PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Validation of an algorithm to identify heart failure hospitalizations in patients with diabetes within the Veterans Health Administration
AUTHORS	Presley, Caroline; Min, Jea Young; Chipman, Jonathan; Greevy, Robert; Grijalva, Carlos; Griffin, Marie R; Roumie, Christianne L.

VERSION 1 – REVIEW

REVIEWER	Blánaid Hicks Centre for Public Health, School of Medicine, Dentistry and Biomedical science, Queen's University Belfast. Belfast, N. Ireland
REVIEW RETURNED	28-Nov-2017

GENERAL COMMENTS	This is a validation study of an algorithm using both primary discharge codes and diagnosis related group codes to identify heart failure in a cohort of patients with diabetes from the Veterans Health Administration. The authors report that using their algorithm exhibited high PPV (89.7%) , NPV and specificity but low sensitivity. This is a very well conducted study. The manuscript is very well written and thorough. The methods and results reported are very clear and appropriate. The authors have also conducted a number of sensitivity analyses, adding to the study. I have only a few very minor comments. 1. The authors in the discussion highlight the low sensitivity observed. The correctly explain why this is less of an issue for studies of comparative effectiveness with HF as an outcome. I note that this is in contrast to the study by Floyd et al using VA data where sensitivity was 90% for HF (albeit chronic heart failure). It could be of interest to the readership to also include in the discussion the reason for the low sensitivity observed in this study. 2. Why did the authors not include ICD 10 codes? Was it possible to include these? If so there might be some added benefit as it may increase the generalizability to other settings where these are utilized. 3. The authors use ejection fraction to classify HF hospitalizations. This can be collect up to a year prior to the HF event. How many of these values where collected at hospitalization? If for many they were collect before the event is it possible there is some misclassification here?

REVIEWER	Kanan Patel University of California San Francisco, USA
REVIEW RETURNED	02-Dec-2017

GENERAL COMMENTS	In the manuscript entitled "Validation of an algorithm to identify heart failure hospitalizations in patients with diabetes within the Veterans Health Administration" the authors described validation of algorithm using primary discharge diagnosis using ICD-9 codes and diagnosis related group codes to identify hospitalization due to decompensated heart failure among diabetic patients in VA system. They demonstrated good positive predictive value, negative predictive value and specificity but poor sensitivity. In addition when they used top five discharge diagnosis instead of using only primary discharge diagnosis, sensitivity improved but very poor PPV and lower specificity which showed limitation of administrative data. The manuscript addressed important issue. Few comments related to the current manuscript are listed as below. 1. In abstract, page 2, participants section: Description of diabetic patients is missing. 2. Methods, study population, page 5: Authors mentioned about inclusion criteria of "visit at least once every 180 days" and "were diagnosed with diabetes (at least one prescription filled for an antidiabetic medication) between 2001 and 2008". Does it means any visit – including IP, ED or outpatient visit? Why selected population with diagnosis of diabetes between 2001-2008 instead of 2001-2012? 3. Methods, study events, page 5: Can authors explain if there was any particular reason for choosing 4:1 ratio for algorithm positive to algorithm negative criteria? 4. There were > 10,000 eligible hospitalization to select 500 sample hospitalization. Why authors included individuals with multiple hospitalization more than once instead of using 500 unique individuals with hospitalization. 5. Results, page 8, 1st sentence: "Of 10,766 eligible hospitalizations in TVHS between 2001 and 2012, a total of 500 hospitalizations were sampled." Please clearly describe if 10,766 eligible hospitalization due to diabetic population or not? And what is n of unique individuals among 10,766 and 500 sampled hospitalization? 6. Results, page 8, 2nd para: authors mentioned that majority of the patients were older adults, aged >=65 years but as showed in Table 2, 48% patients were aged <65 years old. 7. Minor comments: • Table 4: It would be interesting to see values of PPV, NPV, sensitivity and specificity by any ICD-9 code and any DRG code. • Results, page 11, line 10: there was a typo error in sensitivity numbers. It should be 45.1% instead of 41.5%. • In discussion, page 11, para 1, lines 18-26: Use abbreviation of DRG, PPV, NPV.

VERSION 1 – AUTHOR RESPONSE

January 9, 2018

Trish Groves, MBBS, MRCPsych

Editor-in-Chief
BMJ Open

Dear Dr. Groves,

Thank you for further consideration of the manuscript "Validation of an algorithm to identify heart failure hospitalizations in patients with diabetes within the Veterans Health Administration" at BMJ Open. We appreciate your insightful comments and suggestions, and the opportunity to revise this manuscript. All the authors have contributed significantly and approved the final manuscript. The authors accept full responsibility for the design and conduct of the study, had access to the data, and agreed upon the decision to publish. This work has not been published before or concurrently submitted to another journal for publication. The manuscript has been significantly reworked in light of the editors and reviewers' comments and we hope that you find the manuscript much improved.

Caroline Presley on behalf of the Authors:

Reviewer 1
1. The authors in the discussion highlight the low sensitivity observed. The correctly explain why this is less of an issue for studies of comparative effectiveness with HF as an outcome. I note that this is in contrast to the study by Floyd et al using VA data where sensitivity was 90% for HF (albeit chronic heart failure). It could be of interest to the readership to also include in the discussion the reason for the low sensitivity observed in this study.

The study completed by Floyd et al reported their algorithm had a 90% sensitivity for chronic (prevalent) heart failure; their algorithm relied on the presence of one ICD-9 code for heart failure recorded in the inpatient or the outpatient setting in the preceding 12 to 24 months. We believe the lower sensitivity in our study compared with that study may be due to the stringent criteria for the heart failure algorithm as a clinical outcome; 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 Framingham criteria to adjudicate potential heart failure events. Nevertheless, as noted in our results page 11 lines 16-17 we achieve an improved sensitivity when we expand the algorithm to include ICD-9 codes as one of the first five diagnosis codes. We have added a comparison to the Floyd paper in the discussion page 12 lines 8-14.

2. Why did the authors not include ICD 10 codes? Was it possible to include these? If so there might be some added benefit as it may increase the generalizability to other settings where these are utilized.

Validation of ICD-10 codes for heart failure is important future work, however the Veterans Health Administration system did not transition to ICD-10 until after the completion of our study period, October 1, 2015. Therefore, ICD-10 codes were not included in the validation.

3. The authors use ejection fraction to classify HF hospitalizations. This can be collect up to a year prior to the HF event. How many of these values where collected at hospitalization? If for many they were collect before the event is it possible there is some misclassification here?

Of the patients who had a confirmed HF hospitalization and an EF measurement, 55.8% had an EF measurement during the study hospitalization; the remainder had an EF measurement present in the year prior to study hospitalization. We have added this to the manuscript results on page 9 lines 17-20 and acknowledged this limitation in the discussion section in page 14 lines 9-11.

Reviewer 2

1. In abstract, page 2, participants section: Description of diabetic patients is missing.

Thank you for bringing this to our attention. We have added the inclusion criteria of patients to the participants section of the abstract (bolded below).
"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."

2. Methods, study population, page 5: Authors mentioned about inclusion criteria of "visit at least once every 180 days" and "were diagnosed with diabetes (at least one prescription filled for an antidiabetic medication) between 2001 and 2008". Does it means any visit – including IP, ED or outpatient visit? Why selected population with diagnosis of diabetes between 2001-2008 instead of 2001-2012?

A visit could be an outpatient encounter, emergency department visit, or a hospitalization. We have clarified this language in the text to state that inclusion criteria include "an outpatient encounter, emergency department visit, hospitalization, or medication refill at least once every 180 days" on page 5 lines 14-15.
The underlying national patient cohort included patients who were initiated on an antidiabetic medication between 2001 and 2008; we did not have ongoing addition of patients to the cohort after 2008. For the patients in the underlying cohort, follow up data for these patients was available through 2012 which allowed for sampling of hospitalizations through 2012. We have added language to clarify this in the methods section on page 5 lines 10-12.

3. Methods, study events, page 5: Can authors explain if there was any particular reason for choosing 4:1 ratio for algorithm positive to algorithm negative criteria?

We sampled a greater number of algorithm positive cases because we prioritized estimating the PPV with precision. We have added the following line to the manuscript methods in page 5 line 26 to page 6 lines 1-2.
The 500 patients were sampled with a 4:1 algorithm positive:negative ratio to allow measuring PPV with greater precision. This strategy also allowed calculation of sensitivity and negative predictive value.

4. There were > 10,000 eligible hospitalization to select 500 sample hospitalization. Why authors included individuals with multiple hospitalization more than once instead of using 500 unique individuals with hospitalization.

Thank you for raising this point as it was a subtle, but important part of the design. We have added the following line on page 6 lines 9-13 to the manuscript.
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 the entire population.

5. Results, page 8, 1st sentence: "Of 10,766 eligible hospitalizations in TVHS between 2001 and 2012, a total of 500 hospitalizations were sampled." Please clearly describe if 10,766 eligible hospitalization due to diabetic population or not? And what is n of unique individuals among 10,766 and 500 sampled hospitalization?

The reviewer raises an important point. This local cohort is derived from our national cohort of patients with diabetes who started oral hypoglycemic medications (N=411,055 veterans). The local

data sample of 10,766 in the diabetes cohort thus could have had a diabetes medication started before or after this independent selection of the study hospitalization. We have revised the text on page 5 lines 10-12 in the methods to more clearly explain that this sample is derived from a national cohort of patients with diabetes. In our patient sample, 430 of 497 patients (86.5%) had diagnosis of diabetes documented in the two years prior to the study hospitalization. Thus, the majority of the population has diabetes and the results of the algorithm are most reflective of a population with diabetes. This is included in the results section on page 9 lines 9-11.

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). This text has been added to page 8 line 24-26 in the results.

6. Results, page 8, 2nd para: authors mentioned that majority of the patients were older adults, aged >=65 years but as showed in Table 2, 48% patients were aged <65 years old.
Agreed, the distribution of patients is approximately half < 65 years and half >=65 years. We have revised our results description to more accurately represent the age distribution of patients in our study. See revised results page 9 line 5-6 noting that the median age is 65 with the IQR of (58, 75) .

7. Minor comments:
a. Table 4: It would be interesting to see values of PPV, NPV, sensitivity and specificity by any ICD-9 code and any DRG code.
We present the PPV, NPV, sensitivity, and specificity for hospitalizations meeting only DRG code criteria for the algorithm; this is presented in Table 4. Because of the limited number of these hospitalizations (n=8), we were unable to examine the two possible DRG codes separately. We provide the PPV for individual ICD-9 codes in Appendix Table A1. We were unable to calculate the NPV, sensitivity, specificity for individual ICD-9 codes because these calculations require inclusion of algorithm-negative hospitalizations. If an included hospitalization had an ICD-9 code from the HF algorithm present in the primary diagnosis code position, this hospitalization was defined as an algorithm-positive hospitalization.

b. Results, page 11, line 10: there was a typo error in sensitivity numbers. It should be 45.1% instead of 41.5%.
Thank you for bringing this to our attention. We have corrected this in the text.

c. In discussion, page 11, para 1, lines 18-26: Use abbreviation of DRG, PPV, NPV.
We have made these edits.

Sincerely,

Caroline Presley, M.D., M.D.
GRECC, VA Tennessee Valley Healthcare System and
Section of Internal Medicine-Pediatrics, Vanderbilt University Medical Center
1310 24th Avenue South
Nashville, TN 37212
Phone 615-873-8012; Fax 615-873-7241
Caroline.a.presley@vanderbilt.edu; Caroline.presley@va.gov

VERSION 2 – REVIEW

REVIEWER	Dr Blánaid Hicks Centre for Public Health, Queen's University Belfast, N. Ireland, UK
REVIEW RETURNED	15-Jan-2018

GENERAL COMMENTS	I thank the authors for their responses to my queries. The authors have answered these clearly and added, where appropriate, further information to the manuscript. Therefore I am happy to recommend this study for publication.