocardial infarction, initial episode of care; diabetes with ophthalmic manifestations, type II or unspecified type, uncontrolled; anxiety state, unspecified; and atrioventricular block, complete. Primary discharge diagnosis codes for the two hospitalizations that did not include a heart failure 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% confidence interval, 86.8, 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 hospitalizations 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 hospitalizations 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 hospitalizations with different ICD-9 primary discharge diagnosis codes. The PPV of the algorithm limited to hospitalizations with 428.x codes (Heart failure) that fulfilled both ICD-9 and DRG code criteria was highest, 92.8% (89.3, 95.3), Appendix Table A1. For hospitalizations with 428.x codes that only fulfilled ICD-9 code criteria, PPV was 85.3% (75.0, 91.8). For hospitalizations with ICD-9 code of 402.x (Hypertensive heart disease with heart failure), the PPV of the algorithm was 83.3% (43.6, 97.0) for both hospitalizations 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 hospitalizations with a primary discharge diagnosis code of 404.x (Hypertensive heart disease and chronic kidney disease with heart failure) or 425.x (Cardiomyopathy). The PPV was 50.0% (15.0, 85.0) for hospitalizations with a 404.x code that met both ICD-9 and DRG code criteria and 0% (0, 79.3) for hospitalizations with 404.x code that met only ICD-9 criteria. In our sample, no hospitalizations with an ICD-9 code of 425.x met both ICD-9 and DRG code criteria. The PPV for hospitalizations 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 hospitalization, 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 hospitalizations were reclassified as algorithm-positive hospitalizations using this

alternate algorithm. Of these, four hospitalizations were confirmed heart failure hospitalizations by chart review (events discussed above), and 12 hospitalizations were confirmed non-heart failure hospitalizations. This alternate algorithm had higher sensitivity, 81.7% (59.9, 100.0) vs. 41.5% (25.1, 65.1), but had poor PPV, 41.6% (24.5, 58.6) vs. 89.7% (86.8, 92.7), and lower specificity, 86.4% (79.6, 93.3) vs. 99.4% (99.2, 99.6), compared with the original heart failure hospitalization study algorithm, Appendix Table A2.

Discussion

Our algorithm to identify hospitalizations due to decompensated heart failure in a sample of Veterans with diabetes used both primary discharge diagnosis and diagnosis-related group codes and demonstrated high positive predictive value (89.7%), negative predictive value (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%).[9-11] 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.[2] 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 hospitalizations.

Our algorithm, which focused on primary diagnoses, has a good PPV (89.7%), is highly specific (99.4%), but has poor sensitivity (45.1%); while, an alternate 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 heart failure as an outcome for antidiabetic medications. In these studies, high specificity outcome definitions help minimize 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 hospitalizations selected as algorithm-positive are very likely due to a true heart failure hospitalization.

An algorithm with higher sensitivity may be more appropriate if one is seeking to capture heart failure as a co-morbidity 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 hospitalizations that fulfilled both the primary discharge diagnosis code and DRG code criteria, and had the lowest PPV for hospitalizations 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 heart failure hospitalizations.

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%.[12,18-20] Algorithms including additional ICD-9 codes have shown varying performance with PPV ranging from 77 to 99%.[19,21-23] By including multiple ICD-9 codes in our algorithm, we were able to compare positive predictive values for individual ICD-9 codes. The algorithm performed best for hospitalizations with ICD-9 code 428.x and had lowest PPV for ICD-9 codes 404.x and 425.x, although the number of hospitalizations with the latter two codes was limited. While we did not evaluate an algorithm that included ICD-10 codes, our data suggests that I50.x (Heart failure) and I11.0 (Heart failure due to hypertension), which

correspond to the 428.x and 402.x ICD-9 codes, will perform best to identify heart failure hospitalizations.

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 standardized Framingham heart failure criteria for our adjudications, and complemented those data with heart failure classifications based on ejection fraction 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 minimize this potential issue by using a standardized abstraction process. However, we did not calculate the reliability of our reviews. This study was limited to a sample of hospitalizations within VHA healthcare system and the sample was predominantly older males, which may limit the generalizability of the study findings to other settings.

Implications

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

Funding: This project was funded by the by VA Clinical Science research and Development investigator initiated grant CX000570-06 (Roumie). Dr. Roumie was supported in part by Center for Diabetes Translation Research P30DK092986. Dr. Min was supported by the *Clinical and Translational Science Award (CTSA) No. TL1TR000447-09 from the National Center for Advancing Translational Sciences.* Dr. Presley was supported by the Office of Academic Affiliations VA Quality Scholars program.

Data sharing statement: No additional data available

Contributorship statement: 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. Caroline Presley contributed to the design of the study, collection of data, analysis or interpretation of data, drafting of the manuscript, and final approval of the submission. Jonathan Chipman contributed to data analysis and interpretation, critical revision of the manuscript, and final approval of the submission. Robert Greevy contributed to collection of data, analysis or interpretation of data, critical revision of the manuscript, and final approval of the submission. Jea Young Min, Carlos Grijalva, and Marie Griffin contributed to the design of the study, critical revision of the manuscript, and final approval of the submission. Christianne Roumie contributed to the design of the study, data analysis and interpretation, drafting the manuscript, and final approval of the submission.

Conflicts of interest statement: The authors due not have any conflicts of interest to report.

References

1. Nichols GA, Gullion CM, Koro CE, Ephross SA, Brown JB. 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, Blondon M, Moore KP, Boyko EJ, Smith NL. 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, Ash AS, Berlowitz DR. Identifying hypertension-related comorbidities from administrative data: what's the optimal approach? American journal of medical quality : the official journal of the American College of Medical Quality 2004;19:201-6.
7. Brophy MT, Snyder KE, Gaehde S, Ives C, Gagnon D, Fiore LD. Anticoagulant use for atrial fibrillation in the elderly. Journal of the American Geriatrics Society 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. Clinical journal of the American Society of Nephrology : CJASN 2016;11:1954-61.
9. McCullough PA, Philbin EF, Spertus JA, Kaatz S, Sandberg KR, Weaver WD. Confirmation of a heart failure epidemic: findings from the Resource Utilization Among Congestive Heart Failure (REACH) study. Journal of the American College of Cardiology 2002;39:60-9.
10. Owan TE, Hodge DO, Herges RM, Jacobsen SJ, Roger VL, Redfield MM. Trends in prevalence and outcome of heart failure with preserved ejection fraction. The New England journal of medicine 2006;355:251-9.
11. Philbin EF, Rocco TA, Jr., Lynch LJ, Rogers VA, Jenkins P. Predictors and determinants of hospital length of stay in congestive heart failure in ten community hospitals. The Journal of heart and lung transplantation : the official publication of the International Society for Heart Transplantation 1997;16:548-55.
12. 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.
13. McKee PA, Castelli WP, McNamara PM, Kannel WB. The natural history of congestive heart failure: the Framingham study. The New England journal of medicine 1971;285:1441-6.
14. 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. Journal of the American College of Cardiology 2013;62:e147-239.
15. Camplain R, Kucharska-Newton A, Cuthbertson CC, Wright JD, Alonso A, Heiss G. 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.
16. Lumley T. Analysis of Complex Survey Samples. Journal of Statistical Software 2004;9:19.
17. Brown LD, Cai TT, DasGupta A. Interval Estimation for a Binomial Proportion. Statistical Science 2001;16:101-17.

18. 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.
19. Goff DC, Jr., Pandey DK, Chan FA, Ortiz C, Nichaman MZ. Congestive heart failure in the United States: is there more than meets the I(CD code)? The Corpus Christi Heart Project. Archives of internal medicine 2000;160:197-202.
20. 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.
21. Havranek EP, Masoudi FA, Westfall KA, Wolfe P, Ordin DL, Krumholz HM. Spectrum of heart failure in older patients: results from the National Heart Failure project. Am Heart J 2002;143:412-7.
22. Go AS, Lee WY, Yang J, Lo JC, Gurwitz JH. Statin therapy and risks for death and hospitalization in chronic heart failure. Jama 2006;296:2105-11.
23. 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.

Table 1: Framingham Criteria for Heart Failure, the Reference Standard for Classification of Hospitalizations [a]

Major Criteria	Minor Criteria
Paroxysmal nocturnal dyspnea or orthopnea	Night cough
Elevated jugular venous pressure	Dyspnea with exertion
Heart failure treatment-related 10 pound weight loss in preceding 5 days	Non-heart failure treatment-related 10 pound weight loss in preceding 5 days
S3 gallop	Hepatomegaly
Hepatojugular reflex	Bilateral ankle edema
Rales, crackles	Pleural effusion (on imaging)
Cardiomegaly (on imaging)	Pulmonary vascular engorgement (on imaging)
Pulmonary edema (on imaging)	Tachycardia (heart rate >120 beats/min)

[a] A hospitalization was classified as heart failure if it met a minimum of two major or one major and two minor criteria.

Table 2: Characteristics of Sampled Hospitalized Patients Based on Veterans Health Administration Data [a]

	All Patients (N=497)
Age in years, Mean (Standard deviation [SD])	66.1 (11.4)
Age groups, n (%)	
<55 years old	66 (13.3)
55 - 64 years old	174 (35.0)
65 - 74 years old	124 (25.0)
≥ 75 years old	133 (26.8)
Sex, n (%) Male	491 (98.8)
Race, n (%)	
White**, %**	373 (75.1)
Black**, %**	101 (20.3)
Other, %	23 (4.6)
Hypertension, n (%)	416 (83.7)
Hyperlipidemia, n (%)	292 (58.8)
Atherosclerotic Cardiovascular Disease, n (%)	307 (61.8)
Type 2 Diabetes, n (%)	430 (86.5)
Chronic Kidney Disease: Stage 3-5, n (%)	206 (41.5)
Body Mass Index (kg/m^2), Mean (SD)	31.3 (7.3)
Hemoglobin A1C (%), Mean (SD)	6.98 (1.6)

[a] Covariate data were collected from administrative sources, Veterans Health Administration data linked to Medicare and Medicaid data, for the 730 days preceding the study hospitalization.

Table 3: Positive and Negative Predictive Value, Sensitivity, Specificity for Overall Heart Failure Hospitalization Identification Algorithm [a], Weighted Analysis

	Confirmed HF hospitalization, sum weight [b] (n) [c]	**Confirmed non-HF hospitalization,** sum weight (n)	**Total hospitalizations,** sum weight (n)	**Performance metric** (95% Confidence interval, CI) [d]
HF algorithm positive	513 (354)	59 (45)	572 (399)	Positive predictive value 89.7 (86.8, 92.7)
HF algorithm negative	624 (6)	9,570 (92)	10,194 (98)	Negative predictive value 93.9 (89.1, 98.6)
Total	1,138 (360)	9,628 (137)	10,766 (497)	
Validity measure	Sensitivity (95% CI)	Specificity (95% CI)		
	45.1 (25.1, 65.1)	99.4 (99.2, 99.6)		

[a] The heart failure 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 diagnosis-related group (DRG) code 127 or 291-293.

[b] sum weight represents the number of hospitalizations from the overall study population that would have fallen into each category when inverse probability of sampling weights were applied to the study sample

[c] n represents the actual number of charts reviewed that were true positives, false positives, false negatives, or true negatives in each category

[d] To create 95% confidence intervals, we used a Taylor Series linearization to calculate standard errors with sampling weights

Table 4: Positive and Negative Predictive Value, Sensitivity, Specificity for Components of Heart Failure Algorithm

	Number of algorithm-positive hospitalizations, sum weight [a] (n) [b]	**Positive Predictive Value** (95% Confidence Interval [CI])[c]	**Negative Predictive Value** (95% CI)	**Sensitivity** (95% CI)	**Specificity** (95% CI)
All	572 (399)	89.7 (86.8, 92.7)	93.9 (89.1, 98.6)	45.1 (25.1, 65.1)	99.4 (99.2, 99.6)
ICD-9 and DRG	477 (304)	92.1 (89.1, 95.1)	93.9 (89.1, 98.6)	41.3 (21.6, 61.0)	99.6 (99.4, 99.7)
ICD-9 only	87 (87)	79.3 (70.7, 87.9)	93.9 (89.1, 98.6)	19.9 (4.8, 35.0)	99.6 (99.4, 99.8)
DRG only	8 (8)	62.5 (28.4, 96.6)	93.9 (89.1, 98.6)	0.79 (0.16, 1.75)	99.9 (99.9, 100)

[a] sum weight represents the number of hospitalizations from the overall study population that would have fallen into each category when inverse probability of sampling weights were applied to the study sample

[b] n represents the actual number of charts reviewed that were true positives, false positives, false negatives, or true negatives in each category

[c] To create 95% confidence intervals, we used a Taylor Series linearization to calculate standard errors with sampling weights

Appendix

Table A1: Positive Predictive Values for Individual ICD-9 Codes

	Algorithm-positive events fulfilling ICD-9 and DRG code criteria		Algorithm-positive events fulfilling only ICD-9 code criteria	
ICD-9 Code	Hospitalizations, N	Positive Predictive Value[a], (95% Confidence Interval [CI])[b]	Hospitalizations, N	Positive Predictive Value[a], (95% CI)
428.x Heart failure	293	92.8 (89.3, 95.3)	68	85.3 (75.0, 91.8)
428.0 Congestive heart failure unspecified	229	93.0 (89.7, 96.3)	55	89.1 (78.2, 94.9)
428.1 Left heart failure	0		0	
428.20 Systolic heart failure unspecified	5	80.0 (37.6, 96.4)	0	
428.21 Acute systolic heart failure	2	100 (34.2, 100.0)	2	50.0 (9.5, 90.5)
428.22 Chronic systolic heart failure	9	90.0 (70.1, 100.0)	1	0 (0, 79.3)
428.23 Acute on chronic systolic heart failure	14	100.0 (78.5, 100.0)	5	100.0 (56.6, 100.0)
428.30 Diastolic heart failure unspecified	7	85.7 (48.7, 97.4)	0	

428.31 Acute diastolic heart failure	1	100.0 (20.7, 100.0)	0	
428.32 Chronic diastolic heart failure	8	62.5 (30.6, 86.3)	1	100.0 (20.7, 100.0)
428.33 Acute on chronic diastolic heart failure	7	100.0 (64.6, 100.0)	1	100.0 (20.7, 100.0)
428.40 Combined systolic and diastolic heart failure	3	100.0 (43.9, 100.0)	0	
428.41 Acute combined systolic and diastolic heart failure	1	0 (0, 79.3)	1	0 (0, 79.3)
428.42 Chronic combined systolic and diastolic heart failure	0		0	
428.43 Acute on chronic combined systolic and diastolic heart failure	8	100.0 (67.6, 100.0)	1	100. 0 (20.7, 100.0)
428.9 Heart failure unspecified	0		1	0 (0, 79.3)
425.x Cardiomyopathy	0		12	50.0 (25.4, 74.6)
425.1 Hypertrophic obstructive cardiomyopathy	0		2	0 (0, 65.8)
425.2 Obscure cardiomyopathy of Africa	0		0	
425.3 Endocardial fibroelastosis	0		0	

425.4 Other primary cardiomyopathy	0		8	62.5 (30.6, 86.3)
425.5 Alcoholic cardiomyopathy	0		0	
425.7 Metabolic cardiomyopathy	0		0	
425.8 Cardiomyopathy in other diseases classified elsewhere	0		0	
425.9 Secondary cardiomyopathy unspecified	0		2	50.0 (9.5, 90.5)
404.x Hypertensive heart disease and chronic kidney disease with heart failure	4	50.0 (15.0, 85.0)	1	0 (0, 79.3)
404.01 Malignant hypertensive heart and chronic kidney disease with heart failure	0		1	0 (0, 79.3)
404.03 Malignant hypertensive heart and chronic kidney disease with heart failure with chronic kidney disease stage V or end stage renal disease	0		0	
404.11 Benign hypertensive heart and chronic kidney disease with heart failure and with chronic kidney disease stage I – stage IV or unspecified	0		0	

404.13 Benign hypertensive heart and chronic kidney disease with heart failure and with chronic kidney disease stage V or end stage renal disease	0		0	
404.91 Hypertensive heart disease and chronic kidney disease unspecified with heart failure and with chronic kidney disease stage I – stage IV or unspecified	3	66.7 (20.8, 93.9)	0	
404.93 Hypertensive heart disease and chronic kidney disease unspecified with heart failure and with chronic kidney disease stage V or end stage renal disease	1	0 (0, 79.3)	0	
402.x Hypertensive heart disease with heart failure	6	83.3 (43.6, 97.0)	6	83.3 (43.6, 97.0)
402.01 Malignant hypertensive heart disease with heart failure	1	0 (0, 79.3)	2	100.0 (34.2, 100.0)
402.11 Benign hypertensive heart disease with heart failure	0		0	
402.91 Hypertensive heart disease unspecified	5	100.0 (56.5, 100.0)	4	75.0 (30.0, 95.4)

with heart failure		
398.91 Rheumatic heart failure	0	0

[a] Positive predictive values were calculated by unweighted analysis. Sampling weights were not needed as each analysis was completed within a given sampling stratum.

[b] Wilson's formula was used to calculate 95% confidence interval

Table A2: Sensitivity analysis – Positive and negative predictive value, sensitivity, specificity of alternate algorithm allowing heart failure (HF) or cardiomyopathy codes in any discharge diagnosis position, weighted analysis

	Confirmed HF hospitalization, sum weight[a] (n)[b]	**Confirmed non-HF hospitalization,** sum weight (n)	**Total hospitalizations,** sum weight (n)	**Predictive value (95% Confidence interval, CI)[c]**
HF algorithm positive	929 (358)	1307 (57)	2236 (415)	Positive predictive value 41.5% (24.5, 58.6)
HF algorithm negative	208 (2)	8322 (80)	8530 (82)	Negative predictive value 97.6% (94.2, 100.0)
Total	1137 (360)	9629 (137)	10766 (497)	
	Sensitivity (95% CI) 81.7% (59.9, 100.0)	Specificity (95% CI) 86.4% (79.6, 93.3)		

[a] sum weight represents the number of hospitalizations from the overall study population that would have fallen into each category when inverse probability of sampling weights were applied to the study sample

[b] n represents the actual number of charts reviewed that were true positives, false positives, false negatives, or true negatives in each category

[c] To create 95% confidence intervals, Stata uses a Taylor Series linearization to calculate standard errors with sampling weights

Manuscript: Validation of an algorithm to identify heart failure hospitalizations in patients with diabetes within the Veterans Health Administration system

STROBE 2007 (v4) Statement—Checklist of items that should be included in reports of *cross-sectional studies*

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	2
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4
Objectives	3	State specific objectives, including any prespecified hypotheses	4-5
Methods			
Study design	4	Present key elements of study design early in the paper	5
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	5-6
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants	5
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	5-7
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	5-7
Bias	9	Describe any efforts to address potential sources of bias	*7-8*
Study size	10	Explain how the study size was arrived at	5
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	7-8
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	7-8
		(*b*) Describe any methods used to examine subgroups and interactions	7-8

		(*c*) Explain how missing data were addressed	7-8
		(*d*) If applicable, describe analytical methods taking account of sampling strategy	7-8
		(*e*) Describe any sensitivity analyses	7-8
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	8
		(b) Give reasons for non-participation at each stage	8
		(c) Consider use of a flow diagram	n/a
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	8-9
		(b) Indicate number of participants with missing data for each variable of interest	8-9
Outcome data	15*	Report numbers of outcome events or summary measures	9
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	9
		(*b*) Report category boundaries when continuous variables were categorized	n/a
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	n/a
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	10-11
Discussion			
Key results	18	Summarise key results with reference to study objectives	11
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	13
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	11-13
Generalisability	21	Discuss the generalisability (external validity) of the study results	13
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	14

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

Validation of an algorithm to identify heart failure hospitalizations in patients with diabetes within the Veterans Health Administration

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-020455.R1
Article Type:	Research
Date Submitted by the Author:	10-Jan-2018
Complete List of Authors:	Presley, Caroline; VA - Tennessee Valley Healthcare System, Geriatric Research Education and Clinical Center; Vanderbilt University Medical Center, Medicine Min, Jea Young ; VA - Tennessee Valley Healthcare System, Geriatric Research Education and Clinical Center; Vanderbilt University Medical Center, Health Policy Chipman, Jonathan; Vanderbilt University Medical Center, Biostatistics Greevy, Robert; Vanderbilt University Medical Center, Biostatistics Grijalva, Carlos; VA - Tennessee Valley Healthcare System, Geriatric Research Education and Clinical Center; Vanderbilt University Medical Center, Health Policy Griffin, Marie R; VA - Tennessee Valley Healthcare System, Geriatric Research Education and Clinical Center; Vanderbilt University Medical Center, Medicine Roumie, Christianne L.; VA - Tennessee Valley Healthcare System, Geriatric Research Education and Clinical Center; Vanderbilt University Medical Center, Medicine
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Diabetes and endocrinology, Cardiovascular medicine
Keywords:	Validation study, Pharmacoepidemiology, General diabetes < DIABETES & ENDOCRINOLOGY, Heart failure < CARDIOLOGY

Validation of an algorithm to identify heart failure hospitalizations in patients with diabetes within the Veterans Health Administration

Caroline Presley, M.D., MPH[1,2]; Jea Young Min, Pharm.D., M.P.H[1]; Jonathan Chipman, M.S.[3]; Robert A. Greevy, Jr, PhD[3]; Carlos G. Grijalva, M.D., M.P.H.[1,4]; Marie R. Griffin, M.D., M.P.H.[1,2,4]; Christianne L. Roumie, M.D., M.P.H.[1,2]

1. Veterans Health Administration-Tennessee Valley Healthcare System Geriatric Research Education Clinical Center (GRECC), HSR&D Center, Nashville, TN
2. Department of Medicine, Vanderbilt University Medical Center, Nashville, TN.
3. Department of Biostatistics, Vanderbilt University Medical Center, Nashville TN
4. Department of Health Policy, Vanderbilt University Medical Center, Nashville, TN.

Corresponding author: Caroline Presley, M.D., M.P.H.

1310 24th Avenue South, GRECC, Nashville, TN 37212

Phone: 615-873-8012; Fax: 615-873-7981

Caroline.a.presley@vanderbilt.edu; caroline.presley@va.gov

Word count: 3296

Abstract

Objectives: We aimed to validate an algorithm using both primary discharge diagnosis (ICD-9) and diagnosis-related group (DRG) codes to identify hospitalizations due to decompensated heart failure in a population of patients with diabetes within the Veterans Health Administration system.

Design: Validation study

Setting: Veterans Health Administration - Tennessee Valley Healthcare System

Participants: We identified and reviewed a stratified, random sample of hospitalizations between 2001 and 2012 within a single Veterans Health Administration healthcare system of adults who received regular VHA care and were initiated on an antidiabetic medication between 2001 and 2008. We sampled 500 hospitalizations; 400 hospitalizations that fulfilled algorithm criteria, 100 that did not. Of these, 497 had adequate information for inclusion. The mean patient age was 66.1 years (Standard deviation 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 hospitalization was due to heart failure, we performed chart abstraction using Framingham criteria as the referent standard. We calculated the positive predictive value (PPV), negative predictive value, sensitivity, and specificity for the overall algorithm and each component (primary diagnosis code [ICD-9], DRG code, or both).

Results: The algorithm had a positive predictive value of 89.7% (95% confidence interval: 86.8, 92.7), negative predictive value of 93.9% (89.1, 98.6), sensitivity of 45.1% (25.1, 65.1), and specificity of 99.4% (99.2, 99.6). The PPV was highest for hospitalizations that fulfilled both the ICD-9 and DRG algorithm criteria (92.1% [89.1, 95.1]), and lowest for hospitalizations that fulfilled only DRG algorithm criteria (62.5% [28.4, 96.6]).

Conclusions: Our algorithm, which included primary discharge diagnosis and diagnosis-related group codes, demonstrated excellent positive predictive value for identification of

hospitalizations due to decompensated heart failure among patients with diabetes in the Veterans Health Administration system.

Strengths and Limitations of this Study

- This is the first study to validate an algorithm using both primary discharge diagnosis (ICD-9) and diagnosis-related group (DRG) codes to identify hospitalizations due to decompensated heart failure within the Veterans Health Administration system.
- We applied a sampling strategy that allowed weighted estimations to extrapolate findings to our underlying study population.
- We used standardized Framingham heart failure criteria for our adjudications; we performed a complete validation assessment, contrasted with other studies that have only reported positive predictive values.
- Study limitations include potentially limited generalizability of findings to other settings, and data abstraction by chart review may be subject to error.
- The validation of this algorithm will facilitate future study of the risk of heart failure hospitalizations associated with antidiabetic medication regimens in Veterans Health Administration patients with diabetes, especially in comparative effectiveness studies.

Introduction

Patients with diabetes are up to two and a half times more likely to develop heart failure than those without diabetes.[1] Several mechanisms may play a role in this increased risk of heart failure including diabetic cardiomyopathy, as well as co-morbid hypertension and atherosclerotic cardiovascular disease.[2] Thiazolidinediones have been shown to increase heart failure risk in patients with type 2 diabetes (T2DM).[3] Little evidence exists on the risk of heart failure outcomes associated with use of common first and second line antidiabetic medications (i.e. metformin, sulfonylurea, insulin), as heart failure has been an infrequent primary outcome in clinical trials.[4]

Observational studies using administrative data are an important alternative to randomized clinical trials to evaluate the risk of heart failure, including hospitalizations due to decompensated heart failure, 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 minimize misclassification of outcomes, it is necessary to validate algorithms that identify decompensated heart failure as the primary reason for hospital admission, not as a preexisting comorbidity or a complication that developed during the course of hospitalization.

Although algorithms to identify heart failure events have been validated in the Veterans Health Administration (VHA) system, these included both inpatient and outpatient encounters and did not specifically focus on events resulting from decompensated heart failure.[5-7] Additionally, these algorithms only relied on International Classification of Diseases, 9th revision [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 hospitalizations due to decompensated heart failure has not been tested within VHA.[2,8] Such algorithms have performed well in academic and community health systems (PPV 83-96%).[9-11] We aimed to validate an algorithm using both

primary discharge diagnosis (ICD-9) and DRG codes to identify hospitalizations due to decompensated heart failure in a population of patients with diabetes within the VHA system.

Methods

Study Design

This was a validation study of an algorithm to identify heart failure hospitalizations 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 observational cohort of Veterans who were initiated on an oral hypoglycemic medication between 2001 and 2008 (N=411,055); 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 outpatient encounter, emergency department visit, hospitalization, or medication refill at least once every 180 days), were diagnosed with diabetes (at least one prescription filled for an oral hypoglycemic medication) between 2001 and 2008, and were hospitalized in TVHS between 2001 and 2012. For this study, a patient's diagnosis of diabetes could have occurred before or after the included study hospitalization to allow adequate sampling of hospitalizations meeting heart failure algorithm criteria.

Study events

The algorithm identified hospitalizations with a primary discharge diagnosis code (ICD-9) of heart failure 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, Appendix Table A1), and/or a diagnosis-related group (DRG) code for heart failure (127, used prior to fiscal year 2008; 291-293, used after fiscal year 2008). We sampled 500 hospitalizations 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 hospitalizations from the following strata: hospitalizations fulfilling both ICD-9 and DRG code criteria, only ICD-9 code criteria, and only DRG code criteria, as well as, algorithm-negative hospitalizations. The probability of selection within strata was used to calculate sampling weights in each stratum (i.e. weights = (# of hospitalizations in the sampling strata) / (# of hospitalizations sampled from that strata)). We weighted observations so the stratified sample accurately reflected the underlying study population of hospitalizations. An individual could be included in the study more than once if they had multiple hospitalizations sampled. The HF algorithm operates on each hospitalization independently, thus a random sample hospitalizations (as opposed to patients who may have a mix of algorithm positive and negative hospitalizations over time) was needed for unbiased estimates of the algorithm's performance on identifying HF in hospitalizations for this population.

Data collection

Data were abstracted from the VHA's electronic medical record using standardized forms by an Internal Medicine physician, blinded to heart failure algorithm status. We used the standardized Framingham criteria, to classify hospitalizations as decompensated heart failure.[13] The presence or absence of symptoms, signs, and radiologic features of heart failure were abstracted from the electronic medical record from within the first 24 hours of the admission date to avoid capturing signs or symptoms of heart failure not present upon admission. A hospitalization met criteria for heart failure 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 classify heart failure hospitalizations as heart failure with reduced ejection fraction (HFrEF, EF ≤ 40%), heart failure with preserved ejection fraction (HFpEF, EF ≥ 50%), or borderline HFpEF (EF 41-49%) according to American College of Cardiology Foundation/American Heart Association guidelines.[15] The ejection fraction measurement collected during or in closest proximity (up to one year prior) to the study

hospitalization was used. If multiple assessments were present, the ejection fraction measurement from an echocardiogram was used if available, followed by measurements from cardiac catheterization or a nuclear medicine study, respectively. Furthermore, heart failure hospitalizations were classified as incident (new-onset heart failure) or prevalent (exacerbation of chronic heart failure). For this, the investigator examined the electronic medical record for the two years preceding the study hospitalization to determine if the patient had a prior diagnosis of or hospitalization for heart failure.[16]

Covariates

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

Statistical Analysis

Descriptive statistics were used to characterize the study sample and hospitalizations including type of heart failure and incident or prevalent classification for confirmed heart failure hospitalizations.

Using the chart review classification based on Framingham criteria as the reference standard, we calculated the positive predictive value (PPV, proportion of algorithm-positive cases confirmed as heart failure) for the overall algorithm and each component (primary diagnosis code [ICD-9], DRG code, or both). Chart review classifications for each hospitalization were treated as statistically independent, as they were determined using only data collected from each discrete hospitalization. We also calculated the negative predictive value (NPV, proportion of algorithm-negative cases confirmed as non-heart failure), sensitivity (proportion of heart failure hospitalizations correctly identified by the algorithm), and specificity (proportion of non-heart failure hospitalizations 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 hospitalizations. To create 95% confidence intervals, a Taylor Series linearization was used to calculate standard errors with sampling weights.[17] We calculated positive predictive values for each distinct ICD-9 code included in the algorithm for hospitalizations that met both ICD-9 and DRG code criteria, as well as, for hospitalizations 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% confidence intervals due to smaller sample sizes.[18]

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

Statistical analyses were performed using Stata Statistical Software: Release 14, College Station, TX: StataCorp LP.

Results

Of 10,766 eligible hospitalizations in TVHS between 2001 and 2012, a total of 500 hospitalizations were sampled. Of the 500 sampled hospitalizations, 324 unique patients were represented only once (i.e. contributed only 1 hospitalization for review); the remaining 176 hospitalizations were from patients who contributed more than one hospitalizations (range 2-9).

Of the algorithm-positive hospitalizations, 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 hospitalizations, three had insufficient documentation to assess Framingham criteria (one algorithm-positive, two algorithm-negative); thus, 497 hospitalizations were included.

The patients were on average 66.1 years old (Standard deviation [SD] 11.4) with a median age of 65 years (interquartile range [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%), hyperlipidemia (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 hospitalization. Mean hemoglobin A1c was 6.96% (SD 1.6).

Of 497 hospitalizations reviewed, 360 (72.4%) fulfilled Framingham criteria for decompensated heart failure. Of these 360, 127 (35.3%) were incident heart failure events, 229 (63.6%) were prevalent events, and four (1.1%) had insufficient documentation for this determination. Additionally, 186 of the 360 heart failure hospitalizations (51.7%) were classified as HFrEF; 86 (23.9%) were HFpEF; 36 (10.0%) were HFpEF borderline; and 52 (14.4%) did not have ejection fraction data available. Of patients who had a confirmed HF hospitalization and available EF data, 172 of 308 (55.8%) patients had their EF assessed during the study hospitalization; the remainder had an assessment of EF during the year prior to the study hospitalization.

Overall, we found 354 true positive hospitalizations due to heart failure, 45 false positives, six false negatives, and 92 true negatives. Of the six heart failure algorithm-negative hospitalizations that fulfilled Framingham criteria, four had a heart failure 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 hospitalizations included: subendocardial infarction, initial episode of care; diabetes

with ophthalmic manifestations, type II or unspecified type, uncontrolled; anxiety state, unspecified; and atrioventricular block, complete. Primary discharge diagnosis codes for the two hospitalizations that did not include a heart failure 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% confidence interval, 86.8, 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 hospitalizations 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 hospitalizations 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 hospitalizations with different ICD-9 primary discharge diagnosis codes. The PPV of the algorithm limited to hospitalizations with 428.x codes (Heart failure) that fulfilled both ICD-9 and DRG code criteria was highest, 92.8% (89.3, 95.3), Appendix Table A1. For hospitalizations with 428.x codes that only fulfilled ICD-9 code criteria, PPV was 85.3% (75.0, 91.8). For hospitalizations with ICD-9 code of 402.x (Hypertensive heart disease with heart failure), the PPV of the algorithm was 83.3% (43.6, 97.0) for both hospitalizations 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 hospitalizations with a primary discharge diagnosis code of 404.x (Hypertensive heart disease and chronic kidney disease with heart failure) or 425.x (Cardiomyopathy). The PPV was 50.0% (15.0, 85.0) for hospitalizations with a 404.x code that met both ICD-9 and DRG code criteria and 0% (0, 79.3) for hospitalizations with 404.x code that met only ICD-9 criteria. In our sample, no hospitalizations with an ICD-9 code of 425.x met both

ICD-9 and DRG code criteria. The PPV for hospitalizations 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 hospitalization, 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 hospitalizations were reclassified as algorithm-positive hospitalizations using this alternate algorithm. Of these, four hospitalizations were confirmed heart failure hospitalizations by chart review (events discussed above), and 12 hospitalizations were confirmed non-heart failure hospitalizations. This alternate algorithm had higher sensitivity, 81.7% (59.9, 100.0) vs. 45.1% (25.1, 65.1), but had poor PPV, 41.6% (24.5, 58.6) vs. 89.7% (86.8, 92.7), and lower specificity, 86.4% (79.6, 93.3) vs. 99.4% (99.2, 99.6), compared with the original heart failure hospitalization study algorithm, Appendix Table A2.

Discussion

Our algorithm to identify hospitalizations due to decompensated heart failure 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%).[9-11] 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.[2] 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 hospitalizations.

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.[5] 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 heart failure as the primary diagnosis code and/or a DRG code for heart failure, and rigorous use of the Framingham criteria to adjudicate potential heart failure 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 heart failure as an outcome for antidiabetic medications. In these studies, high specificity outcome definitions help minimize 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 hospitalizations selected as algorithm-positive are very likely due to a true heart failure hospitalization. An algorithm with higher sensitivity may be more appropriate if one is seeking to capture heart failure as a co-morbidity 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 hospitalizations that fulfilled both the primary discharge diagnosis code and DRG code criteria, and had the lowest PPV for hospitalizations 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 heart failure hospitalizations.

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%.[20,22-24] By including multiple ICD-9 codes in our algorithm, we were able to compare positive predictive values for individual ICD-9 codes. The algorithm performed best for hospitalizations with ICD-9 code 428.x and had lowest PPV for ICD-9 codes 404.x and 425.x, although the number of hospitalizations with the latter two codes was limited. While we did not evaluate an algorithm that included ICD-10 codes, our data suggests that I50.x (Heart failure) and I11.0 (Heart failure due to hypertension), which correspond to the 428.x and 402.x ICD-9 codes, will perform best to identify heart failure hospitalizations.

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 standardized Framingham heart failure criteria for our adjudications, and

complemented those data with heart failure classifications based on ejection fraction 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 minimize this potential issue by using a standardized abstraction process. However, we did not calculate the reliability of our reviews. This study was limited to a sample of hospitalizations within VHA healthcare system and the sample was predominantly older males, which may limit the generalizability of the study findings to other settings. Additionally, misclassification of HF hospitalizations by EF may exist as we used EF assessments from up to one year prior to the study hospitalization; though 55.8% of assessments were completed during the study hospitalization.

Implications

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

Funding: This project was funded by the by VA Clinical Science research and Development investigator initiated grant CX000570-06 (Roumie). Dr. Roumie was supported in part by Center for Diabetes Translation Research P30DK092986. Dr. Min was supported by the *Clinical and Translational Science Award (CTSA) No. TL1TR000447-09 from the National Center for Advancing Translational Sciences.* Dr. Presley was supported by the Office of Academic Affiliations VA Quality Scholars program.

Data sharing statement: No additional data available

Contributorship statement: 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. Caroline Presley contributed to the design of the study, collection of data, analysis or interpretation of data, drafting of the manuscript, and final approval of the submission. Jonathan Chipman contributed to data analysis and interpretation, critical revision of the manuscript, and final approval of the submission. Robert Greevy contributed to collection of data, analysis or interpretation of data, critical revision of the manuscript, and final approval of the submission. Jea Young Min, Carlos Grijalva, and Marie Griffin contributed to the design of the study, critical revision of the manuscript, and final approval of the submission. Christianne Roumie contributed to the design of the study, data analysis and interpretation, drafting the manuscript, and final approval of the submission.

Conflicts of interest statement: The authors due not have any conflicts of interest to report.

References

1. Nichols GA, Gullion CM, Koro CE, Ephross SA, Brown JB. 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, Blondon M, Moore KP, Boyko EJ, Smith NL. 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, Ash AS, Berlowitz DR. Identifying hypertension-related comorbidities from administrative data: what's the optimal approach? American journal of medical quality : the official journal of the American College of Medical Quality 2004;19:201-6.
7. Brophy MT, Snyder KE, Gaehde S, Ives C, Gagnon D, Fiore LD. Anticoagulant use for atrial fibrillation in the elderly. Journal of the American Geriatrics Society 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. Clinical journal of the American Society of Nephrology : CJASN 2016;11:1954-61.
9. McCullough PA, Philbin EF, Spertus JA, Kaatz S, Sandberg KR, Weaver WD. Confirmation of a heart failure epidemic: findings from the Resource Utilization Among Congestive Heart Failure (REACH) study. Journal of the American College of Cardiology 2002;39:60-9.
10. Owan TE, Hodge DO, Herges RM, Jacobsen SJ, Roger VL, Redfield MM. Trends in prevalence and outcome of heart failure with preserved ejection fraction. The New England journal of medicine 2006;355:251-9.
11. Philbin EF, Rocco TA, Jr., Lynch LJ, Rogers VA, Jenkins P. Predictors and determinants of hospital length of stay in congestive heart failure in ten community hospitals. The Journal of heart and lung transplantation : the official publication of the International Society for Heart Transplantation 1997;16:548-55.
12. Roumie CL, Hung AM, Greevy RA, et al. Comparative effectiveness of sulfonylurea and metformin monotherapy on cardiovascular events in type 2 diabetes mellitus: a cohort study. Annals of internal medicine 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, Kannel WB. The natural history of congestive heart failure: the Framingham study. The New England journal of medicine 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. Journal of the American College of Cardiology 2013;62:e147-239.
16. Camplain R, Kucharska-Newton A, Cuthbertson CC, Wright JD, Alonso A, Heiss G. 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. Journal of Statistical Software 2004;9:19.
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, Jr., Pandey DK, Chan FA, Ortiz C, Nichaman MZ. Congestive heart failure in the United States: is there more than meets the I(CD code)? The Corpus Christi Heart Project. Archives of internal medicine 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, Wolfe P, Ordin DL, Krumholz HM. 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, Lo JC, Gurwitz JH. 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.

Table 1: Framingham Criteria for Heart Failure, the Reference Standard for Classification of Hospitalizations [a]

Major Criteria	Minor Criteria
Paroxysmal nocturnal dyspnea or orthopnea	Night cough
Elevated jugular venous pressure	Dyspnea with exertion
Heart failure treatment-related 10 pound weight loss in preceding 5 days	Non-heart failure treatment-related 10 pound weight loss in preceding 5 days
S3 gallop	Hepatomegaly
Hepatojugular reflex	Bilateral ankle edema
Rales, crackles	Pleural effusion (on imaging)
Cardiomegaly (on imaging)	Pulmonary vascular engorgement (on imaging)
Pulmonary edema (on imaging)	Tachycardia (heart rate >120 beats/min)

[a] A hospitalization was classified as heart failure if it met a minimum of two major or one major and two minor criteria.

Table 2: Characteristics of Sampled Hospitalized Patients Based on Veterans Health Administration Data [a]

	All Patients (N=497)
Age in years, Mean (Standard deviation [SD])	66.1 (11.4)
Age groups, n (%)	
<55 years old	66 (13.3)
55 - 64 years old	174 (35.0)
65 - 74 years old	124 (25.0)
≥ 75 years old	133 (26.8)
Sex, n (%) Male	491 (98.8)
Race, n (%)	
White**, %**	373 (75.1)
Black**, %**	101 (20.3)
Other, %	23 (4.6)
Hypertension, n (%)	416 (83.7)
Hyperlipidemia, n (%)	292 (58.8)
Atherosclerotic Cardiovascular Disease, n (%)	307 (61.8)
Type 2 Diabetes, n (%)	430 (86.5)
Chronic Kidney Disease: Stage 3-5, n (%)	206 (41.5)
Body Mass Index (kg/m^2), Mean (SD)	31.3 (7.3)
Hemoglobin A1C (%), Mean (SD)	6.98 (1.6)

[a] Covariate data were collected from administrative sources, Veterans Health Administration data linked to Medicare and Medicaid data, for the 730 days preceding the study hospitalization.

Table 3: Positive and Negative Predictive Value, Sensitivity, Specificity for Overall Heart Failure Hospitalization Identification Algorithm [a], Weighted Analysis

	Confirmed HF hospitalization, sum weight [b] (n) [c]	**Confirmed non-HF hospitalization,** sum weight (n)	**Total hospitalizations,** sum weight (n)	**Performance metric** (95% Confidence interval, CI) [d]
HF algorithm positive	513 (354)	59 (45)	572 (399)	Positive predictive value 89.7 (86.8, 92.7)
HF algorithm negative	624 (6)	9,570 (92)	10,194 (98)	Negative predictive value 93.9 (89.1, 98.6)
Total	1,138 (360)	9,628 (137)	10,766 (497)	
Validity measure	Sensitivity (95% CI)	Specificity (95% CI)		
	45.1 (25.1, 65.1)	99.4 (99.2, 99.6)		

[a] The heart failure 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 diagnosis-related group (DRG) code 127 or 291-293.

[b] sum weight represents the number of hospitalizations from the overall study population that would have fallen into each category when inverse probability of sampling weights were applied to the study sample

[c] n represents the actual number of charts reviewed that were true positives, false positives, false negatives, or true negatives in each category

[d] To create 95% confidence intervals, we used a Taylor Series linearization to calculate standard errors with sampling weights

Table 4: Positive and Negative Predictive Value, Sensitivity, Specificity for Components of Heart Failure Algorithm

	Number of algorithm-positive hospitalizations, sum weight [a] (n) [b]	**Positive Predictive Value** (95% Confidence Interval [CI])[c]	**Negative Predictive Value** (95% CI)	**Sensitivity** (95% CI)	**Specificity** (95% CI)
All	572 (399)	89.7 (86.8, 92.7)	93.9 (89.1, 98.6)	45.1 (25.1, 65.1)	99.4 (99.2, 99.6)
ICD-9 and DRG	477 (304)	92.1 (89.1, 95.1)	93.9 (89.1, 98.6)	41.3 (21.6, 61.0)	99.6 (99.4, 99.7)
ICD-9 only	87 (87)	79.3 (70.7, 87.9)	93.9 (89.1, 98.6)	19.9 (4.8, 35.0)	99.6 (99.4, 99.8)
DRG only	8 (8)	62.5 (28.4, 96.6)	93.9 (89.1, 98.6)	0.79 (0.16, 1.75)	99.9 (99.9, 100)

[a] sum weight represents the number of hospitalizations from the overall study population that would have fallen into each category when inverse probability of sampling weights were applied to the study sample

[b] n represents the actual number of charts reviewed that were true positives, false positives, false negatives, or true negatives in each category

[c] To create 95% confidence intervals, we used a Taylor Series linearization to calculate standard errors with sampling weights

Appendix

Table A1: Positive Predictive Values for Individual ICD-9 Codes

	Algorithm-positive events fulfilling ICD-9 and DRG code criteria		Algorithm-positive events fulfilling only ICD-9 code criteria	
ICD-9 Code	Hospitalizations, N	Positive Predictive Value[a], (95% Confidence Interval [CI])[b]	Hospitalizations, N	Positive Predictive Value[a], (95% CI)
428.x Heart failure	293	92.8 (89.3, 95.3)	68	85.3 (75.0, 91.8)
428.0 Congestive heart failure unspecified	229	93.0 (89.7, 96.3)	55	89.1 (78.2, 94.9)
428.1 Left heart failure	0		0	
428.20 Systolic heart failure unspecified	5	80.0 (37.6, 96.4)	0	
428.21 Acute systolic heart failure	2	100 (34.2, 100.0)	2	50.0 (9.5, 90.5)
428.22 Chronic systolic heart failure	9	90.0 (70.1, 100.0)	1	0 (0, 79.3)
428.23 Acute on chronic systolic heart failure	14	100.0 (78.5, 100.0)	5	100.0 (56.6, 100.0)
428.30 Diastolic heart failure unspecified	7	85.7 (48.7, 97.4)	0	

428.31 Acute diastolic heart failure	1	100.0 (20.7, 100.0)	0	
428.32 Chronic diastolic heart failure	8	62.5 (30.6, 86.3)	1	100.0 (20.7, 100.0)
428.33 Acute on chronic diastolic heart failure	7	100.0 (64.6, 100.0)	1	100.0 (20.7, 100.0)
428.40 Combined systolic and diastolic heart failure	3	100.0 (43.9, 100.0)	0	
428.41 Acute combined systolic and diastolic heart failure	1	0 (0, 79.3)	1	0 (0, 79.3)
428.42 Chronic combined systolic and diastolic heart failure	0		0	
428.43 Acute on chronic combined systolic and diastolic heart failure	8	100.0 (67.6, 100.0)	1	100. 0 (20.7, 100.0)
428.9 Heart failure unspecified	0		1	0 (0, 79.3)
425.x Cardiomyopathy	0		12	50.0 (25.4, 74.6)
425.1 Hypertrophic obstructive cardiomyopathy	0		2	0 (0, 65.8)
425.2 Obscure cardiomyopathy of Africa	0		0	
425.3 Endocardial fibroelastosis	0		0	

425.4 Other primary cardiomyopathy	0		8	62.5 (30.6, 86.3)
425.5 Alcoholic cardiomyopathy	0		0	
425.7 Metabolic cardiomyopathy	0		0	
425.8 Cardiomyopathy in other diseases classified elsewhere	0		0	
425.9 Secondary cardiomyopathy unspecified	0		2	50.0 (9.5, 90.5)
404.x Hypertensive heart disease and chronic kidney disease with heart failure	4	50.0 (15.0, 85.0)	1	0 (0, 79.3)
404.01 Malignant hypertensive heart and chronic kidney disease with heart failure	0		1	0 (0, 79.3)
404.03 Malignant hypertensive heart and chronic kidney disease with heart failure with chronic kidney disease stage V or end stage renal disease	0		0	
404.11 Benign hypertensive heart and chronic kidney disease with heart failure and with chronic kidney disease stage I – stage IV or unspecified	0		0	

404.13 Benign hypertensive heart and chronic kidney disease with heart failure and with chronic kidney disease stage V or end stage renal disease	0		0	
404.91 Hypertensive heart disease and chronic kidney disease unspecified with heart failure and with chronic kidney disease stage I – stage IV or unspecified	3	66.7 (20.8, 93.9)	0	
404.93 Hypertensive heart disease and chronic kidney disease unspecified with heart failure and with chronic kidney disease stage V or end stage renal disease	1	0 (0, 79.3)	0	
402.x Hypertensive heart disease with heart failure	6	83.3 (43.6, 97.0)	6	83.3 (43.6, 97.0)
402.01 Malignant hypertensive heart disease with heart failure	1	0 (0, 79.3)	2	100.0 (34.2, 100.0)
402.11 Benign hypertensive heart disease with heart failure	0		0	

402.91 Hypertensive heart disease unspecified with heart failure	5	100.0 (56.5, 100.0)	4	75.0 (30.0, 95.4)
398.91 Rheumatic heart failure	0		0	

[a] Positive predictive values were calculated by unweighted analysis. Sampling weights were not needed as each analysis was completed within a given sampling stratum.

[b] Wilson's formula was used to calculate 95% confidence interval

Table A2: Sensitivity analysis – Positive and negative predictive value, sensitivity, specificity of alternate algorithm allowing heart failure (HF) or cardiomyopathy codes in any discharge diagnosis position, weighted analysis

	Confirmed HF hospitalization, sum weight[a] (n)[b]	**Confirmed non-HF hospitalization,** sum weight (n)	**Total hospitalizations,** sum weight (n)	**Predictive value (95% Confidence interval, CI)[c]**
HF algorithm positive	929 (358)	1307 (57)	2236 (415)	Positive predictive value 41.5% (24.5, 58.6)
HF algorithm negative	208 (2)	8322 (80)	8530 (82)	Negative predictive value 97.6% (94.2, 100.0)
Total	1137 (360)	9629 (137)	10766 (497)	
	Sensitivity (95% CI) 81.7% (59.9, 100.0)	Specificity (95% CI) 86.4% (79.6, 93.3)		

[a] sum weight represents the number of hospitalizations from the overall study population that would have fallen into each category when inverse probability of sampling weights were applied to the study sample

[b] n represents the actual number of charts reviewed that were true positives, false positives, false negatives, or true negatives in each category

[c] To create 95% confidence intervals, Stata uses a Taylor Series linearization to calculate standard errors with sampling weights

Manuscript: Validation of an algorithm to identify heart failure hospitalizations in patients with diabetes within the Veterans Health Administration system

STROBE 2007 (v4) Statement—Checklist of items that should be included in reports of *cross-sectional studies*

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	2
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4
Objectives	3	State specific objectives, including any prespecified hypotheses	4-5
Methods			
Study design	4	Present key elements of study design early in the paper	5
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	5-6
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants	5
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	5-7
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	5-7
Bias	9	Describe any efforts to address potential sources of bias	*7-8*
Study size	10	Explain how the study size was arrived at	5
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	7-8
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	7-8
		(*b*) Describe any methods used to examine subgroups and interactions	7-8

		(*c*) Explain how missing data were addressed	7-8
		(*d*) If applicable, describe analytical methods taking account of sampling strategy	7-8
		(*e*) Describe any sensitivity analyses	7-8
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	8
		(b) Give reasons for non-participation at each stage	8
		(c) Consider use of a flow diagram	n/a
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	8-9
		(b) Indicate number of participants with missing data for each variable of interest	8-9
Outcome data	15*	Report numbers of outcome events or summary measures	9
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	9
		(*b*) Report category boundaries when continuous variables were categorized	n/a
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	n/a
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	10-11
Discussion			
Key results	18	Summarise key results with reference to study objectives	11
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	13
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	11-13
Generalisability	21	Discuss the generalisability (external validity) of the study results	13
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	14

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.