An evidence-based data science perspective on the prediction of heart failure readmissions

Kenneth J. Locey^{1*}, Thomas A. Webb^{1,2**}, Bala Hota^{1,2,3***}

Affiliations

¹ Center for Quality, Safety and Value Analytics, Rush University Medical Center, Chicago, Illinois, 60612, USA.

² Knowledge Management Services, Rush University Medical Center, Chicago, Illinois, 60612, USA.

³ Department of Internal Medicine, Division of Infectious Diseases, Rush Medical College, Chicago, Illinois, 60612, USA.

* Corresponding author. email: Kenneth_J_Locey@rush.edu

Coauthor Contact emails

** Thomas_A_Webb@rush.edu

*** Bala_Hota@rush.edu

ABSTRACT

The prevention of unplanned 30-day readmissions of patients discharged with a diagnosis of heart failure (HF) remains a profound challenge among hospital enterprises. Despite the many models and indices developed to predict which HF patients will readmit for any unplanned cause within 30 days, predictive success has been meager. Using simulations of HF readmission models and the diagnostics most often used to evaluate them (C-statistics, ROC curves), we demonstrate common factors that have contributed to the lack of predictive success among studies. We reveal a greater need for precision and alternative metrics such as partial C-statistics and precision-recall curves and demonstrate via simulations how those tools can be used to better gauge predictive success. We suggest how studies can improve their applicability to hospitals and call for a greater understanding of the uncertainty underlying 30-day all-cause HF readmission. Finally, using insights from sampling theory, we suggest a novel uncertainty-based perspective for predicting readmissions and non-readmissions.

Over the past decade, hospitals have been motivated to decrease 30-day all-cause readmissions among heart failure (HF) patients.¹ These HF readmissions (HFR) are often deemed preventable and failing to decrease them carries financial penalties.¹ Consequently, to aid hospital enterprises in reducing HFR, hundreds of studies have used or developed indices and models to predict HFR, 30-day all-cause or otherwise.²⁻⁴ These HFR studies have varied greatly in their readmission time frames (30d – 1yr) and methodologies (risk indices, statistical models, machine learning), as well as in the data underpinning their predictions (clinical, administrative, psychosocial).⁴⁻⁸ However, the degrees of performance and applicability needed to make transformative progress in predicting HFR have been notoriously difficult to achieve^{4,9}, begging the question of why.

While the variation in methods, data, and performance among HFR studies has been thoroughly reviewed by others^{3,4,10}, few explanations have been given for the shared lack of success or the commonalities that have contributed to it. In the current work, we identify and discuss common factors contributing to the difficulty in predicting HFR, 30-day all-cause or otherwise. We demonstrate how the choice of diagnostic metrics, data, and models can not only impede progress but also prevent the development of useful tools and actionable solutions. In each case, we suggest alternatives and potential paths forward. Finally, we recast the prediction of HFR with an uncertainty-based perspective.

ROC curves and C-statistics — HFR studies often demonstrate their success via receiver operating characteristic (ROC) curves.^{4,7,10,11} ROC curves relate the true positive rate (TPR) to the false positive rate (FPR) as a classifier is applied across diagnostic thresholds (Table 1, Fig 1a). In HFR studies, thresholds are often measures of risk or probabilities above which an index discharge is classified as resulting in readmission. ROC curves represent the trade-off between correctly classifying discharges that result in readmission and incorrectly classifying discharges that do not. To quantify their ROC results, HFR studies use C-statistics (aka C-index) representing the area under the ROC curve (AUROC).^{4,7,10,11} These C-statistics range from 0 to 1, with values near 0.5 being no better than random and values greater than 0.9 signifying

outstanding diagnostic power (Fig 1a).¹² For context, C-statistics (i.e., AUROC) for 30-day all-cause HFR studies typically range between 0.5 and 0.7.^{7,10,11,13-18}

ROC curves and C-statistics provide a standard way to report results and compare studies. However, these metrics provide an incomplete view of diagnostic success and can hide diagnostic failures. For example, only a small portion of the ROC curve may be useful because only a small range of FPR may be acceptable.¹⁹ Among HFR studies, FPR often exceeds 0.25 before the TPR exceeds 0.5.^{11,20,21} Simply put; the false alarm rate often reaches 25% before 50% of discharges resulting in readmission are correctly classified. If an FPR greater than 0.25 were unacceptable, then a C-statistic or AUROC based on the full ROC curve would be misleading. Consequently, HFR studies may benefit from partial AUROC, i.e., the area under the ROC curve corresponding to an acceptable range of FPR (Fig 1b). Partial AUROC measures and corresponding partial C-statistics have been developed by others, included into analytical libraries, and have been used in other healthcare studies.^{19,22}

Whether using partial C-statistics for a maximally acceptable FPR or not, there is a more deeply concerning detriment to using ROC curves as a standard of performance. Specifically, ROC curves do not reveal the probability of making a correct positive prediction (i.e., precision) and hence, whether a prediction for readmission can be trusted. Likewise, the degree of diagnostic failure obscured by ROC curves can grow as data become imbalanced, i.e., as non-readmissions outnumber readmissions. We address these issues below.

Precision and imbalanced data — Also known as the positive predictive value (PPV), precision is the fraction of positive predictions that were actually correct (Table 1). Surprisingly, many HFR studies never mention or measure precision.^{2-4,7,8,13-16,18,20,23-27} However, without precision there is little way to establish confidence in whether or not a prediction for readmission can be trusted. Among the comparatively few studies that report it, precision typically ranges from 0.09 to 0.44, meaning that 56 – 91% of predictions for readmission are often incorrect.^{3,21,24,28,29} If hospitals acted on the results of such predictions, they would invest their limited time and resources into expectations that are, more often than not, false alarms.

Precision and TPR are often inversely related; an increase in one will produce a decrease in the other.¹⁹ Consequently, if HFR studies give greater attention to precision, they will inevitably find that models with impressive ROC curves can be highly imprecise. Simulations

can reveal how this occurs (Fig 1a-c, Appendix 1). In the analyses behind Figure 1, applying a hypothetical high performing model to simulated HFR outcomes produced an AUROC of 0.92 and a TPR of 0.9 at a maximally allowed FPR of 0.25 (Fig 1b). If this were a real model applied to real HFR data, the result would be exemplary among HFR studies. Yet, the corresponding precision of this model was only 47% (Fig 1c). Another hypothetical model designed to be similar in performance to real HFR models (AUROC = 0.66) produced a TPR of 0.47 at an FPR of 0.25, with a precision of only 32% (Fig 1a-c).

The tradeoff between accurately capturing readmissions and making reliable predictions of readmission (TPR vs. precision) is one complicating factor for predictive HFR studies. The other factor is class imbalance, i.e., large differences in the numbers of readmissions and non-readmissions. In HFR studies, 30-day all-cause readmissions typically comprise less than 25% of HF index discharges.^{2,7,13,17,18,23} This class imbalance poses challenges for machine learning algorithms and some HFR studies have taken measures to correct for it.^{3,17} However, class imbalance also has a less frequently acknowledged effect. Specifically, if non-readmissions greatly outnumber readmissions, then ROC curves will fail to reflect how greatly false positives outnumber true positives, causing precision to suffer in a way that ROC curves and their associated metrics cannot capture (Fig 2).

In addition to giving greater attention to precision and class imbalance, HFR studies may benefit from incorporating precision-recall curves (PRCs) (Fig 1c, Fig 2a-b). PRCs reveal the relationship between precision and TPR, thereby capturing the tradeoff between correctly classifying discharges that result in readmission and avoiding false positives. PRCs are recommended over ROC curves for use with imbalanced data and can be summarized by calculating the area under the curve (AUPRC) (Fig 1c).^{19,30,31} Noting this, a small number of HFR studies have used PRCs and reported their respective AUC values.^{30,31} As an example of the insight gained by using PRCs, we applied a hypothetical model to two data sets of simulated HFR outcomes (Fig 2a-b, Appendix 1). These two data sets only varied in their ratios of non-readmissions to readmissions (1:1, 4:1). While the resulting ROC curves were identical, PRCs reveal disparate degrees of performance.

HFR studies can provide greater insight and pursue greater rigor when developing, evaluating, and comparing predictive models by 1) using PRC's in combination with both ROCs and partial AUROC and partial C-statistics, 2) examining a greater diversity of diagnostic

measures, and by 3) establishing standards for a minimum TPR (e.g., 0.5) and a maximum FPR (e.g., 0.25). In doing so, HFR studies could attempt to optimize precision and TPR, while not exceeding an acceptable FPR.

Meeting the needs of healthcare providers — Ideally, predictive HFR studies would be guided by the abilities of providers to readily apply an approach within time sensitive needs and without excluding patients. However, HFR studies are often vulnerable to biases, include biases in model methodology and validation, data attrition and patient exclusion, reliability in new populations, ease of implementation in clinical settings and patient management.^{4,32} Predictive HFR studies are also often retrospective, i.e., using data that are only available after the point of intervention has passed.^{16,32} Consequently, without considering the needs and limitations of healthcare providers, HFR studies can fail to foresee biases that impede the ability of hospitals to implement an approach. In particular, the advancing sophistication of predictive HFR studies may exacerbate these biases.

Predictive HFR studies have begun using advanced forms of deep learning such as convolutional neural networks, deep-unified neural networks, and hybrid topic recurrent neural networks.³¹⁻³⁴ These tools represent some of the most powerful algorithms that data science has to offer but require high levels of technical expertise in the design of network architecture, choice of hyperparameters, and model validation. When married to massive amounts of patient data, advanced forms of machine learning may require computational resources that most hospitals lack. And, when requiring data that many hospitals may not readily have (e.g., complete assays of laboratory results and clinician notes), advanced machine learning methods can be self-limiting in regards to practical application.³²

Authors of HFR studies could increase their impact by aiming for real-time predictions and by avoiding the above biases.^{4,16,32} Going further, studies could increase their usefulness to hospitals by collaborating with clinical teams to clarify how an approach can be used in practice, if at all, and at what point in patient management. Likewise, predictive HFR studies could increase their usefulness among researchers and hospital data scientists by adopting higher standards of reproducibility. In particular, HFR studies rarely make their analytical source code and permissible data available via public repositories (e.g., via GitHub). By offering public repositories containing permissible data (e.g., simulated pseudo-data) and well-documented

source code built with open-source tools (e.g., Python, R), HFR studies can enable others to use, test, and improve upon existing methods.

Inherent uncertainty of 30-day all cause HFR — Predicting HF-related mortality, HFR beyond 30 days, or predicting which patients have HF often leads to greater success than predicting 30-day all-cause HFR.^{4,5,10,13,14,20,21,31,33,35} As opposed to using a strong diagnostic such as elevated cardiac troponin to predict whether a patient has HF or will expire within one year, predicting 30-day all-cause HFR often means using data with a weak signal to predict whether a patient will readmit for any unplanned cause within a relatively short period of time. The inherent uncertainties of this problem make it a grand predictive challenge and whether it outstrips the classification problems to which machine learning is often successfully applied is unknown. However, the most sophisticated deep learning algorithm will fail when making predictions about random outcomes (e.g., coin flips). Though prior studies have shown that 30-day all-cause HFR is not entirely random, the overall failure to breach C-statistics above 0.8 and precision above 0.5, despite greatly varied data and increasingly advanced methods, suggests a degree of stochasticity that is poorly understood and rarely discussed.

Authors have suggested that increasingly sophisticated machine learning applied to a greater volume of relevant data will improve 30-day all-cause HFR predictions.^{9,11,29} Given the march of advancing methods and increasingly large data sets, many have invested in this assumption. But, to our knowledge, that assumption has never been tested and it remains to be seen what data are most useful for predicting 30-day all-cause HFR and what models will push the field beyond incremental success. It seems clear, however, that hospitals are penalized on what seems to be the most difficult aspect of HFR to predict. Consequently, progress towards predictive power and the continuation of policies that penalize against readmissions need a clearer understanding of the inherent stochasticity of 30-day all-cause HFR.

Missed opportunities — The common goal of HFR studies is to predict which HF patients will readmit or, at least, which are at greatest risk. Considering the precision and accuracy needed and the urgent needs of hospitals, additional goals should be considered. For example, HF patients that do not readmit may be easier to predict than those that do. Among HFR studies that report positive predictive values (PPV; precision) and negative predictive values (NPV), only 19% –

44% of predictions for readmission were correct, while 80 – 96% of predictions for non-readmission were correct.^{21,24,28} However, no HFR study has explicitly asked whether or why predicting negative outcomes (non-readmissions) is easier than predicting positive ones (readmissions) or how the ability to predict non-readmissions can be used.

Statistical sampling theory and the inspection of diagnostic metrics (TPR, FPR, PPV, NPV) suggests that the class imbalance of HFR data should cause false negatives to have less of an impact on the probability of predicting a non-readmission (NPV) than false positives have on PPV (Fig 2b-c, Table 1). In short, confident predictions for non-readmissions may be naturally easier to achieve. Additionally, and also related to class imbalance, a given HFR data set is bound to be richer in information on negative outcomes (non-readmission) than positive ones, because non-readmissions are more common than readmissions. Taken together, predicting non-readmissions should be less sensitive to misclassification and offer more data on which confident predictions can be based.

Given that success towards predicting HFR has advanced little in the past decade with respect to the sophistication of the tools applied or the data used, a more strategic use of predictive analytics is needed. To this end, we suggest an uncertainty-based perspective. Specifically, if predictions for non-readmission are often precise (e.g., $NPV > 0.8$), then predictions with a high probability of non-readmission should aptly be treated as “high certainty for non-readmission”. Likewise, because predictions for readmission are typically incorrect (i.e., $PPV < 0.4$), predictions normally regarded as indicating “high risk” should more aptly be treated as predictions of “high uncertainty”.

Stratifying patients by uncertainty rather than presumably greater risk may allow hospitals to more effectively distribute limited resources (e.g., outpatient care, phone calls, virtual visits). Rather than overstretch resources by assuming that patients of presumably high risk are truly at high risk, hospitals could look deeper into their “high uncertainty” cohort for signs of greater certainty. Additionally, perhaps filtering out the “high certainty for non-readmission” cohort from HF discharge data will help correct the class imbalance in HF data sets and allow algorithms to focus their training on patients of greater uncertainty. Integrating precise predictions for non-target cases is bound to be more useful than ignoring them and, when “high risk” patients readmit at a frequency less than 50%, it is almost certainly more appropriate and useful to consider them as patients of high uncertainty.

Beyond HFR — The challenges of predicting 30-day all-cause readmissions and the penalties imposed on hospitals for not decreasing them, extend beyond patients with an HF discharge diagnosis. Hospitals face the same challenges for patients discharged with pneumonia, acute myocardial infarction, and other clinical conditions or procedures.¹ We suspect that predicting 30-day all-cause readmissions for most any discharge diagnosis is likely fraught with similar challenges. In each case, hospitals need easily integrated tools that make real-time predictions for strongly stochastic events from all-to-often weakly diagnostic data, and quickly enough to make informed decisions.

Beyond aiming for larger datasets and increasingly sophisticated models, we expect studies that eventually overcome these predictive challenges will be aided by: 1) a greater focus on precision (PPV, NPV), 2) a better choice of diagnostic metrics (e.g., PRCs, partial AUCs), 3) greater attention to practical application and reproducibility, 4) an exploration of additional goals (e.g., prediction of non-readmissions) and perhaps, 5) the adoption of an uncertainty-based perspective that clarifies the inherent stochasticity of 30-day all-cause readmissions and, likewise, that allows researchers and hospitals to best deal with it.

References

1. McIlvennan, C.K., Eapen, Z.J. and Allen, L.A., 2015. Hospital readmissions reduction program. *Circulation*, 131(20), pp.1796-1803.
2. Chamberlain, R.S., Sond, J., Mahendraraj, K., Lau, C.S. and Siracuse, B.L., 2018. Determining 30-day readmission risk for heart failure patients: The Readmission After Heart Failure scale. *International journal of general medicine*, 11, p.127.
3. Guo, A., Pasque, M., Loh, F., Mann, D.L. and Payne, P.R., 2020. Heart Failure Diagnosis, Readmission, and Mortality Prediction Using Machine Learning and Artificial Intelligence Models. *Current Epidemiology Reports*, pp.1-8.

4. Shin, S., Austin, P.C., Ross, H.J., Abdel-Qadir, H., Freitas, C., Tomlinson, G., Chicco, D., Mahendiran, M., Lawler, P.R., Billia, F. and Gramolini, A., 2020. Machine learning vs. conventional statistical models for predicting heart failure readmission and mortality. *ESC Heart Failure*.
5. Ross, J.S., Mulvey, G.K., Stauffer, B., Patlolla, V., Bernheim, S.M., Keenan, P.S. and Krumholz, H.M., 2008. Statistical models and patient predictors of readmission for heart failure: a systematic review. *Archives of internal medicine*, 168(13), pp.1371-1386.
6. Krumholz, H.M., Chen, Y.T., Wang, Y., Vaccarino, V., Radford, M.J. and Horwitz, R.I., 2000. Predictors of readmission among elderly survivors of admission with heart failure. *American heart journal*, 139(1), pp.72-77.
7. Frizzell, J.D., Liang, L., Schulte, P.J., Yancy, C.W., Heidenreich, P.A., Hernandez, A.F., Bhatt, D.L., Fonarow, G.C. and Laskey, W.K., 2017. Prediction of 30-day all-cause readmissions in patients hospitalized for heart failure: comparison of machine learning and other statistical approaches. *JAMA cardiology*, 2(2), pp.204-209.
8. Brown, J.R., Alonso, A., Mazimba, S., Warman, E.N. and Bilchick, K.C., 2020. Improved 30 day heart failure rehospitalization prediction through the addition of device-measured parameters. *ESC Heart Failure*, 7(6), pp.3762-3771.
9. Gottdiener, J.S. and Fohner, A.E., 2020. Risk prediction in heart failure: new methods, old problems.
10. Kansagara, D., Englander, H., Salanitro, A., Kagen, D., Theobald, C., Freeman, M. and Kripalani, S., 2011. Risk prediction models for hospital readmission: a systematic review. *Jama*, 306(15), pp.1688-1698.
11. Golas, S.B., Shibahara, T., Agboola, S., Otaki, H., Sato, J., Nakae, T., Hisamitsu, T., Kojima, G., Felsted, J., Kakarmath, S. and Kvedar, J., 2018. A machine learning model to predict the risk

of 30-day readmissions in patients with heart failure: a retrospective analysis of electronic medical records data. BMC medical informatics and decision making, 18(1), pp.1-17.

12. Mandrekar, J.N., 2010. Receiver operating characteristic curve in diagnostic test assessment. Journal of Thoracic Oncology, 5(9), pp.1315-1316.

13. Fleming, L.M., Gavin, M., Piatkowski, G., Chang, J.D. and Mukamal, K.J., 2014. Derivation and validation of a 30-day heart failure readmission model. The American journal of cardiology, 114(9), pp.1379-1382.

14. Yazdan-Ashoori, P., Lee, S.F., Ibrahim, Q. and Van Spall, H.G., 2016. Utility of the LACE index at the bedside in predicting 30-day readmission or death in patients hospitalized with heart failure. American heart journal, 179, pp.51-58.

15. Ahmad, F.S., French, B., Bowles, K.H., Sevilla-Cazes, J., Jaskowiak-Barr, A., Gallagher, T.R., Kangovi, S., Goldberg, L.R., Barg, F.K. and Kimmel, S.E., 2018. Incorporating patient-centered factors into heart failure readmission risk prediction: A mixed-methods study. American heart journal, 200, pp.75-82.

16. Amarasingham, R., Moore, B.J., Tabak, Y.P., Drazner, M.H., Clark, C.A., Zhang, S., Reed, W.G., Swanson, T.S., Ma, Y. and Halm, E.A., 2010. An automated model to identify heart failure patients at risk for 30-day readmission or death using electronic medical record data. Medical care, pp.981-988.

17. Awan, S.E., Bennamoun, M., Sohel, F., Sanfilippo, F.M. and Dwivedi, G., 2019. Machine learning-based prediction of heart failure readmission or death: implications of choosing the right model and the right metrics. ESC heart failure, 6(2), pp.428-435.

18. Awan, S.E., Bennamoun, M., Sohel, F., Sanfilippo, F.M., Chow, B.J. and Dwivedi, G., 2019. Feature selection and transformation by machine learning reduce variable numbers and improve prediction for heart failure readmission or death. PloS one, 14(6), p.e0218760.

19. Carrington, A.M., Fieguth, P.W., Qazi, H., Holzinger, A., Chen, H.H., Mayr, F. and Manuel, D.G., 2020. A new concordant partial AUC and partial c statistic for imbalanced data in the evaluation of machine learning algorithms. BMC medical informatics and decision making, 20(1), pp.1-12.
20. Sudharshan, S., Novak, E., Hock, K., Scott, M.G. and Geltman, E.M., 2017. Use of biomarkers to predict readmission for congestive heart failure. The American journal of cardiology, 119(3), pp.445-451.
21. Roshanghalb, A., Mazzali, C. and Lettieri, E., 2020. Composite outcomes of mortality and readmission in patients with heart failure: retrospective review of administrative datasets. Journal of Multidisciplinary Healthcare, 13, p.539.
22. Chen, C.Y., Lin, W.C. and Yang, H.Y., 2020. Diagnosis of ventilator-associated pneumonia using electronic nose sensor array signals: solutions to improve the application of machine learning in respiratory research. Respiratory research, 21(1), p.45.
23. Krumholz, H.M., Chaudhry, S.I., Spertus, J.A., Mattera, J.A., Hodshon, B. and Herrin, J., 2016. Do non-clinical factors improve prediction of readmission risk? results from the Tele-HF study. JACC: Heart Failure, 4(1), pp.12-20.
24. Leong, K.T.G., Wong, L.Y., Aung, K.C.Y., Macdonald, M., Cao, Y., Lee, S., Chow, W.L., Doddamani, S. and Richards, A.M., 2017. Risk stratification model for 30-day heart failure readmission in a multiethnic South East Asian community. The American journal of cardiology, 119(9), pp.1428-1432.
25. El Morr, C., Ginsburg, L., Nam, S. and Woollard, S., 2017. Assessing the performance of a modified LACE index (LACE-rt) to predict unplanned readmission after discharge in a community teaching hospital. Interactive Journal of Medical Research, 6(1), p.e7183.

26. McKinley, D., Moye-Dickerson, P., Davis, S. and Akil, A., 2019. Impact of a pharmacist-led intervention on 30-day readmission and assessment of factors predictive of readmission in African American men with heart failure. *American journal of men's health*, 13(1), p.1557988318814295.

27. Kleiner Shochat, M., Fudim, M., Shotan, A., Blondheim, D.S., Kazatsker, M., Dahan, I., Asif, A., Rozenman, Y., Kleiner, I., Weinstein, J.M. and Panjraht, G., 2018. Prediction of readmissions and mortality in patients with heart failure: lessons from the IMPEDANCE-HF extended trial. *ESC heart failure*, 5(5), pp.788-799.

28. Zai, A.H., Ronquillo, J.G., Nieves, R., Chueh, H.C., Kvedar, J.C. and Jethwani, K., 2013. Assessing hospital readmission risk factors in heart failure patients enrolled in a telemonitoring program. *International journal of telemedicine and applications*, 2013.

29. Zolfaghar, K., Meadem, N., Teredesai, A., Roy, S.B., Chin, S.C. and Muckian, B., 2013, October. Big data solutions for predicting risk-of-readmission for congestive heart failure patients. In *2013 IEEE International Conference on Big Data* (pp. 64-71). IEEE.

30. Davis, J. and Goadrich, M., 2006, June. The relationship between Precision-Recall and ROC curves. In *Proceedings of the 23rd international conference on Machine learning* (pp. 233-240).

31. Xiao, C., Ma, T., Dieng, A.B., Blei, D.M. and Wang, F., 2018. Readmission prediction via deep contextual embedding of clinical concepts. *PloS one*, 13(4), p.e0195024.

32. Mahajan, S.M., Heidenreich, P., Abbott, B., Newton, A. and Ward, D., 2018. Predictive models for identifying risk of readmission after index hospitalization for heart failure: a systematic review. *European Journal of Cardiovascular Nursing*, 17(8), pp.675-689.

33. Kwon, J.M., Kim, K.H., Jeon, K.H. and Park, J., 2019. Deep learning for predicting in-hospital mortality among heart disease patients based on echocardiography. *Echocardiography*, 36(2), pp.213-218.

34. Liu, X., Chen, Y., Bae, J., Li, H., Johnston, J. and Sanger, T., 2019, November. Predicting Heart Failure Readmission from Clinical Notes Using Deep Learning. In 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) (pp. 2642-2648). IEEE.
35. Evans, R.S., Benuzillo, J., Horne, B.D., Lloyd, J.F., Bradshaw, A., Budge, D., Rasmusson, K.D., Roberts, C., Buckway, J., Geer, N. and Garrett, T., 2016. Automated identification and predictive tools to help identify high-risk heart failure patients: pilot evaluation. Journal of the American Medical Informatics Association, 23(5), pp.872-878.

Table 1. Diagnostic measures related to receiver operating characteristic (ROC) curves and precision-recall curves (PRC). P = positives = readmissions. N = negatives = non-readmissions. TP = True positives, FP = False positives, TN = True negatives, FN = False negatives.

Measure	Equation	Description relative to HF readmissions (HFR)	Note
True positive rate (TPR), sensitivity, recall	$TPR = TP/P = TP/(TP+FN)$	Fraction of positives correctly classified	No established minimum acceptable TPR for HFR
False positive rate (FPR), False alarm rate	$FPR = FP/N = FP/(FP+TN)$	Fraction of non-readmissions incorrectly classified as positives	No established maximum acceptable FPR for HFR
True negative rate (TNR), specificity	$TNR = TN/N = TN/(TN + FP)$	Fraction of non-readmissions correctly classified as negatives	Rarely used in HFR studies
False negative rate (FNR)	$FNR = FN/P = FN/(FN+TP)$	Fraction of readmissions incorrectly classified as negatives	Rarely used in HFR studies
Precision; positive predictive value (PPV)	$PPV = TP/(TP+FP)$	Fraction of predictions for readmission that were correct.	Rarely used in HFR studies
Negative predictive value (NPV)	$NPV = TN/(TN+FN)$	Fraction of predictions for non-readmission that were correct.	Rarely used in HFR studies

Figure 1. Receiver operating characteristic (ROC) curves and precision-recall curves (PRC) resulting from analyzing a single simulated dataset of 10^5 binary outcomes (20% positives, 80% negatives) with three hypothetical models. These include a model programmed to make unbiased random guesses (black lines), a model programmed to produce results typical of many HFR models (blue lines), and a model programmed to represent an HFR model of exceptionally high diagnostic power (red lines). See Appendix 1 for modeling details. **A)** Examples of ROC curves and AUROC values. As expected, the random model approximates the 1:1 line (AUROC = 0.5). **B)** Same ROC curves as in A but showing partial AUROC ($pAUC_c$) values using a maximally acceptable false positive rate of 0.25. Solid lines represent the range of FPR below 0.25. Note how $pAUC_c = AUROC = 0.5$ for the random model, but $pAUC_c < AUROC$ for other models. **C)** PRCs and their AUPRC values. Solid lines represent FPR values below 0.25, as in B. This figure can be recreated using the supplemental Fig1.py python script.

Figure 2. Receiver operating characteristic (ROC) curves and precision-recall curves (PRC) resulting from the application of a single model (typical of results for HFR models) to two simulated datasets ($n = 10^5$ binary outcomes). The ratio of negative to positive outcomes was 1:1 for the balanced data set and 4:1 for the imbalanced data set. The latter ratio represents that ~20% of index discharges in HFR datasets result in 30-day all-cause readmission. See Appendix 1 for modeling details. **A)** ROC curves typical of HFR model results; these are identical for imbalanced and balanced data sets. **B)** PRCs reveal the influence of imbalanced data on precision (i.e., positive predictive value, PPV). The AUPRC for balanced data is more than twice that for imbalanced data. **C)** PRC curves focused on negative outcomes, i.e., negative predictive value (NPV) vs. true negative rate (TNR). AUPRC for imbalanced data is substantially greater than that for balanced data. This figure can be recreated using the supplemental Fig2.py python script.