# Promising Administrative Measures of Heart Failure and Future Directions

**Natalia Festa, MD, MBA**[1,2], **Jason H. Wasfy, MD, MPhil**[3,4], **Lidia MVR Moura, MD, PhD, MPH**[5,6]

[1.]National Clinician Scholars Program at Yale University, New Haven, Connecticut.

[2.]Department of Internal Medicine, Yale School of Medicine, New Haven, Connecticut.

[3.]Harvard Medical School, Boston, Massachusetts.

[4.]Cardiology Division, Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts.

[5.]Department of Neurology, Massachusetts General Hospital, Boston, Massachusetts.

[6.]Department of Neurology, Harvard Medical School, Boston, Massachusetts.

In the United States, heart failure affects approximately 6.5 million adults and its prevalence is expected to increase with population aging.[1] Unfortunately, there are not yet standardized approaches to describe and track its epidemiological burden within the Medicare population.[1] Diagnosis codes within Medicare claims are often used to identify heart failure cases, clinical subtypes, and associated outcomes for purposes including population health management, reimbursement policy, and pharmacoepidemiologic research.[2,3] Because the accuracy of administrative diagnosis codes may vary with clinical and policy conventions, their performance characteristics require ongoing evaluation.[4] While clinical databases are not well-suited to describe national trends, national claims often lack the granularity to describe clinical and epidemiologic trends accurately. This challenge is common to other diseases and syndromes.[5]

The linked study by Bates and colleagues evaluates the accuracy with which International Classification of Diseases, 10th revision (ICD-10) diagnosis codes from inpatient episodes identify acute heart failure hospitalizations and clinical subtypes (e.g., heart failure with preserved ejection fraction).[6] The investigators randomly sampled the medical records of 200 traditional Medicare beneficiaries who were hospitalized with a principal or secondary (second position, only) heart failure diagnosis on inpatient administrative claims between October 2015 and December 2017. They evaluated the ability of principal and secondary heart failure ICD-10 codes to identify acute heart failure hospitalization against a primary reference standard of a recorded heart failure diagnosis by a treating physician in the medical record. They repeated their validation using additional reference standard

Correspondence: Natalia Festa, MD, National Clinician Scholars Program at Yale University, 333 Cedar Street, PO Box 208088, New Haven, CT 06510, natalia.festa@yale.edu.

**Conflict of Interest Disclosures**: The authors report no conflicts of interest.

definitions for heart failure hospitalization, including the Modified Framingham Heart Failure Criteria and adjudication by a cardiovascular fellow. They next validated these select ICD-10 diagnosis codes against reference standard definitions for heart failure subtypes.

Bates and colleagues demonstrated that principal ICD-10 diagnosis codes identify acute heart failure hospitalizations with excellent positive predictive value (point estimate 0.98, 95% confidence interval 0.95–1.00) against the primary reference standard of a recorded heart failure diagnosis by a treating physician in the medical record. The positive predictive value for secondary ICD-10 diagnosis codes was poor (0.66, 0.58–0.74) and attenuated predictive values for the combination of primary and secondary diagnosis codes (0.76, 0.70–0.82). The combined (0.60, 0.53–0.66) and separate positive predictive values for primary and secondary ICD-10 diagnosis codes were poorest when validated against the Modified Framingham Heart Failure Diagnostic criteria. The investigators also observed an excellent positive predictive value for ICD-10 heart failure diagnosis codes against the primary reference standard for systolic dysfunction, using an ejection fraction threshold of 50% (0.90, 0.82–0.98), but this declined considerably when using an ejection fraction threshold of 40% (0.72, 0.60–0.85). The observed positive predictive value of ICD-10 diagnosis codes for diastolic dysfunction was excellent against the primary reference standard (0.92, 0.95–1.00), whereas positive predictive values were fair-to-poor for persons with mixed systolic and diastolic dysfunction or unspecified heart failure.

Prior to this study, information regarding the accuracy of heart failure diagnosis codes was limited to those within the 9th revision of the ICD (ICD-9). Generally, these studies have reported poorer accuracy.[7,8] The improved performance of ICD-10 codes demonstrated by Bates and colleagues could be explained by the authors' decision to restrict their sample to beneficiaries with a first- or second-position diagnostic code. Restricting the codes in this way may increase the prevalence of heart failure within their sample and select for persons who are likelier to have true heart failure. The reference standards used by Bates and colleagues may have also affected the prevalence of adjudicated heart failure within the study sample, as compared to those used in ICD-9 validations studies, such as the Modified Framingham criteria or standalone ejection fraction cutpoints.[7,8] Apart from methodological differences, evolving payment incentives and clinical and administrative conventions may have improved the accuracy of heart failure ICD-10 diagnosis codes.

It is important to consider that positive predictive value, the performance characteristic reported by Bates and colleagues, is tied to underlying disease prevalence. Because the sample selection criteria for this study yielded beneficiaries with a high pretest probability of heart failure, positive predictive values for ICD-10 codes may appear more favorable than if tested within the general traditional Medicare population. Because single measures of accuracy can be misleading,[9] it would be ideal for future validation studies to report multiple performance characteristics, including negative predictive value, sensitivity, specificity, and calibration metrics. As an increasing proportion of beneficiaries enroll in Medicare Advantage, it will also be useful to understand how time-varying reimbursement incentives, as well as differences in population characteristics and care management,[10] could affect the accuracy of heart failure diagnostic claims from inpatient encounters within privately managed plans.

The performance of ICD-10 claims differentiating heart failure presentations and outcomes has multiple important applications. For example, the high positive predictive values with which ICD-10 codes differentiated systolic and diastolic dysfunction should be useful to provider organizations seeking to improve population health management. As goal-directed medical therapy differs for heart failure subtypes, health systems may be able to use validated ICD-10 codes to improve guideline-concordant care within their patient populations. The poor positive predictive values for clinically ambiguous presentations, including combined systolic and diastolic dysfunction, are consistent with prior studies that have suggested poor detection of heart failure with mid-range or mildly reduced ejection fraction.[11]

That Bates and colleagues observed attenuated positive predictive values for secondary diagnosis codes is relevant to researchers, provider organizations, and payers using administrative information to identify trends in heart failure and heart failure outcomes. This finding complements recent evidence that metadata (data describing the characteristics of administrative claims) regarding the setting or frequency of diagnostic claims affect their validity and accuracy.[5] Further validation of claims-based algorithms for heart failure should test whether accuracy is improved by incorporating metadata regarding claim position and the setting in which claims were assigned.

Bates and colleagues raise salient considerations regarding validation methodology. Despite the utility of real-world evidence,[12] it is susceptible to bias when both the validated and reference standard definitions are derived from real-world data sources. As noted by the study authors, heart failure diagnoses are clinical and, therefore, susceptible to heterogeneity in presentation, diagnosis, and management. These uncertainties affect the quality of the reference standard definitions from medical record abstraction and may be heightened by variations in the quality of clinical documentation. These factors could explain the poorer interrater reliability and performance of ICD-10 codes when validated against the Modified Framingham Heart Failure criteria, which may have greater specificity for clinical heart failure as compared to the broader primary reference standard. These discrepancies in the accuracy of ICD-10 codes against distinct reference standards suggest that the choice of reference standard is important to the reproducibility of future validation studies.

Validated approaches to improving the identification of heart failure subtypes in administrative claims are essential to both population health management and research. Bates and colleagues make an important and timely contribution by addressing knowledge gaps regarding the accuracy with which ICD-10 diagnosis codes identify common heart failure presentations and outcomes among Medicare beneficiaries with primary or secondary heart failure diagnosis codes on hospital claims. Their study also provides useful insight into the effects of reference standard selection when using real-world evidence, while providing a valuable framework for future studies to validate administrative measures against multiple reference standard permutations. Future research should extend these initial findings and investigate multiple performance characteristics of administrative claims across populations that vary with respect to their underlying heart failure prevalence and health insurance plans, among other characteristics. As Bates and colleagues demonstrated the importance of claim position to the performance of heart failure ICD-10 diagnosis codes, future work should also

incorporate additional dimensions of metadata, including care setting and the frequency of heart failure claims.

## Acknowledgements

## References

1. Jackson SL, Tong X, King RJ, Loustalot F, Hong Y, Ritchey MD. National Burden of Heart Failure Events in the United States, 2006 to 2014. Circ Hear Fail. 2018;11(12):e004873. doi:10.1161/CIRCHEARTFAILURE.117.004873

2. Desai RJ, Mahesri M, Chin K, et al. Epidemiologic Characterization of Heart Failure with Reduced or Preserved Ejection Fraction Populations Identified Using Medicare Claims. Am J Med. 2020;134(4):e241–e251. doi:10.1016/J.AMJMED.2020.09.038 [PubMed: 33127370]

3. Patorno E, Pawar A, Franklin JM, et al. Empagliflozin and the Risk of Heart Failure Hospitalization in Routine Clinical Care. Circulation. 2019;139(25):2822–2830. doi:10.1161/CIRCULATIONAHA.118.039177 [PubMed: 30955357]

4. Festa N, Price M, Weiss M, et al. Evaluating The Accuracy Of Medicare Risk Adjustment For Alzheimer's Disease And Related Dementias. Health Aff. 2022;41(9):1324–1332. doi:10.1377/hlthaff.2022.00185

5. Festa N, Price M, Moura LMVR, et al. Evaluation of Claims-Based Ascertainment of Alzheimer Disease and Related Dementias Across Health Care Settings. JAMA Heal Forum. 2022;3(4):e220653. doi:10.1001/JAMAHEALTHFORUM.2022.0653

6. Bates BA, Akhabue E, Akhabue M, Mukherjee A, Hiltner E, Rock J, Wilton B, Mittal G, Visaria A, Rua M et al. Validity of ICD-10 Diagnosis Codes for Identification of Acute Heart Failure Hospitalization and Heart Failure with Reduced vs. Preserved Ejection Fraction in a National Medicare Sample. Circ Cardiovac Qual Outcomes. 2022;In Press. doi:10.1161/CIRCOUTCOMES.122.009078

7. Presley CA, Min JY, Chipman J, et al. Validation of an algorithm to identify heart failure hospitalisations in patients with diabetes within the veterans health administration. BMJ Open. 2018;8(3):20455. doi:10.1136/BMJOPEN-2017-020455

8. Desai RJ, Lin KJ, Patorno E, et al. Development and Preliminary Validation of a Medicare Claims-Based Model to Predict Left Ventricular Ejection Fraction Class in Patients With Heart Failure. Circ Cardiovasc Qual Outcomes. 2018;11(12):e004700. doi:10.1161/CIRCOUTCOMES.118.004700 [PubMed: 30562067]

9. Adhikari S, Normand SL, Bloom J, Shahian D, Rose S. Revisiting Performance Metrics for Prediction with Rare Outcomes. Stat Methods Med Res. 2021;30(10):2352. doi:10.1177/09622802211038754 [PubMed: 34468239]

10. Figueroa JF, Blumenthal DM, Feyman Y, et al. Differences in Management of Coronary Artery Disease in Patients With Medicare Advantage vs Traditional Fee-for-Service Medicare Among Cardiology Practices. JAMA Cardiol. 2019;4(3):265–271. doi:10.1001/JAMACARDIO.2019.0007 [PubMed: 30785590]

11. Savarese G, Stolfo D, Sinagra G, Lund LH. Heart failure with mid-range or mildly reduced ejection fraction. Nat Rev Cardiol 2021 192. 2021;19(2):100–116. doi:10.1038/s41569-021-00605-5 [PubMed: 34489589]

12. Sherman RE, Anderson SA, Dal Pan GJ, et al. Real-World Evidence — What Is It and What Can It Tell Us? N Engl J Med. 2016;375(23):2293–2297. doi:10.1056/NEJMSB1609216/SUPPL_FILE/NEJMSB1609216_DISCLOSURES.PDF [PubMed: 27959688]