



EDITORIAL COMMENT

The trials and tribulations of risk prediction studies in heart failure



Desafios e dificuldades em estudos de previsão de risco de insuficiência cardíaca

Daniel T. Mathew, Sanjiv J. Shah*

Division of Cardiology, Department of Medicine, and Bluhm Cardiovascular Institute, Northwestern University Feinberg School of Medicine, Chicago, IL, USA

Available online 29 August 2022

Heart failure (HF) is a leading worldwide cause of morbidity and mortality; therefore, it is of great clinical importance to try to predict which patients with HF are at increased risk for adverse outcomes (e.g., HF hospitalization, death) to appropriately triage high-risk patients for closer monitoring or advanced HF therapies.¹ There are several established risk calculators to estimate mortality and morbidity for patients with HF and reduced ejection fraction (HFrEF) such as the Seattle Heart Failure Model,² the Heart Failure Survival Score,³ and the Meta-Analysis Global Group in Chronic Heart Failure (MAGGIC) risk model.⁴ Published in 2014, the MAGGIC risk score was based on 39 372 HF patients from 30 studies and predicts mortality regardless of underlying ejection fraction (EF). Using variables readily encountered in outpatient clinical practice, the MAGGIC risk calculator is a practical tool for clinical evaluation.

Notably, however, neither B-type natriuretic peptide (BNP) nor N-terminal pro-BNP (NT-proBNP) was included in the MAGGIC risk score despite numerous studies showing their ability to predict a variety of adverse outcomes in HF across the EF spectrum.⁵ Furthermore, the MAGGIC risk

score was derived from studies that predated the advent of modern guideline-directed medical therapy for HFrEF, and did not include patients from many parts of the world, particularly lower- and middle-income countries.⁴ Finally, the original MAGGIC risk score was developed to predict all-cause mortality, and surprisingly few follow-up studies have examined the utility of the MAGGIC risk score in predicting HF hospitalizations. Thus, evaluation of the performance of the MAGGIC risk score for risk prediction in contemporary HFrEF populations in these countries, for both HF hospitalization and death, and with and without inclusion of NT-proBNP, remains an unmet need.

In this issue of the Portuguese Journal of Cardiology, Rohen et al. address these concerns by conducting a prospective, longitudinal observational study in 93 consecutive patients with chronic HF and EF <50% who were followed for an average duration of approximately 7-8 months.⁶ The baseline MAGGIC score was calculated and compared to baseline NT-proBNP values in the prediction of adverse events. The primary end point in the study was time-to-first event for either HF hospitalization (20 events) or cardiac death (three events). The study participants were relatively low risk (82% were New York Heart Association class I or II) with low MAGGIC risk scores (median 16, which corresponds to the lowest quartile in the original MAGGIC risk score

* Corresponding author.

E-mail address: sanjiv.shah@northwestern.edu (S.J. Shah).

study), even though 37% had an HF hospitalization in the 12 months prior to enrollment into the study. Both the MAGGIC risk score and NT-proBNP had relatively high sensitivity (87% and 82%, respectively) but both had low specificity (37% and 57%), leading to low area under the receiver operating curve of 0.59 and 0.67. Despite these test characteristics, both MAGGIC and NT-proBNP independently predicted the composite outcome, and the authors concluded that NT-proBNP is a better predictor than MAGGIC and that it is additive to MAGGIC based on their results.⁶

The strengths of this study include its prospective design, the inclusion of HF hospitalization (in addition to mortality) as one of the outcomes, the very high frequency of guideline-directed medical therapy for HFrEF (renin-angiotensin system inhibitor 98%, mineralocorticoid antagonist 90%, and beta-blocker 95%), and the Brazilian patient population (in whom similar studies are generally lacking).

However, several aspects of the study remain unanswered or could be improved. Most importantly, the issue remains of the generalizability of the authors' single center study to the overall Brazilian population. Despite a prior study demonstrating a very high prevalence of symptomatic HF in the region of Brazil where the study took place (9.3% prevalence in individuals >45 years of age in the DIGITALIS study),⁷ the present study only included 93 consecutive patients seen in the outpatient HF clinic at a university hospital over a one-year period. This number of consecutive HFrEF patients seems quite low for an outpatient HF clinic and may reflect a referral bias to an academic program. The authors state that patients with terminal illnesses other than HF, with a short life expectancy, were excluded from the study; however, the authors did not include a CONSORT diagram.⁸ It is therefore unclear how many patients were excluded for this reason. If the number is large, it could account for the low risk nature of the study cohort, as described above. It is also unclear what proportion of the patients had had mildly reduced HF or improved EF (which is likely to be high given the mean EF of 39% in the study population). Furthermore, it is not clear whether the low use of defibrillators and cardiac resynchronization therapy (only 3.2%) reflects the relatively healthy patient population, a relatively low number of patients with EF <30%, or low use of these devices in Brazilian patients, given the paucity of data on these devices in HFrEF in Brazil (e.g., in the BREATHE registry of HF patients in Brazil, use of these devices was not reported⁹). Finally, both the MAGGIC risk score and the study by Rohen et al. predated the advent of SGLT2 inhibitors as a key pillar of HFrEF therapy¹⁰; thus, the findings could be different in the current and future era where there will likely be rapid uptake of these drugs in HFrEF, especially as they become generic medications.

Besides generalizability, the study is limited by the relatively short duration of follow-up (mean follow-up of only 219 days) — a limitation noted by the authors. The major limitation of the study, however, is the comparison of the MAGGIC risk score and NT-proBNP as risk predictors. First, the authors state that "... NT-proBNP outperformed the MAGGIC score." While this statement may be true, the authors justify it by showing in their multivariable risk prediction model that the increase in risk of the primary outcome (cardiac death or HF hospitalization) was only 10% for every one-point increase in the MAGGIC risk score

vs. five-fold (i.e., 500% increased risk) in patients with a baseline NT-proBNP >1000 pg/ml.⁶ This is an unequal comparison between the two risk predictors. At the very least, the authors should have compared patients with MAGGIC risk score >12 (the optimal cut-point in their study for MAGGIC) vs. NT-proBNP >1000 pg/ml. However, a more equal comparison would have been a one standard deviation increase in each predictor, which would have enabled a more balanced comparison of effect sizes for the two predictors for the primary outcome. Second, the authors state in the abstract to their paper that "the addition of biomarkers improved the accuracy of the [MAGGIC] score."⁶ While they do show differences in Kaplan-Meier survival curves when combining MAGGIC score above and below 12 points, and NT-proBNP above and below 1000 pg/ml, a better way to examine the additive effects of the two risk prediction methods (MAGGIC risk score and NT-proBNP) would have been to determine whether the addition of one risk predictor to the other results in a statistically significant improvement in model performance. For example, the authors could have compared the Cox regression model for MAGGIC risk score vs. MAGGIC risk score plus NT-proBNP using a likelihood ratio test, comparison of Harrell's C indices, or comparing model goodness of fit measures.¹¹ The opposite also could have been done (comparison of NT-proBNP alone to NT-proBNP plus the MAGGIC risk score to see if addition of the MAGGIC score to NT-proBNP results in statistically significant improvement in model performance). Additional metrics such as the net reclassification index could have also been used to examine the benefit or lack thereof when adding one risk predictor over to another.¹² Thus, the question of whether these risk predictors are additive (vs. duplicative) in Brazilian HFrEF and HFmrEF patients remains unanswered.

Despite the aforementioned limitations, the study by Rohen et al. is important because it emphasizes the need for additional studies to provide improved risk prediction for HF patients in Brazil and elsewhere in the world. As mentioned by the authors, current HF risk scores consistently overestimate mortality (in the present study, for example, predicted mortality was 7% at one year compared to 3.2% observed mortality over a mean follow-up duration of 219 days). Overestimation of mortality may be due to several reasons, but most importantly may have to do with the advent and uptake of several risk-lowering therapies in HFrEF patients that were not available at the time of the studies used to create the HF risk scores. For example, in the original MAGGIC risk score derivation publication, only approximately 60-70% of patients were on renin-angiotensin system inhibitors and only 24-40% of patients were on beta-blockers (and neither angiotensin receptor-neprilysin inhibitors or SGLT2 inhibitors, which both reduce cardiovascular death in HFrEF, were available yet).

In addition, while risk scores may work well at the population level, they often do not perform well in the individual patient.¹³ Indeed, in the study by Rohen et al., while both the MAGGIC risk score and NT-proBNP had relatively high sensitivity, the specificity of the two risk predictors was low, as mentioned above. This means that in patients with HF, EF <50%, and a MAGGIC risk score <12 or NT-proBNP <1000 pg/ml, HF hospitalization or death are unlikely. However, in patients with values of these predictors higher than these

thresholds (which is common), predicting adverse outcomes is still difficult given the low specificity of both predictors. Further research is necessary to identify low-cost predictors of HF hospitalization and death in these patients so that they can be triaged appropriately for more intensive monitoring and/or advanced HF therapies. Further study of patients with divergent risk predictors and outcomes would also be helpful to improve risk prediction. In the study by Rohen et al., 25% of patients had an NT-proBNP <225 pg/ml. Did any of these patients have adverse outcomes? If so, examining these patients and their outcomes in detail could help improve risk prediction approaches. Alternatively, patients with high MAGGIC risk scores and/or NT-proBNP values who do not experience adverse outcomes could also be studied to determine whether there are protective factors (biological or socioeconomic [e.g., access to care, social support]) that could be leveraged to improve outcomes in other high-risk patients.

Ultimately, however, risk prediction of HF hospitalization and mortality in HF patients will always be challenging given the fuzzy nature of these outcomes. For HF hospitalization, the cause of HF exacerbation and decision to hospitalize is dependent on numerous reasons, many of which are not conventional or measurable risk factors. Medication and dietary non-compliance, both of which can lead to HF exacerbation, can be measured but can be variable over time in the individual patient. The decision to hospitalize a patient with HF is also quite variable and depends on the patient (e.g., willingness to be hospitalized, social support, access to outpatient intravenous diuretics) and the healthcare provider (e.g., HF specialists may have a higher threshold for hospitalization compared to general practitioners). Even mortality can be challenging for risk prediction. Despite the binary nature of the outcome, competing causes of death (e.g., deaths due to non-cardiac comorbidities, accidental deaths) may occur and may be unrelated to the HF diagnosis. Thus, no risk prediction algorithm will ever be perfect; clinical acumen, provision of medical and social support for HF patients, and individualization and optimization of medical therapy will ultimately be most important in the care of HF patients.

In conclusion, Rohen et al. have provided one of the first prospective studies of risk prediction in Brazilian patients with HF and EF <50% that has included the MAGGIC risk score and both HF hospitalization and death outcomes, which is an important advance in the care of Brazilian HF patients. Nevertheless, their study demonstrates the trials and tribulations of risk prediction studies in HF, which remain challenging.

Conflicts of interest

DTM has no relevant disclosures. SJS has received research grants from the National Institutes of Health (U54 HL160273, R01 HL107577, R01 HL127028, R01 HL140731,

R01 HL149423), Actelion, AstraZeneca, Corvia, Novartis, and Pfizer; and has received consulting fees from Abbott, Actelion, AstraZeneca, Amgen, Aria CV, Axon Therapies, Bayer, Boehringer-Ingelheim, Boston Scientific, Bristol-Myers Squibb, Cardiora, Coridea, CVRx, Cycleron, Cytokinetics, Edwards Lifesciences, Eidos, Eisai, Imara, Impulse Dynamics, Intellia, Ionis, Ironwood, Lilly, Merck, MyoKardia, Novartis, Novo Nordisk, Pfizer, Prothena, Regeneron, Rivus, Sanofi, Shifamed, Tenax, Tenaya, and United Therapeutics.

References

1. Zhao HL, Cui W. Prognostic risk scores for patients with heart failure. *Br J Hosp Med (Lond)*. 2022;83:1–16.
2. Levy WC, Mozaffarian D, Linker DT, et al. The Seattle Heart Failure Model: prediction of survival in heart failure. *Circulation*. 2006;113:1424–33.
3. Aaronson KD, Schwartz JS, Chen TM, et al. Development and prospective validation of a clinical index to predict survival in ambulatory patients referred for cardiac transplant evaluation. *Circulation*. 1997;95:2660–7.
4. Pocock SJ, Ariti CA, McMurray JJ, et al. Predicting survival in heart failure: a risk score based on 39 372 patients from 30 studies. *Eur Heart J*. 2013;34:1404–13.
5. Don-Wauchope AC, Santaguida PL, Oremus M, et al. Incremental predictive value of natriuretic peptides for prognosis in the chronic stable heart failure population: a systematic review. *Heart Fail Rev*. 2014;19:521–40.
6. Rohan FM, de Avila DX, Lemos CMC, et al. The MAGGIC risk score in the prediction of death or hospitalization in patients with heart failure: comparison with natriuretic peptides. *Rev Port Cardiol*. 2022 (in press).
7. Jorge AL, Rosa ML, Martins WA, et al. The prevalence of stages of heart failure in primary care: a population-based study. *J Card Fail*. 2016;22:153–7.
8. von Elm E, Altman DG, Egger M, et al. The Strengthening of Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med*. 2007;147:573–7.
9. Albuquerque DC, Neto JD, Bacal F, et al. I Brazilian Registry of Heart Failure – clinical aspects, care quality and hospitalization outcomes. *Arq Bras Cardiol*. 2015;104:433–42.
10. Heidenreich PA, Bozkurt B, Aguilar D, et al. 2022 AHA/ACC/HFSA Guideline for the Management of Heart Failure: a report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. *Circulation*. 2022;145:e895–1032.
11. Klein JP, Moeschberger ML. *Survival analysis: techniques for censored and truncated data*, vol 1230. New York: Springer; 2003.
12. Leening MJ, Vedder MM, Wittteman JC, et al. Net reclassification improvement: computation, interpretation, and controversies: a literature review and clinician's guide. *Ann Intern Med*. 2014;160:122–31.
13. Allen LA, Matlock DD, Shetterly SM, et al. Use of risk models to predict death in the next year among individual ambulatory patients with heart failure. *JAMA Cardiol*. 2017;2:435–41.