Editorial

The Quality Chasm Between Administrative Coding and Accurate Phenotyping of Heart Failure

SUMEET PAWAR, MD,[1] TARIQ AHMAD, MD, MPH,[1,2] AND NIHAR R. DESAI, MD, MPH[1,2]

New Haven, Connecticut

Administrative claims data are routinely gathered as part of clinical care, primarily for purposes of billing, and they include sociodemographic information, health care resource utilization and limited but salient clinical data. This information is used by providers, payers, health services researchers, and policymakers for diverse purposes, including disease surveillance, quality improvement programs and understanding variations in outcomes and potential disparities in care. The easy accessibility of claims data has made it a pivotal gauge for local and national performance-improvement campaigns as well as evaluation of health care policy initiatives. Alongside these factors, there has been a growing interest in pragmatic trials that leverage information collected as part of routine practice so as to randomize patients with heart failure (HF) and generate new knowledge in a cost-effective manner.[1]

Despite its widespread use to address clinical questions and impact health care policy, there has been an ongoing concern about the accuracy and validity of administrative data as they pertain to describing HF. A key conviction underlying the use of administrative data is their ability to identify accurately and consistently patients with HF and classify them as having preserved vs reduced ejection fraction. The expectation, therefore, would be that these codes have clear and direct relationships to patients' measured ejection fractions according to echocardiography. The ability to perform such a "validation" would be foundational scholarship, with wide-ranging implications for hospitals and health systems, professional organizations, clinical researchers, and policymakers.

It is, therefore, timely that this issue of *Journal of Heart Failure* includes an analysis by Heidenreich and colleagues that examines the accuracy of administrative coding using the International Classification of Diseases, Ninth Revision (ICD-9) codes for identifying systolic and diastolic HF in 43,044 HF hospitalizations from 114 Veterans Affairs (VA) hospitals across the country between 2006 and 2013. The administrative codes were verified by comparing them with left ventricular ejection fractions (LVEF) obtained at the same admission. Quite surprisingly, about 1 of 3 patients with hospitalization for HF had specific codes for systolic (18%) or diastolic (17%) HF; the majority (65%) were classified as "not otherwise specified." Furthermore, the sensitivity of the systolic HF code for correctly identifying patients with LVEF <40% was low, at 29%. Even more dismal was the positive predictive value of 76%, which translated to a quarter of the patients labeled as having systolic HF by administrative codes not having reduced ejection fractions (EF). Similar findings were noted for the diastolic HF codes, with a sensitivity of 29% and a positive predictive value of 78%. The authors noted marginal improvements in sensitivity and positive predictive value changes in cut points toward more extreme values of LVEF for both systolic and diastolic HF.

Another key finding of the study was that there was substantial variation in the sensitivity of administrative codes for systolic HF, ranging from 0 to 81% across hospitals. The temporal trends for the accuracy of the codes between 2009 and 2013 demonstrated improvement in sensitivity for both systolic and diastolic HF; the positive predictive value worsened for systolic HF but improved for diastolic HF. A multivariate analysis did not show any association between the HF codes and 1-year mortality after adjusting for LVEF.

Patients with HF are split evenly, half having preserved ejection fraction (HFpEF) and the other half having reduced ejection fraction (HFrEF).[2] However, there appears to be a decline in the incidence rates of HFrEF and an associated increase in HFpEF.[3] The phenotypic heterogeneity of HFpEF further complicates accurate identification of this subtype and allows for the potential of erroneously including other causes of dyspnea. The risk of developing either type of HF varies significantly by age, sex, race, and history of coronary artery disease.[4] Although there are guideline-

From the [1]From the Section of Cardiovascular Medicine, New Haven, Connecticut and [2]Center for Outcomes Research and Evaluation (CORE) at Yale University School of Medicine, New Haven, Connecticut.

Manuscript received April 18, 2019; revised manuscript accepted April 18, 2019.

Reprint requests: Nihar Desai, MD, MPH, 333 Cedar Street, Yale University School of Medicine, New Haven, CT 06510. E-mail: nihar.desai@yale.edu

See page 492 for disclosure information.

1071-9164/$ - see front matter

https://doi.org/10.1016/j.cardfail.2019.04.010

directed medical therapies that result in improvements in morbidity, mortality and quality of life for HFrEF, the same cannot be said about HFpEF. Therefore, it is of concern that in this large study from the VA, two-thirds of the patients did not have coding diagnoses of either systolic or diastolic HF; instead, the unspecified ICD codes for HF were used, despite clear documentation of ventricular function in the electronic health record system. This category of not otherwise specified (NOS) does not have any clinical relevance or any targeted therapies yet constitutes the largest group of patients in the current study. Presumably, this cohort may not receive appropriate clinical-decision alerts, referral to disease management or guideline-based therapies.

These results should be considered in the light of several limitations. It is important to note that the VA population consists mainly of males and does not help us to understand HF in women. Second, given the non-diagnosis-related group-based payment model at the VA, the financial pressures that would incentivize private hospitals to improve the coding of HF do not exist at the VA. Third, this study included only patients with principal diagnoses of HF, therefore, the results need to be taken in the context of this assumption. The prevalence of HF has increased, so the number of hospitalizations with primary diagnoses of HF has decreased, while those with secondary diagnoses have increased or remained stable.[5] Even more perplexing is why "trained abstractors" at the VA misclassify HF codes despite knowing the LVEFs. The authors do not provide any answers regarding the reasons for misclassification, but one reason for this might be that direct chart review is not performed routinely by the abstractors.

With these considerations, what are the implications of this study? First, the accuracy of ICD-9 HF codes in the VA population appears to be too poor for use in performance measurement and quality-improvement work. There are some data to suggest that these findings might extend beyond the VA, but this needs to be confirmed by definitive studies[6,7] because we need to be able to monitor the quality of health failure care and disease incidence and prevalence as well as undertake efforts to address disparities and unwarranted variations in care practices and outcomes. If the administrative data that have been the cornerstone of these efforts are of insufficient sensitivity and specificity, patients, providers, payers, and policymakers will lose. Of note, this is in contrast to the case of myocardial infarction, where ICD codes among Medicare beneficiaries show a sensitivity of greater than 90% when verified by direct chart reviews.[8] Second, there is substantial variation in the quality of administrative coding. There are hospitals where the sensitivity and specificity of the coding for HF more closely reflect clinical reality. Understanding what resources, oversight mechanisms, operational infrastructure, and enabling structures contribute to improved performance and whether they can be spread and scaled to other institutions may help in the short and intermediate terms. Third, the number of "not otherwise specified" codes is alarming. The total number of coding diagnoses in the ICD-9 handbook is 13,000, but it is now much higher with ICD-10 (68,000 codes). This increase in the number of codes was thought to improve specificity and provide granular representation of various disease states. For any clinician who has attempted to code a diagnosis, the number of NOS codes is striking and occurs at the cost of redundancy. Finally, the results question the reliance of health services research on claims data. The current study is a cautionary tale; health services research is only as good as the quality of the data themselves, and the same applies even if sophisticated statistical methods such as machine-learning tools are deployed to analyze administrative data.[9] These data illustrate the importance of and illuminate the potential for data derived directly from the electronic medical records to be a source of "truth," especially in the intermediate and longer term timeframes. Instead of relying on ICD codes for identifying disease states, we can leverage higher dimensional data that are collected in the electronic medical records and include, but are not limited to, the temporal trends in LVEF, N-terminal pro−B-type natriuretic peptide and concurrent utilization of guideline-determined medical therapy to assess the quality of HF care and health-system performance.

Natural language processing is a tool that has the potential to identify accurately patients with HFrEF and HFpEF and to assess variations in care patterns and opportunities for performance improvement. With this approach, it is possible to extract the LVEF directly from the medical records rather than relying on human abstractors.[10,11] Furthermore, Amazon Web Services has recently announced a HIPAA-compliant machine-learning service named Amazon Comprehend Medical; it is capable of parsing information in the medical records, such as patient diagnoses, treatments and more.[12] Such tools have the potential to deliver rapid, scalable performance measurement and provide real-time feedback to providers, administrators and even policymakers. As our understanding of these tools continues to evolve, there is immense potential for changing practice that is not limited to accurate identification of HFrEF and HFpEF but, instead, phenotyping HF beyond ejection fraction by integrating higher dimensions of data that include information derived from electronic medical records, biosensors, "omics," as well as patient-reported outcome measures.[13] Given the significant clinical heterogeneity of HFpEF, the ability to group pathophysiologically similar phenotypes can help us to identify patients who respond to targeted therapies.[14] The VA, with its computerized patient record system (CPRS) that includes granular information about individual patients, is uniquely poised to leverage the multidimensional, longitudinal data recorded in its electronic medical records for health services research if it is done in a more systematic manner. Certainly, the current system is neither accurate nor precise in its delineation of HF. If we are to bring precision medicine to our patients, we need to go beyond contemporary claims data. The goal—a crucial and foundational step toward improving HF care—is to

close the massive chasm between what is captured in administrative data and what is reality.

Disclosures

No conflicts of interest to declare.

References

1. Ford I, Norrie J. Pragmatic trials. N Engl J Med 2016;375:454–63.
2. Dunlay SM, Roger VL, Redfield MM. Epidemiology of heart failure with preserved ejection fraction. Nat Rev Cardiol 2017;14:591–602.
3. Tsao CW, Lyass A, Enserro D, Larson MG, Ho JE, Kizer JR, et al. Temporal trends in the incidence of and mortality associated with heart failure with preserved and reduced ejection fraction. JACC Heart Fail 2018;6:678–85.
4. Pandey A, Omar W, Ayers C, LaMonte M, Klein L, Allen NB, et al. Sex and race differences in lifetime risk of heart failure with preserved ejection fraction and heart failure with reduced ejection fraction. Circulation 2018;137:1814–23.
5. Blecker S, Paul M, Taksler G, Ogedegbe G, Katz S. Heart failure-associated hospitalizations in the United States. J Am Coll Cardiol 2013;61:1259–67.
6. Bovitz T, Gilbertson DT, Herzog CA. Administrative data and the philosopher's stone: Turning heart failure claims data into quantitative assessment of left ventricular ejection fraction. Am J Med 2016;129:223–5.
7. McCormick N, Lacaille D, Bhole V, Avina-Zubieta JA. Validity of heart failure diagnoses in administrative databases: a systematic review and meta-analysis. PLOS ONE 2014;9:e104519.
8. Kiyota Y, Schneeweiss S, Glynn RJ, Cannuscio CC, Avorn J, Solomon DH. Accuracy of Medicare claims-based diagnosis of acute myocardial infarction: estimating positive predictive value on the basis of review of hospital records. Am Heart J 2004;148:99–104.
9. Miller PE, Pawar S, Vaccaro B, McCullough M, Rao P, Ghosh R, et al. Predictive abilities of machine learning techniques may be limited by dataset characteristics: insights from the UNOS database. J Card Fail 2019;25:479–83.
10. Kim Y, Garvin JH, Goldstein M, Hwang T, Redd A, Bolton D, et al. Extraction of left ventricular ejection fraction information from various types of clinical reports. J Biomed Inform 2017;67:42–8.
11. Meystre SM, Kim Y, Gobbel GT, Matheny ME, Redd A, Bray BE, et al. Congestive heart failure information extraction framework for automated treatment performance measures assessment. J Am Med Inform Assoc 2017;24:e40–e6.
12. Introducing medical language processing with Amazon Comprehend Medical. https://aws.amazon.com/blogs/machine-learning/introducing-medical-language-processing-with-amazon-comprehend-medical/. Accessed 03/01/2019.
13. Ahmad T, Wilson FP, Desai NR. The trifecta of precision care in heart failure: biology, biomarkers, and big data. J Am Coll Cardiol 2018;72:1091–4.
14. Ahmad T, Lund LH, Rao P, Ghosh R, Warier P, Vaccaro B, et al. Machine learning methods improve prognostication, identify clinically distinct phenotypes, and detect heterogeneity in response to therapy in a large cohort of heart failure patients. J Am Heart Assoc 2018;7(8). https://doi.org/10.1016/j.cardfail.2019.01.018.