Editorial

Risk Prediction in Heart Failure: Untranslatable or Lost in Translation?

THOMAS H. MARWICK, MBBS, PhD, MPH, AND QUAN HUYNH, MB, PhD

Melbourne, Australia

The outcome of heart failure (HF) is heterogeneous, depending on severity, etiology, precipitants, comorbidities, social determinants, and treatment variables. Both for social and medical reasons, having some ability to prognosticate is of value.

The Seattle Heart Failure Model (SHFM), based upon clinical, laboratory, and treatment variables, is among the most widely-used HF risk scores. SHFM was described in 1125 patients with heart failure, and validated by the original investigators in 5 populations of nearly 10,000 patients with HF, with a C statistic of 0.73.[1] However, its use has predominantly been in Europe and North America and the evidence of its utility in other populations with HF has not been well defined. This is an important consideration in Asia, where there is a growing burden of HF, fueled by an aging population and a substantial increase in the prevalence of HF risk factors.[2] Shiraishi et al[3] have validated the SHFM in Japan, initially in a single-center population of ~500 patients. In this issue of The Journal, the same investigators further describe the validation and calibration of the SHFM in 2470 Japanese patients in the West Tokyo Heart Failure (WET-HF) and National cerebral and cardiovascular center acute Decompensated heart Failure (NaDEF) registries.[4] In patients with HF with reduced ejection fraction (HFrEF), they show SHFM to have good discrimination (C statistic, 0.75) but overestimate 1-year survival (ie, miscalibration), which could be addressed by simple intercept recalibration. The HFpEF group was more difficult, with modest discrimination and underestimation of 1-year survival, which did not improve following simple intercept recalibration. Differences between the populations, particularly in medication use, might explain this variation.

Implicit in this study is the notion that application of a score to a new population is dependent on similarity of that population to the derivation group. If the model were based on social variables, it would be easy to understand how differences might occur between different cultures. At first thought, we might consider this unlikely to be a consideration with the SHFM, given that it is dominated by objective medical observations, which we might consider to be uniform among people with HF. However, patients hospitalized for HF in Asia tend to present with a more severe clinical profile, and are often younger than their counterparts in the West.[5] Moreover, the etiology and comorbidities of HF are not the same as those in the West; comorbidities have been reported to cluster in 5 patterns, with different consequences on perceived wellbeing and outcomes.[6] We might therefore expect that an Asian HF score might well be different from one in the West, and the results of this study suggest this is indeed the case.

This heterogeneity in the performance of scores is not limited to geographical variation, but may also include treatment subgroups, especially when the results of treatment are variable. SHFM seems to discriminate in patients undergoing cardiac resynchronization therapy (CRT).[7] However, survival was overestimated in a transplant population, particularly among patients with implanted devices.[8]

Heterogeneity in performance might also be a temporal issue. Relatively stable results for the SHFM were reported in the decade between 2004 and 2014, over which time the outcomes of HF were stable.[9] It is possible that as the sacubitril/valsartan combination becomes more widely used outcomes will improve, leading to a temporal drift in calibration. Conversely, the increasing proportion of patients with HFpEF, and the lack of specific treatment, might need a modification of the score to be developed.

As aptly described by Aaronson and Cowger,[10] the results of HF scores are probabilistic, not deterministic. The scores work best when they are applied to patient groups, for example identifying individuals at similar levels of risk to participate in trials, to reflect the average risk of different populations for the purposes of benchmarking between hospitals, or to identify a particular risk subgroup for therapy of limited availability (such as a

From the Imaging Research Group, Baker Heart and Diabetes Institute, Melbourne, Australia.

Manuscript received May 29, 2019; revised manuscript accepted May 29, 2019.

Reprint requests: Tom Marwick, MBBS, PhD, MPH, Baker Heart and Diabetes Institute, 75 Commercial Road, Melbourne, Victoria 3004, Australia. E-mail: tom.marwick@baker.edu.au

1071-9164/$ - see front matter

https://doi.org/10.1016/j.cardfail.2019.05.011

Table 1. Discrimination of HF Risk Scores

Scores (References)	n (patients)	Location	Content	AUC
SHFM[1]	1125	USA	Clinical variables	0.73
	9942			0.69 to 0.81
Japan validation[4]	2470	Japan	Clinical variables	0.75
CRT validation[7]	342	Italy	Clinical variables	0.70
Transplant validation[8]	445	USA	Clinical variables	0.78
European registry[12] – MAGGIC	6,161	Europe	Clinical variables	0.74
GISSI-HF				0.74
CHARM				0.73
SHFM				0.71
US registry[13] – SHFM	10 980	USA	Clinical variables	0.66
MAGGIC				0.69
Heart failure survival score[17]	268 (derivation)	USA	Clinical variables and peak oxygen consumption	0.79
	199 (validation)			0.76
MECKI[18]	2716	Italy	Clinical, echocardiographic and metabolic exercise testing	0.80

CHARM, Candesartan in Heart Failure-Assessment of Reduction in Mortality and Morbidity; CRT, cardiac resynchronization therapy; GISSI-HF, Gruppo Italiano per lo Studio della Sopravvivenza nell'Infarto miocardico; MAGGIC, Meta-analysis Global Group in Chronic Heart Failure; MECKI, metabolic exercise cardiac kidney indexes; SHFM, Seattle Heart Failure Model.

disease management program).[11] When applied to individuals, the scores might provide some grounds for shared decision-making, but the reliability of these predictions seems modest at most. In a recent report from 6161 patients in the European Society of Cardiology HF registry (91.8% 1-year survival), the Meta-analysis Global Group in Chronic Heart Failure (MAGGIC), Gruppo Italiano per lo Studio della Sopravvivenza nell'Infarto miocardico (GISSI-HF), and Candesartan in Heart Failure-Assessment of Reduction in Mortality and Morbidity (CHARM) scores showed similar overall accuracy (area under the curve [AUC] 0.74, 0.74 and 0.73, respectively), superior to the SHFM (AUC = 0.71; P= .018), although the observed-to-predicted survival ratios showed overestimation of mortality by all scores (CHARM: 1.10, GISSI-HF: 1.08, MAGGIC: 1.03), except the SHFM (0.98).[12] In an important study of 10,980 patients with HF, 161 of whom died during follow-up,[13] the C statistic of SHFM was 0.66, and a >50% probability of death in 1 year was identified in only 8 of those who died, and 5 among those who lived for 1 year. The C statistic for the MAGGIC score was 0.69, but although 52 patients who died had a predicted risk of >50%, 63 who lived for >1 year were also in this category. The prediction of death over the medium-term is more feasible than short-term prediction, or the prediction of nonfatal events,[14] especially readmission, probably reflecting the importance of social and medical influences on individual readmission decisions. The fact that these scores are moderately rather than extremely predictive, perhaps explains why <1% of patients in the ESC registry received a prognostic estimate from their enrolling physician.[12]

Existing HF risk scores focus on mortality and do not account for nonfatal events and the changes in quality of life (QoL) associated with these events. This is important, because patients' input into shared decision-making is often influenced by nonfatal events. Findings from a study using a time tradeoff utility have shown that patients with HF, usually those with more severe HF, are willing to trade a substantial proportion of their remaining time for a significantly better QoL.[15] As there may even be changes in patient preferences as clinical status evolves,[16] the notion of incorporating QoL into risk models requires significant future research.

The availability of advanced treatments for heart failure is highly variable around the world, and the desire to optimize selection of patients is understandable. The limited discrimination of most HF scores might be supplemented by physiologic markers (including VO_2, biomarkers, and imaging). The Heart Failure Survival Score (HFSS), which included both clinical variables and peak oxygen consumption, provided an AUC of 0.79±0.03 in the derivation and 0.76±0.04 in the validation sample.[17] The Metabolic Exercise Cardiac Kidney Indexes (MECKI) score integrated clinical, echocardiographic and metabolic exercise testing findings in 2716 patients with HFrEF at 13 Italian centers, and reported AUC values of 0.804 (0.754−0.852) at 1 year.[18] A subsequent comparative study of 259 patients with HFrEF confirmed the good discrimination of the MECKI (AUC 0.8−0.9), with calibration at 1 year, similar or better than the SHFM, MAGGIC, and HFSS.[19] The challenge is that there is a need to understand which of a large number of patients with HF should be further tested with VO_2 and other tests, possibly repeatedly.

Although the discrimination of different scores shows variability (Table 1), we have to question whether a 70%−80% success rate in discriminating death or major events is sufficient for individual patient discussions. In the current era of precision medicine, it should be possible to provide more accurate individual risk prediction than is provided by the standard HF risk scores. When such a well-calibrated algorithm is created, it will doubtless aid individual treatment decisions. Until then we should accept that prediction is a very inexact science—even with recalibration based on race, region, and local practice—and we should beware the pitfalls of underestimation and overestimation of risk.

References

1. Levy WC, Mozaffarian D, Linker DT, et al. The Seattle Heart Failure Model: prediction of survival in heart failure. Circulation 2006;113:1424–33.
2. Mentz RJ, Roessig L, Greenberg BH, et al. Heart failure clinical trials in East and Southeast Asia: understanding the importance and defining the next steps. JACC Heart Fail 2016;4:419–27.
3. Shiraishi Y, Sawano M, Kohno T, et al. Validation of the Seattle Heart Failure Model in Japanese heart failure patients. Int J Cardiol 2016;203:87–9.
4. Shiraishi Y, Kohsaka S, Nagai T, Goda A, Mizuno A, Nagatomo Y, Validation and recalibration of Seattle Heart Failure Model in Japanese acute heart failure patients. J Card Fail in press.
5. Atherton JJ, Hayward CS, Wan Ahmad WA, et al. Patient characteristics from a regional multicenter database of acute decompensated heart failure in Asia Pacific (ADHERE International-Asia Pacific). J Card Fail 2012;18:82–8.
6. Tromp J, Tay WT, Ouwerkerk W, et al. Multimorbidity in patients with heart failure from 11 Asian regions: A prospective cohort study using the ASIAN-HF registry. PLoS Med 2018;15:e1002541.
7. Perrotta L, Ricciardi G, Pieragnoli P, et al. Application of the Seattle Heart Failure Model in patients on cardiac resynchronization therapy. Pacing Clin Electrophysiol 2012;35:88–94.
8. Kalogeropoulos AP, Georgiopoulou VV, Giamouzis G, et al. Utility of the Seattle Heart Failure Model in patients with advanced heart failure. J Am Coll Cardiol 2009;53:334–42.
9. Sacks CA, Jarcho JA, Curfman GD. Paradigm shifts in heart-failure therapy: a timeline. N Engl J Med 2014;371:989–91.
10. Aaronson KD, Cowger J. Heart failure prognostic models: why bother? Circ Heart Fail 2012;5:6–9.
11. Huynh QL, Whitmore K, Negishi K, Marwick TH, Investigators E. Influence of risk on reduction of readmission and death by disease management programs in heart failure. J Card Fail 2019;25:330–9.
12. Canepa M, Tavazzi L, Maggioni AP, Investigators EHLTR. Reply: The Barcelona Bio-HF Calculator: a contemporary web-based heart failure risk score. JACC Heart Fail 2018;6:810–1.
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.
14. Howlett JG. Should we perform a heart failure risk score? Circ Heart Fail 2013;6:4–5.
15. MacIver J, Rao V, Delgado DH, et al. Choices: a study of preferences for end-of-life treatments in patients with advanced heart failure. J Heart Lung Transplant 2008;27:1002–7.
16. Stevenson LW, Hellkamp AS, Leier CV, et al. Changing preferences for survival after hospitalization with advanced heart failure. J Am Coll Cardiol 2008;52:1702–8.
17. Aaronson KD, Schwartz JS, Chen TM, Wong KL, Goin JE, Mancini DM. Development and prospective validation of a clinical index to predict survival in ambulatory patients referred for cardiac transplant evaluation. Circulation 1997;95:2660–7.
18. Agostoni P, Corra U, Cattadori G, et al. Metabolic exercise test data combined with cardiac and kidney indexes, the MECKI score: a multiparametric approach to heart failure prognosis. Int J Cardiol 2013;167:2710–8.
19. Freitas P, Aguiar C, Ferreira A, Tralhao A, Ventosa A, Mendes M. Comparative analysis of four scores to stratify patients with heart failure and reduced ejection fraction. Am J Cardiol 2017;120:443–9.