


Comparative performance of risk prediction indices for mortality or readmission following heart failure hospitalization

Tauben Averbuch¹ , Ali Zafari², Shofiquil Islam³, Shun Fu Lee³, Rajiv Sankaranarayanan^{4,5}, Stephen J Greene⁶, Mamas A Mamas⁷, Ambarish Pandey⁸ and Harriette GC Van Spall^{3,9,10*}

¹Department of Cardiology, University of Calgary, Calgary, AB, Canada; ²Tehran Heart Center, Cardiovascular Diseases Research Institute, Tehran University of Medical Sciences, Tehran, Iran; ³Population Health Research Institute, Hamilton, ON, Canada; ⁴Department of Cardiology, Aintree University Hospital, Liverpool University Hospitals NHS Foundation Trust, Liverpool, UK; ⁵Liverpool Centre for Cardiovascular Sciences, University of Liverpool, Liverpool, UK; ⁶Duke Clinical Research Institute, Durham, North Carolina, USA; ⁷Keele University, Keele, UK; ⁸Department of Medicine, University of Texas Southwestern Medical Center, Dallas, Texas, USA; ⁹Department of Medicine, Faculty of Health Sciences, McMaster University, Hamilton, ON, Canada; and ¹⁰Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, ON, Canada

Abstract

Aims Risk prediction indices used in worsening heart failure (HF) vary in complexity, performance, and the type of datasets in which they were validated. We compared the performance of seven risk prediction indices in a contemporary cohort of patients hospitalized for HF.

Methods and results We assessed the performance of the Length of stay and number of Emergency department visits in the prior 6 months (LE), Length of stay, number of Emergency department visits in the prior 6 months, and admission N-Terminal prohormone of brain natriuretic peptide (NT-proBNP) (LENT), Length of stay, Acuity, Charlson co-morbidity index, and number of Emergency department visits in the prior 6 months (LACE), Get With The Guidelines Heart Failure (GWTG), Readmission Risk Score (RRS), Enhanced Feedback for Effective Cardiac Treatment model (EFFECT), and Acute Decompensated Heart Failure National Registry (ADHERE) risk indices among consecutive patients hospitalized for HF and discharged alive from January 2017 to December 2019 in a network of hospitals in England. The primary composite outcome was 30-day all-cause mortality or readmission. We assessed model discrimination and overall accuracy using the C-statistic (higher values, better) and Brier score (lower values, better), respectively. Among 1206 patients in the cohort, 45.0% were female, mean (SD) age was 76.6 (11.7) years, and mean (SD) left ventricular ejection fraction was 43.0% (11.6). At 30 days, 236 (19.6%) patients were readmitted and 28 (2.3%) patients died, with 264 (21.9%) patients experiencing either readmission or death. The LENT index offered the combination of greatest risk discrimination and accuracy for the primary composite outcome (C-statistic: 0.97; 95% CI 0.96, 0.98; 0.29; Brier score: 0.05). The LE (C-statistic: 0.95; 95% CI 0.93, 0.96; Brier score: 0.06) and LACE (C-statistic: 0.90; 95% CI 0.88, 0.92; Brier score 0.09) indices had high discrimination and accuracy. Discrimination and accuracy were modest with the RRS (C-statistic: 0.65; 95% CI 0.61, 0.69; Brier score: 0.16) and EFFECT (C-statistic: 0.64; 95% CI 0.60, 0.67; Brier score: 0.16) score; and poor with the GWTG-HF (C-statistic: 0.62; 95% CI 0.58, 0.66; Brier score: 0.17) and ADHERE (C-statistic: 0.54; 95% CI 0.50, 0.57; Brier score: 0.17) scores.

Conclusions In a study that compared the performance of seven risk prediction indices in a contemporary cohort of patients hospitalized for HF, the simple LENT index offered the greatest combination of discrimination and accuracy for the primary composite outcome of 30-day all-cause mortality or readmission. This three-variable index -using length of hospital stay, preceding emergency department visits and admission NT-proBNP level- is a practical and reliable way to assess prognosis following hospitalization for HF.

Keywords Acute heart failure ; Risk prediction; 30-day mortality; 30-day readmission

Received: 19 September 2024; Accepted: 3 October 2024

*Correspondence to: Harriette Gillian Christine Van Spall, Population Health Research Institute, 237 Barton Street East, Hamilton, ON L8L 2X2, Canada.

Email: harriette.vanspall@phri.ca

Background

In the first 30 days following hospitalization for heart failure (HF), patients face a 20%–25% risk of all-cause readmission and 4%–10% risk of death.^{1–3} Risk prediction tools can identify hospitalized patients at high risk of death to help guide post-discharge follow-up, transplant referrals, or palliative care.³ However, a majority have not been adequately validated, and their performance, particularly calibration, is often not reported.^{4–8} Validated risk prediction tools in HF are typically complex and require risk calculators as well as multiple variables for computation, rendering them impractical at the point of care.^{4–6,8} A majority of established risk prediction models in HF have been derived and validated using administrative data, which are limited by the accuracy of coding; for example, up to 20% of patients in administrative datasets can have the diagnosis of HF misclassified.⁹

In response to the complexity and limitations of existing HF risk prediction models, we derived and externally validated the LE index, comprising only two variables: length (L) of hospital stay and the number of emergency (E) department visits (E) in the preceding 6 months.^{10,11} To further improve the performance of the LE index, we derived and internally validated the LENT index, which adds points based on admission or discharge N-terminal prohormone of brain natriuretic peptide (NT-proBNP) level to the LE index.¹² The LENT index had better risk classification for 30-day composite all-cause mortality or readmission than did the LE index.¹²

To date, the performance of validated HF risk prediction indices has not been compared in a single cohort. Comparisons have therefore been indirect, based on performance in cohorts or administrative datasets that vary in demographics, comorbidities, background therapies, and quality of care. In this comparative effectiveness cohort study, we aimed to compare the discrimination, calibration, and overall performance of the LENT index with six other risk indices that had previously been validated for the prediction of clinical outcomes among patients hospitalized for HF and discharged alive; these indices included LE, Length of stay, Acuity, Charlson co-morbidity index, and number of Emergency department visits in the prior 6 months (LACE), Acute Decompensated Heart Failure National Registry (ADHERE), Enhanced Feedback for Effective Cardiac Treatment (EFFECT), Get With The Guidelines Heart Failure (GWTG), and Readmission Risk Score (RRS) scores^{4–6,8,10,12} (*Table S1*).

Methods

This study was approved by the Hamilton Integrated Research Ethics Board, Health and Care Research Wales, and the West Midlands—The Black Country Research Ethics Committee.

Study design and population

We undertook a comparative effectiveness study in a cohort of consecutive patients hospitalized for HF and discharged alive from the Liverpool University Hospitals NHS Foundation Trust, comprising three hospitals that serve an ethnically diverse population in England. Patients were included if they had a primary admission diagnosis of decompensated HF from 1 January 2017 to 31 December 2019. Our exclusion criteria were limited and pragmatic. We excluded patients who did not have a primary diagnosis of acute HF and patients who died during the index hospitalization. Study personnel collected demographic and clinical data from the electronic medical record and obtained prior ED visits from the hospital database. Heart failure was identified with the ICD-10 codes I11.0, I25.5, I42.0, I42.9, I50.0, I50.1, and I50.9.

Data collection

We extracted baseline demographics and relevant variables to calculate the risk scores of interest (*Tables S1*). Clinical outcomes were obtained from the electronic medical record of the Liverpool University Hospital Foundation Trust and the integrated primary care record using Egton Medical Information Systems (EMIS) and e-Xchange. Risk indices were selected based on prior validation in patients hospitalized for HF, although the outcomes varied. The ADHERE score was originally derived and validated for inpatient mortality. The GWTG-HF and RRS were originally derived and validated for 30-day mortality and 30-day readmission, respectively. Risk indices validated only in ambulatory patients were not included. Outcomes were unblinded and obtained through chart review.

Outcomes

The primary composite outcome was 30-day all-cause mortality or readmission. Secondary outcomes included the components of the primary outcome.

Statistical analysis

Given observed 30-day event rates in the Canadian cohort in which the LENT index was derived (20.3% readmission and 21.1% mortality or readmission), we estimated that we would need at least 1000 patients to observe approximately 200 events (mortality or readmission) and allow for meaningful analysis.^{3,12}

We represented continuous variables as means with standard deviations and medians with interquartile ranges and categorical variables with percentages and numbers. We

computed the ADHERE, EFFECT, GWTG-HF, LACE, LE, LENT, and RRS indices using their published scoring systems (Table S2). All laboratory values or vital signs used in each risk model were recorded at admission as in the original derivation studies. We compared the distribution of variables between patients who were readmitted or died with those who experienced neither outcome using the Rao–Scott chi-square test for categorical variables, and the Wilcoxon rank sums test for continuous variables.

We fitted multivariable logistic regression models for each risk score, with the risk score as the predictor and the composite of all-cause mortality or readmission, all-cause mortality, or all-cause readmission as the outcome. We assessed the odds of each clinical outcome per 1-point increase for all models with odds ratios and 95% confidence intervals (CI). We compared the risk discrimination of each model using the C-statistic, reported with 95% CIs. The C-statistic, a measure of a model's ability to discriminate between patients at low and high risk of a given outcome, ranges from 0 to 1, with higher scores indicating better discrimination. We assessed model calibration - consistency between the estimated risk and observed risk - with the Hosmer–Lemeshow (H-L) *P*-value, with the null hypothesis stating that the model is well-calibrated. A H-L *P*-value below 0.05 indicates inadequate calibration. We assessed model accuracy—a combination of discrimination and calibration—with the Brier score, which has a scale from 0 to 1, where lower scores indicate superior model performance. No adjustments were made for multiple comparisons. We performed our analyses with SAS version 9.4 (SAS Institute Inc, Cary, NC, USA).

Results

Baseline characteristics

We screened 1333 patients for inclusion and excluded 127 due to death in hospital. Baseline characteristics of the 1206 patients enrolled in this study are depicted in Table 1. The mean (SD) age was 76.6 (11.7) years, 45.0% of patients were female, and the mean (SD) left ventricular ejection fraction was 43.0% (11.6). The burden of co-morbidities was representative of usual practice and similar to the LENT derivation cohort.^{12,13} Among the patients in our cohort, 33.9% had diabetes, 59.4% had hypertension, 57.0% had atrial fibrillation, and 22.5% had previously undergone percutaneous coronary intervention (PCI) or coronary artery bypass grafting (CABG). Data were available to compute risk scores for all patients included in the study, with no missing values.

Of the patients in the study, 29.1% had an intermediate or high-risk ADHERE score. The mean (SD) EFFECT score was 78.0 (21.5), GWTG-HF score was 39.2 (6.6), LACE index was

10.7 (3.3), LE index was 4.6 (2.0), LENT index was 5.6 (2.6), and RRS was 21.4 (4.3) (Figure 1 A–G). At discharge, 42.9% of patients were prescribed an angiotensin-converting enzyme (ACE)-inhibitor, angiotensin receptor blocker (ARB), or angiotensin receptor neprilysin inhibitor (ARNI), 75.5% of patients were prescribed a beta-blocker, 19.2% were prescribed a mineralocorticoid receptor antagonist (MRA), and 94% of patients were prescribed a loop diuretic (Table 1).

Clinical outcomes

Of the 1206 patients in the cohort, 264 (21.9%) were either readmitted or died, 236 (19.6%) were readmitted, and there were 28 (2.3%) all-cause deaths within 30 days of discharge. Patients who were either readmitted or died within 30 days of discharge exhibited higher mean scores of the seven risk prediction indices in comparison to those who did not (Table 1). Event rates were similar to the LENT derivation cohort.¹² Outcome data were available for all patients in our cohort.

Performance of risk prediction tools

Composite 30-day all-cause mortality or readmission

Figure 2 and Table 2 illustrate the model performance of each risk index. The combination of discrimination, calibration and accuracy for the primary outcome were highest for the LENT index (C-statistic: 0.97; 95% CI 0.96, 0.98; Hosmer–Lemeshow *P*-value: 0.29; Brier score: 0.05) (Table 2, Central Illustration, Figure 2A). Discrimination and accuracy were high but calibration poor for the LE index (C-statistic: 0.95; 95% CI 0.93, 0.96; Hosmer–Lemeshow *P*-value: <0.01; Brier score: 0.06). Of the remaining risk scores, discrimination and accuracy were modest and calibration good with the RRS (C-statistic: 0.65; 95% CI 0.61, 0.69; Hosmer–Lemeshow *P*-value: 0.39; Brier score: 0.16). Discrimination and accuracy were poorest with the ADHERE score (C-statistic: 0.54; 95% CI 0.50, 0.57; Brier score: 0.17). The Hosmer–Lemeshow *P*-value could not be computed for the ADHERE score as it is categorical with three levels (low, intermediate, high).

30-day all-cause readmission

Figure 2 and Table 2 illustrate the model performance of each risk index. Discrimination and overall accuracy were highest, but calibration limited for the LENT (C-statistic: 0.96; 95% CI 0.94, 0.97; Hosmer–Lemeshow *P*-value: <0.01; Brier score: 0.06) index. The combination of discrimination, calibration, and accuracy were high for the LACE index (C-statistic: 0.89; 95% CI 0.87, 0.91; Hosmer–Lemeshow *P*-value: 0.10; Brier score: 0.09) (Table 2, Figure 2B). Of the remaining risk scores, discrimination was highest although modest, calibration good, and accuracy modest for the RRS (C-statistic: 0.63; 95% CI 0.59, 0.67; Hosmer–Lemeshow *P*-value: 0.58; Brier

Table 1 Baseline characteristics of 1206 patients discharged alive following hospitalization for HF

| | Overall cohort (N = 1206) | Readmitted or dead at 30 days (N = 264) | Readmitted at 30 days (N = 236) | Neither readmitted nor dead at 30 days (N = 942) |
|---|------------------------------|--|------------------------------------|---|
| Demographics | | | | |
| Age (years), mean (SD) | 76.6 (11.7) | 78.2 (10.6) | 77.8 (10.7) | 76.2 (11.9) |
| Sex | | | | |
| Male, n (%) | 543 (45.0) | 122 (46.2) | 113 (47.9) | 421 (44.7) |
| Female, n (%) | 663 (55.0) | 142 (53.8) | 123 (52.1) | 521 (55.3) |
| Vital signs | | | | |
| Systolic blood pressure at admission (mmHg), mean (SD) | 133.1 (23.7) | 131.1 (24.3) | 132.7 (24.2) | 133.6 (23.6) |
| Diastolic blood pressure at admission (mmHg), mean (SD) | 73.0 (13.9) | 72.7 (13.8) | 73.1 (13.7) | 73.1 (13.9) |
| Heart rate (beats per minute) at admission, mean (SD) | 83.9 (20.8) | 85.0 (20.0) | 84.8 (20.2) | 83.6 (21.0) |
| Respiratory rate (breaths per minute) at admission, mean (SD) | 21.8 (3.7) | 22.7 (4.0) | 22.9 (4.0) | 21.6 (3.6) |
| Co-morbidities | | | | |
| Left ventricular ejection fraction, n (%) | 43.0 (11.6) | 40.4 (12.0) | 40.3 (11.9) | 43.7 (11.5) |
| Hypertension, n (%) | 716 (59.4) | 160 (60.6) | 143 (60.6) | 556 (59.0) |
| Atrial fibrillation, n (%) | 688 (57.0) | 151 (57.2) | 133 (56.4) | 537 (57.0) |
| Prior heart failure, n (%) | 903 (74.9) | 233 (88.3) | 208 (88.1) | 670 (71.1) |
| Valvular disease, n (%) | 221 (18.3) | 71 (26.9) | 65 (27.5) | 150 (15.9) |
| Obstructive sleep apnoea, n (%) | 75 (6.2) | 11 (4.2) | 11 (4.7) | 64 (6.8) |
| Aortic stenosis, n (%) | 86 (7.1) | 28 (10.6) | 24 (10.2) | 58 (6.2) |
| Diabetes, n (%) | 409 (33.9) | 100 (37.9) | 90 (38.1) | 309 (32.8) |
| Chronic kidney disease, n (%) | 656 (54.4) | 170 (64.4) | 151 (64.0) | 486 (51.6) |
| Moderate to severe kidney disease, n (%) | 276 (22.9) | 96 (36.4) | 82 (34.7) | 180 (19.1) |
| Prior PCI or CABG, n (%) | 271 (22.5) | 72 (27.3) | 65 (27.5) | 199 (21.1) |
| Coronary artery disease, n (%) | 489 (40.5) | 123 (46.6) | 109 (46.2) | 366 (38.9) |
| Prior ICD or CRT, n (%) | 96 (8.0) | 26 (9.8) | 20 (8.5) | 70 (7.4) |
| Chronic pulmonary disease, n (%) | 375 (31.1) | 102 (38.6) | 90 (38.1) | 273 (29.0) |
| Peripheral vascular disease, n (%) | 80 (6.6) | 21 (8.0) | 18 (7.6) | 59 (6.3) |
| Previous TIA, n (%) | 87 (7.2) | 20 (7.6) | 18 (7.6) | 67 (7.1) |
| Previous stroke, n (%) | 126 (10.4) | 23 (8.7) | 19 (8.1) | 103 (10.9) |
| Dementia, n (%) | 84 (7.0) | 20 (7.6) | 15 (6.4) | 64 (6.8) |
| Liver disease, n (%) | 41 (3.4) | 14 (5.3) | 12 (5.1) | 27 (2.9) |
| Moderate to severe liver disease, n (%) | 22 (1.8) | 8 (3.0) | 6 (2.5) | 14 (1.5) |
| Connective tissue disease, n (%) | 48 (4.0) | 10 (3.8) | 10 (4.2) | 38 (4.0) |
| Cancer (any) n (%) | 164 (13.6) | 36 (13.6) | 32 (13.6) | 128 (13.6) |
| Metastatic solid tumour, n (%) | 26 (2.2) | 9 (3.4) | 8 (3.4) | 17 (1.8) |
| Charlson co-morbidity index, mean (SD) | 6.7 (2.8) | 7.8 (3.0) | 7.7 (2.9) | 6.4 (2.7) |
| Medications at discharge | | | | |
| ACE inhibitor, n (%) | 405 (33.6) | 77 (29.2) | 71 (30.1) | 328 (34.8) |
| ARB, n (%) | 81 (6.7) | 9 (3.4) | 9 (3.8) | 72 (7.6) |
| ARNI, n (%) | 31 (2.6) | 3 (1.1) | 3 (1.3) | 28 (3.0) |
| Beta-blocker, n (%) | 911 (75.5) | 204 (77.3) | 184 (78.0) | 707 (75.0) |
| Loop diuretic, n (%) | 1141 (94.6) | 241 (91.3) | 215 (91.1) | 900 (95.5) |
| MRA, n (%) | 232 (19.2) | 49 (18.6) | 46 (19.5) | 183 (19.4) |
| SGLT2 inhibitor, n (%) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Resource utilization | | | | |
| High acuity admission (via ED), n (%) | 624 (51.7) | 200 (75.8) | 180 (76.3) | 424 (45.0) |
| Number of ED visits in preceding 6 months, mean (SD) | 0.7 (1.1) | 2.1 (1.2) | 2.1 (1.2) | 0.3 (0.7) |
| Length of stay (day), mean (SD) | 7.3 (8.4) | 14.2 (11.9) | 14.1 (11.9) | 5.4 (5.9) |
| Investigations^a | | | | |
| Sodium (mmol), mean (SD) | 139.0 (4.4) | 138.2 (5.2) | 138.1 (5.2) | 139.2 (4.2) |
| Creatinine (μmol/L), mean (SD) | 125.9 (59.4) | 150.0 (80.9) | 148.2 (80.4) | 119.2 (49.8) |
| Blood urea nitrogen (mmol/L), mean (SD) | 5.2 (3.5) | 6.5 (4.4) | 6.2 (3.6) | 4.9 (3.1) |
| Haematocrit, mean (SD) | 0.36 (0.1) | 0.35 (0.1) | 0.35 (0.1) | 0.36 (0.1) |
| Log admission NT-proBNP (pg/ml), mean (SD) | 8.3 (0.9) | 9.2 (0.9) | 9.2 (0.9) | 8.1 (0.8) |
| Log discharge NT-proBNP (pg/ml), mean (SD) | 7.6 (0.9) | 8.3 (0.9) | 8.3 (0.9) | 7.4 (0.8) |

(Continues)

Table 1 (continued)

| | Overall cohort (N = 1206) | Readmitted or dead at 30 days (N = 264) | Readmitted at 30 days (N = 236) | Neither readmitted nor dead at 30 days (N = 942) |
|--|------------------------------|--|------------------------------------|---|
| Clinical characteristics | | | | |
| In-hospital cardiac arrest, n (%) | 13 (1.1) | 6 (2.3) | 4 (1.7) | 7 (0.7) |
| Estimated risk | | | | |
| ADHERE score, intermediate or high risk, n (%) | 351 (29.1) | 91 (34.5) | 74 (31.4) | 260 (27.6) |
| EFFECT score, mean (SD) | 78.0 (21.5) | 86.0 (20.7) | 84.1 (19.5) | 75.7 (21.2) |
| GWTG-HF score, mean (SD) | 39.2 (6.6) | 41.1 (7.0) | 40.5 (6.8) | 38.6 (6.4) |
| LACE index, mean (SD) | 10.7 (3.3) | 14.5 (2.5) | 14.5 (2.5) | 9.7 (2.6) |
| LE index, mean (SD) | 4.6 (2.0) | 7.2 (1.4) | 7.2 (1.3) | 3.9 (1.4) |
| LENT index, mean (SD) | 5.6 (2.6) | 9.2 (1.7) | 9.3 (1.7) | 4.6 (1.7) |
| RRS, mean (SD) | 21.4 (4.3) | 23.4 (4.9) | 23.2 (4.7) | 20.9 (4.0) |

ACE, Angiotensin converting enzyme; ADHERE, Acute Decompensated Heart Failure National Registry; ARB, Angiotensin receptor blocker; ARNI, angiotensin receptor neprilysin-inhibitor; CABG, coronary artery bypass grafting; CRT, cardiac resynchronization therapy; ED, emergency department; EFFECT, Enhanced Feedback for Effective Cardiac Treatment; GWTG-HF, Get With The Guidelines Heart Failure; ICD, implantable cardioverter-defibrillator; LACE, Length of stay, Acuity, Charlson co-morbidity index, and number of Emergency department visits in the prior 6 months; LE, Length of stay and number of Emergency department visits in the prior 6 months; LENT, Length of stay, number of Emergency department visits, and NT-proBNP; MRA, Mineralocorticoid receptor antagonist; NT-proBNP, N-terminal prohormone of brain natriuretic peptide; PCI, percutaneous coronary intervention; RRS, Readmission Risk Score; SD, standard deviation; SGLT2, Sodium-glucose Cotransporter-2; TIA, transient ischemic attack.

^aMeasured at admission.

score: 0.15), while discrimination and accuracy were poorest for the ADHERE score (C-statistic: 0.52; 95% CI 0.48, 0.55; Brier score: 0.16).

30-day all-cause mortality

Figure 2 and Table 2 illustrate the model performance of each risk index. Discrimination and overall accuracy were highest but calibration limited for the LENT (C-statistic: 0.89; 95% CI 0.87, 0.91; Hosmer–Lemeshow *P*-value: <0.01; Brier score 0.05) index. The combination of discrimination, calibration, and accuracy were high for the LACE index (C-statistic: 0.84; 95% CI 0.80, 0.87; Hosmer–Lemeshow *P*-value: 0.22; Brier score: 0.05) (Table 2, Figure 2C). Of the remaining risk scores, discrimination and accuracy were highest although modest with the GWTG-HF (C-statistic: 0.73; 95% CI 0.67, 0.79; Hosmer–Lemeshow *P*-value: 0.83; Brier score: 0.05) index with good calibration. Discrimination was poorest with the ADHERE score (C-statistic: 0.61; 95% CI 0.55, 0.67; Brier score: 0.05).

Discussion

In this cohort study that compared the performance of seven risk prediction indices, the LENT index—requiring only length of hospital stay, number of ED visits in preceding 6 months, and admission NT-proBNP for computation—offered the best combination of risk discrimination, calibration, and overall accuracy for the primary composite outcomes of 30-day all-cause mortality or readmission. The two-variable LE index predicted this outcome with similarly high discrimination but with poor calibration, although the overall accuracy was still high. The ease of computation of the LENT index at the point of care over established indices makes it practical for use in clinical settings and can help guide referral to special-

ist, advanced heart failure, or supportive care services. In resource-limited settings in which NT-proBNP cannot be used, the LE index is a reasonable option, given its simplicity, high discrimination, and overall performance.

Among the risk prediction models tested in our study cohort, the LENT index demonstrated the greatest overall performance across all outcomes, particularly for 30-day composite all-cause mortality or readmission, followed by the LE index and LACE index, respectively. Other established risk prediction models exhibited poorer discrimination and overall accuracy for the composite endpoint. The ADHERE, EFFECT, and GWTG-HF scores were derived and validated to predict mortality, and in the case of ADHERE and GWTG-HF, inpatient rather than 30-day mortality; these indices were not derived for readmission, and it is not surprising that their performance was modest for a 30-day outcome that included readmission.^{4–6} Furthermore, few patients in this cohort had high-risk ADHERE scores—the poor performance of the ADHERE score is likely also related to the limited distribution of risk. While the LE and LENT indices had a left-skewed distribution, which may limit their performance, a broader spread of values was observed than with the ADHERE score. The LENT and LE indices demonstrated superior overall accuracy even for the outcome of 30-day mortality. Readmission is more challenging to predict than mortality, as up to 65% of readmissions following HF hospitalization are not due to HF.¹⁴ Additionally, hospitalization decisions may rely on non-clinical factors such as hospital bed availability, available outpatient resources, patient or caregiver preferences and social circumstances. By comparison, the RRS, LE, LACE, and LENT indices were derived and validated for 30-day outcomes that included readmission.^{8,10–12}

The greater performance of the LENT, LE, and LACE indices may be explained by the inclusion of length of stay, which ac-

Figure 1 Distribution of risk as estimated by seven risk prediction indices in a cohort of patients discharged alive following hospitalization for HF. While most models demonstrated a broad distribution of risk, the majority of patients were deemed low risk by the ADHERE score and very few patients had a high-risk ADHERE score. ADHERE, Acute Decompensated Heart Failure National Registry; EFFECT, Enhanced Feedback for Effective Cardiac Treatment; GWTG-HF, Get With The Guidelines Heart Failure; LACE, Length of stay, Acuity, Charlson co-morbidity index, and number of Emergency department visits in the prior 6 months; LE, Length of stay and number of Emergency department visits in the prior 6 months; LENT, Length of stay, number of Emergency department visits, and NT-proBNP; RRS, Readmission Risk Score.

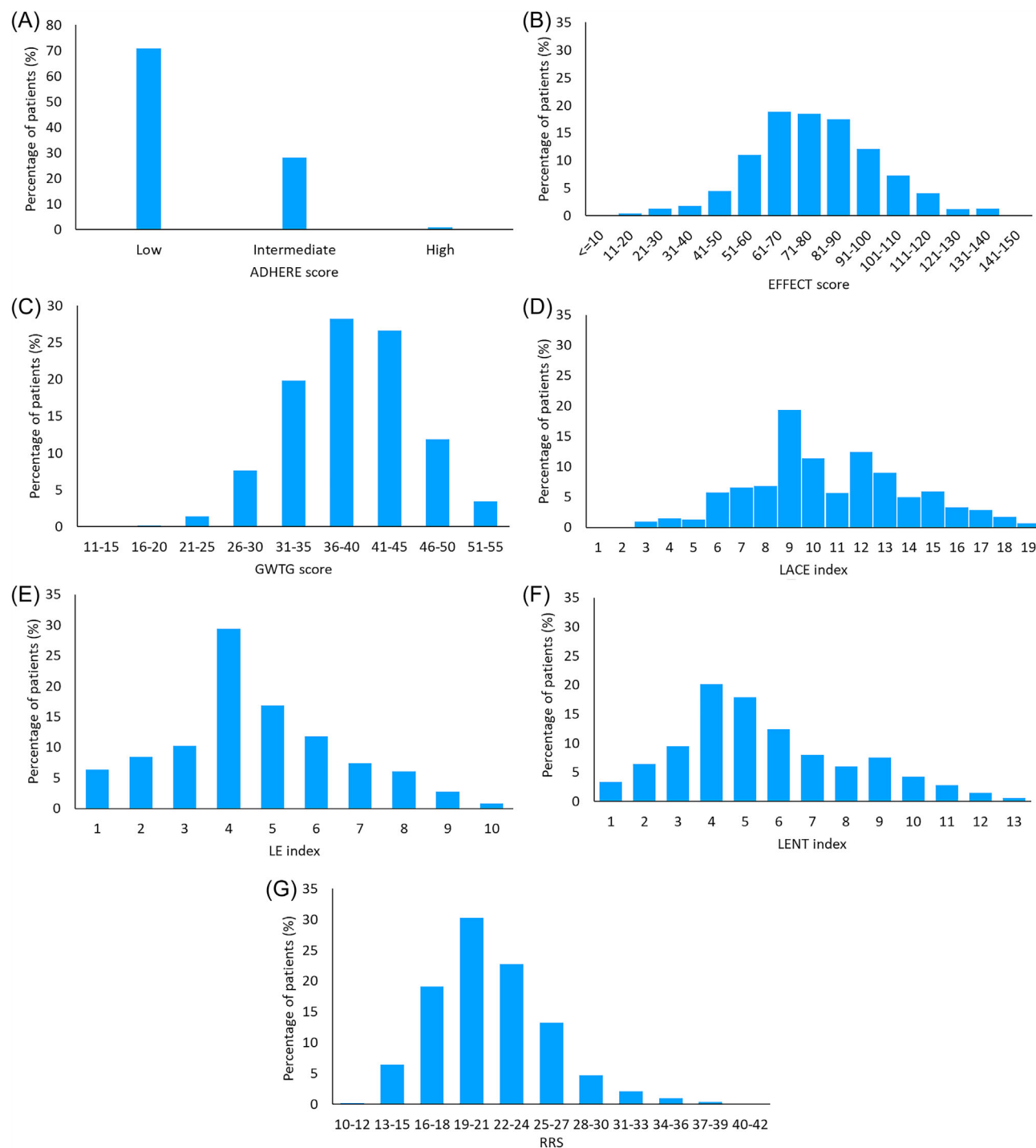


Figure 2 Comparative effectiveness of risk prediction models for 30-day clinical events. (A) Composite all-cause mortality or readmission; (B) All-cause readmission; (C) All-cause mortality. ADHERE, Acute Decompensated Heart Failure National Registry; EFFECT, Enhanced Feedback for Effective Cardiac Treatment; GWTG-HF, Get With The Guidelines Heart Failure; LACE, Length of stay, Acuity, Charlson co-morbidity index, and number of Emergency department visits in the prior 6 months; LE, Length of stay and number of Emergency department visits in the prior 6 months; LENT, Length of stay, number of Emergency department visits in the prior 6 months, and NT-proBNP; RRS, Readmission Risk Score. The C statistic is a measure of model discrimination, ranging from 0 to 1, with higher scores indicating greater discrimination. The Brier score is a measure of model accuracy ranging from 0 to 1, with lower scores indicating greater accuracy. The Hosmer-Lemeshow test evaluates model calibration, with the null hypothesis that the model is well-calibrated.

*High-risk vs low-risk ADHERE score;

†Intermediate-risk vs low-risk ADHERE score.

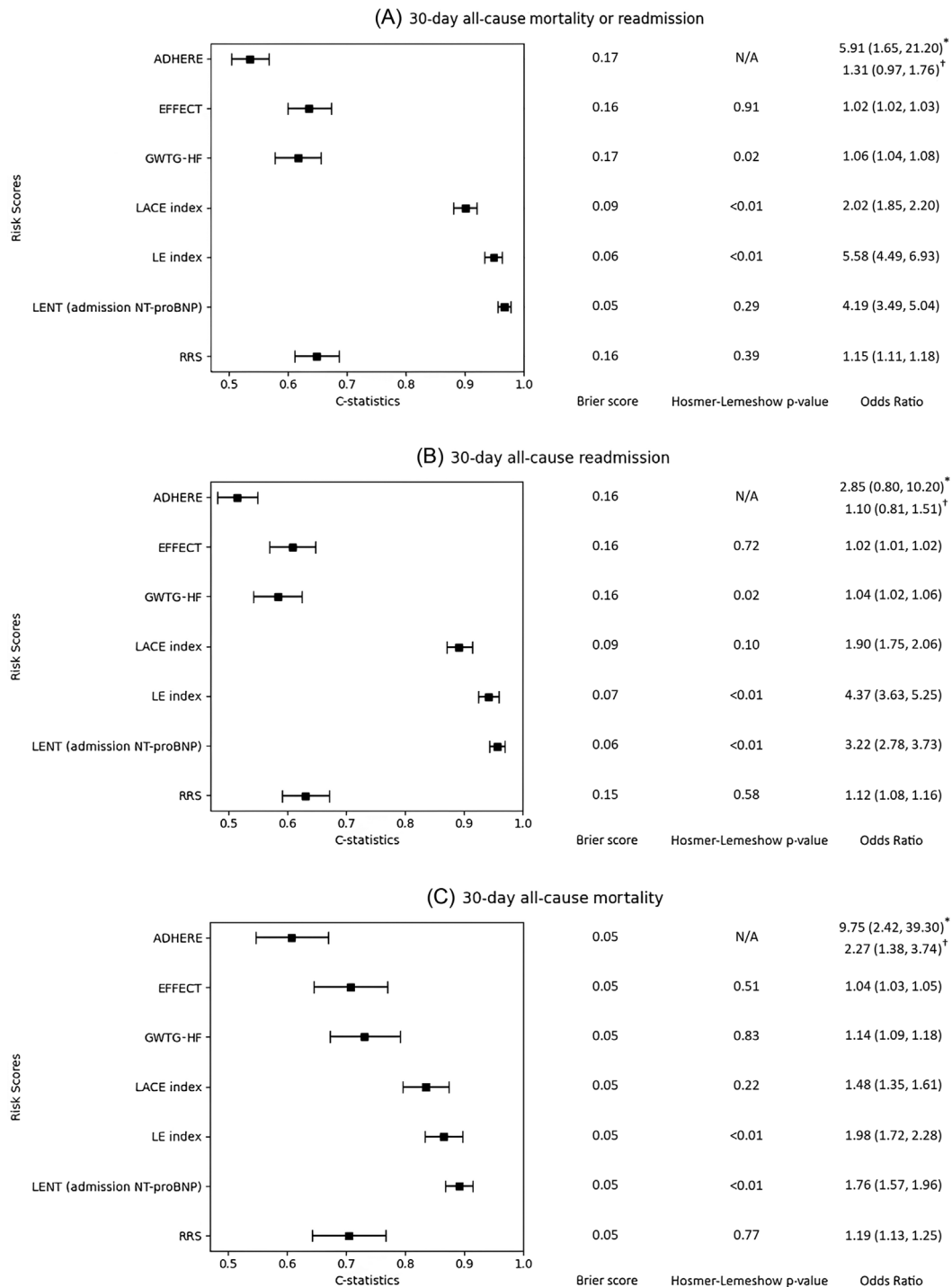


Table 2 The performance of HF risk indices for predicting 30-day outcomes following hospitalization for HF

| Risk score | | OR (95% CI) per 1 point increase | C-statistic ^a (95% CI) | Hosmer–Lemeshow <i>P</i> -value ^b | Brier score ^c |
|---|---------------------------|----------------------------------|-----------------------------------|--|--------------------------|
| 30-day all-cause mortality or readmission | | | | | |
| ADHERE | High vs. low risk | 5.91 (1.65, 21.20) | 0.54 (0.50, 0.57) | N/A | 0.17 |
| | Intermediate vs. low risk | 1.31 (0.97, 1.76) | | | |
| EFFECT | | 1.02 (1.02, 1.03) | 0.64 (0.60, 0.67) | 0.91 | 0.16 |
| GWTG-HF | | 1.06 (1.04, 1.08) | 0.62 (0.58, 0.66) | 0.02 | 0.17 |
| LACE index | | 2.02 (1.85, 2.20) | 0.90 (0.88, 0.92) | <0.01 | 0.09 |
| LE index | | 5.58 (4.49, 6.93) | 0.95 (0.93, 0.96) | <0.01 | 0.06 |
| LENT ^d | | 4.19 (3.49, 5.04) | 0.97 (0.96, 0.98) | 0.29 | 0.05 |
| RRS | | 1.15 (1.11, 1.18) | 0.65 (0.61, 0.69) | 0.39 | 0.16 |
| 30-day all-cause readmission | | | | | |
| ADHERE | High vs. low risk | 2.85 (0.80, 10.20) | 0.52 (0.48, 0.55) | N/A | 0.16 |
| | Intermediate vs. low risk | 1.10 (0.81, 1.51) | | | |
| EFFECT | | 1.02 (1.01, 1.02) | 0.61 (0.57, 0.65) | 0.72 | 0.16 |
| GWTG-HF | | 1.04 (1.02, 1.06) | 0.58 (0.54, 0.63) | 0.02 | 0.16 |
| LACE index | | 1.90 (1.75, 2.06) | 0.89 (0.87, 0.91) | 0.10 | 0.09 |
| LE index | | 4.37 (3.63, 5.25) | 0.94 (0.93, 0.96) | <0.01 | 0.07 |
| LENT ^d | | 3.22 (2.78, 3.73) | 0.96 (0.94, 0.97) | <0.01 | 0.06 |
| RRS | | 1.12 (1.08, 1.16) | 0.63 (0.59, 0.67) | 0.58 | 0.15 |
| 30-day all-cause mortality | | | | | |
| ADHERE | High vs. low risk | 9.75 (2.42, 39.30) | 0.61 (0.55, 0.67) | N/A | 0.05 |
| | Intermediate vs. low risk | 2.27 (1.38, 3.74) | | | |
| EFFECT | | 1.04 (1.03, 1.05) | 0.71 (0.64, 0.77) | 0.51 | 0.05 |
| GWTG-HF | | 1.14 (1.09, 1.18) | 0.73 (0.67, 0.79) | 0.83 | 0.05 |
| LACE index | | 1.48 (1.35, 1.61) | 0.84 (0.80, 0.87) | 0.22 | 0.05 |
| LE index | | 1.98 (1.72, 2.28) | 0.87 (0.83, 0.90) | <0.01 | 0.05 |
| LENT ^d | | 1.76 (1.57, 1.96) | 0.89 (0.87, 0.91) | <0.01 | 0.05 |
| RRS | | 1.19 (1.13, 1.25) | 0.70 (0.64, 0.77) | 0.77 | 0.05 |

ADHERE, Acute Decompensated Heart Failure National Registry; EFFECT, Enhanced Feedback for Effective Cardiac Treatment; GWTG-HF, Get With The Guidelines Heart Failure; LACE, Length of stay, Acuity, Charlson co-morbidity index, and number of Emergency department visits in the prior 6 months; LE, Length of stay and number of Emergency department visits in the prior 6 months; LENT, Length of stay, number of Emergency department visits in the prior 6 months, and NT-proBNP; OR, Odds ratio; RRS, Readmission Risk Score.

^aThe C-statistic measures the ability of a model to discriminate between patients who are higher and lower risk. Scores range from 0 to 1, with a higher score indicating better discrimination.

^bThe Hosmer–Lemeshow test compares predicted probabilities with actual outcomes to assess model calibration or how well the model fits the data. The null hypothesis is that the model is well calibrated. A *P*-value below 0.05 suggests poor calibration.

^cThe Brier score quantifies the accuracy of probabilistic predictions, with a scale from 0 (*perfect accuracy*) to 1 (*entirely inaccurate*), where lower scores indicate better model performance.

^dLENT index was computed using admission NT-proBNP level.

counts for cardiac and non-cardiac complications during hospitalization; the other risk indices all used variables present at baseline or upon admission and do not account for changes in clinical status during hospitalization. Furthermore, preceding ED visits, captured in LENT, LE, and LACE, are a reliable marker of declining health or increasing healthcare needs. The improved performance of the LENT index over all other models may also be due to the inclusion of NT-proBNP, which offers important prognostic information in HF and has been shown to increase net risk reclassification relative to the LE index for 30-day mortality or readmission.^{12,15} In the GUIDE-IT trial, NT-proBNP was significantly associated with cardiovascular death or HF hospitalization, with NT-proBNP levels beginning to rise, on average, 200 days before a clinical event.¹⁶ Furthermore, an elevated NT-proBNP at discharge may suggest incomplete decongestion, which is associated with adverse outcomes.¹⁷

The LE and LENT indices offer both improved performance and greater ease of use compared to traditional risk indices. The LE and LENT indices use variables that are easily accessible from the patient record, making them convenient for use

at the point of care without compromising the model's performance. While the ADHERE score, with two variables, offers the same degree of simplicity, its performance in predicting 30-day outcomes was poor. By comparison, the GWTG-HF score contains seven variables, while the RRS includes 20. For clinicians at the point of care faced with several available risk prediction tools, an ideal tool should be available, accurate, and actionable—tools that are perceived as time-consuming without added clinical value are unlikely to be incorporated into everyday practice.¹⁸

The LENT, LE, and LACE indices performed better in this study than in their original cohorts.^{10–12,19} The difference in mean LE, LENT, and LACE indices in patients who did versus did not experience 30-day composite all-cause mortality or readmission—discrimination—was far greater in the present cohort than in the derivation or validation cohorts.^{10–12} The greater separation in risk scores between patients with and without events may reflect differences between this cohort and the previous derivation or validation cohorts. There was a greater difference in length of stay between patients with and without the outcome in the present cohort (14.2

vs. 5.4 days, respectively) than in the LACE HF validation cohort (6 vs. 6 days, respectively).^{11,19} Similarly, the difference in number of ED visits between patients with and without the outcome was larger in this cohort (2.1 vs. 0.3) than in the original LENT cohort (3.2 vs. 2.2).¹² Taken together, this suggests that the patients in the present cohort had a wider range of risk profiles; the groups of patients with and without events were more distinct in this cohort than in the original cohorts. The greater spread of risk—when measured by length of stay and ED visits—between patients with and without the clinical outcomes of interest led to improved discrimination for the LACE, LE, and LENT indices. The other risk indices rely more on patient comorbidities or clinical characteristics at presentation, which did not differ as greatly between patients with and without the outcomes of interest, which may explain why performance for the traditional indices was similar to the derivation cohorts.

Strengths and limitations

Our study compares the performance of several traditional risk indices that have been well-validated in the heart failure population. We included a large, contemporary cohort of patients in a diverse hospital network, with a co-morbidity profile that is representative of usual practice. The clinical outcomes chosen were of interest to both clinicians and health systems, given the Center for Medicare Services' emphasis on predicting and preventing 30-day unplanned readmissions.⁸ We limited our analysis to risk indices that were derived and validated using traditional statistical modelling rather than machine learning so that we could generate confidence intervals around estimates and propose indices that could be implemented widely.²⁰ Nonetheless, limitations should be noted. The LE, LENT, and LACE indices may not perform as well in cohorts with more homogeneous risk, that is, with less separation in length of stay and prior ED visits between patients who are readmitted and those who are not. It is possible that the models were overfitted as their performance varied in prior cohorts. While calibration was inconsistent among the LE, LENT, and LACE indices between outcomes, the Brier score, a measure that weighs predictions against their outcomes, was lower for the LE, LENT, and LACE indices in this validation cohort than in the original derivation cohorts, suggesting reasonable model accuracy in spite of the Hosmer–Lemeshow test. Lastly, in our analysis of 30-day readmission, we did not account for the competing risk of death, which may limit model discrimination.

Conclusions

The 3-variable LENT index, which can be computed at the bedside based on length of hospital stay, number of ED visits in preceding 6 months, and admission NT-proBNP level, of-

fers the best combination of discrimination, calibration, and accuracy for predicting 30-day composite all-cause mortality or readmission following hospitalization for HF. The simple LE and LENT indices offer high risk discrimination with greater ease of use for 30-day composite all-cause mortality or readmission and its component outcomes than the established LACE, GWTG-HF, ADHERE, EFFECT, and RRS risk prediction tools that were originally derived and validated for mortality. Even for mortality, the LE and LENT indices have higher overall accuracy relative to the other risk indices.

Conflict of interest

SJG has received research support from the Duke University Department of Medicine Chair's Research Award, American Heart Association (#929502), Amgen, AstraZeneca, Boehringer Ingelheim, Bristol Myers Squibb, Cytokinetics, Merck, Novartis, Pfizer, and Sanofi; has served on advisory boards or as consultant for Amgen, AstraZeneca, Bayer, Boehringer Ingelheim, Bristol Myers Squibb, Corteria Pharmaceuticals, CSL Vifor, Cytokinetics, Eli Lilly, Lexicon, Merck, Roche Diagnostics, Sanofi, scPharmaceuticals, Tricog Health, and Urovant Pharmaceuticals; and has received speaker fees from Bayer, Boehringer Ingelheim, Cytokinetics, Lexicon, and Roche Diagnostics. HGCV receives funding from Canadian Institutes of Health Research, has served on clinical trial executive committees sponsored by Medtronic, Novo Nordisk, and Bayer. She has received program funding for education from Boehringer Ingelheim and Novartis. RS has received research support from the British Heart Foundation, National Institute for Health Research, Biotronik, Astra Zeneca and has received speaker fees or sponsorship for medical conference attendance from Medtronic, Astra Zeneca, Novartis, Boehringer Ingelheim, Pfizer, Roche Diagnostics, Vifor Pharma and Pharmacosmos. The authors have no relevant conflicts of interest to disclose. AP has received research support from the National Institute of Health, American Heart Association, Applied Therapeutics, SC Pharmaceuticals Roche, Ultromics, and Gilead Sciences; has received honoraria outside of the present study as an advisor/consultant for Tricog Health Inc, Lilly USA, Rivus, Cytokinetics, Roche Diagnostics, Axon therapies, Medtronic, Edward Lifesciences, Science37, Novo Nordisk, Bayer, Merck, Sarfez Pharmaceuticals, Emmi Solutions, Semler Scientific, Ultromics, Merck, Encarda, Kieele Health, Acorai; and has received non-financial support from Pfizer and Merck. Dr. Pandey is also a consultant for Palomarin with stock options.

Funding

This study was funded by peer-reviewed grants held by Dr Van Spall from the Canadian Institutes of Health Research

(Grant number: 135917) and Ontario Ministry of Health and Long-Term Care (Grant number: 6686).

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1. Risk prediction indices previously validated across cohorts of patients hospitalized for HF and included in the current comparative effectiveness study.

Table S2. Calculation of risk indices.

Data S1. Supplementary Information.

References

1. Van Spall HGC, Rahman T, Mytton O, Ramasundarahettige C, Ibrahim Q, Kabali C, *et al.* Comparative effectiveness of transitional care services in patients discharged from the hospital with heart failure: a systematic review and network meta-analysis: comparative effectiveness of transitional care services in patients hospitalized with heart failure. *Eur J Heart Fail* 2017;**19**: 1427-1443. doi:10.1002/ejhf.765
2. Bottle A, Goudie R, Bell D, Aylin P, Cowie MR. Use of hospital services by age and comorbidity after an index heart failure admission in England: an observational study. *BMJ Open* 2016;**6**: e010669. doi:10.1136/bmjopen-2015-010669
3. Van Spall HGC, Lee SF, Xie F, Oz UE, Perez R, Mitoff PR, *et al.* Effect of patient-centered transitional care services on clinical outcomes in patients hospitalized for heart failure: the PACT-HF randomized clinical trial. *JAMA* 2019;**321**:753-761. doi:10.1001/jama.2019.0710
4. Fonarow GC. Risk stratification for in-hospital mortality in acutely decompensated heart failure classification and regression tree analysis. *JAMA* 2005;**293**:572-580. doi:10.1001/jama.293.5.572
5. Lee DS, Austin PC, Rouleau JL, Liu PP, Naimark D, Tu JV. Predicting mortality among patients hospitalized for heart failure: derivation and validation of a clinical model. *JAMA* 2003;**290**:2581. doi:10.1001/jama.290.19.2581
6. Peterson PN, Rumsfeld JS, Liang L, Albert NM, Hernandez AF, Peterson ED, *et al.* A validated risk score for in-hospital mortality in patients with heart failure from the American Heart Association Get With the Guidelines program. *Circ Cardiovasc Qual Outcomes* 2010;**3**:25-32. doi:10.1161/CIRCOUTCOMES.109.854877
7. Di Tanna GL, Wirtz H, Burrows KL, Globe G. Evaluating risk prediction models for adults with heart failure: a systematic literature review. *PLoS ONE* 2020;**15**:e0224135. doi:10.1371/journal.pone.0224135
8. Keenan PS, Normand SLT, Lin Z, Drye EE, Bhat KR, Ross JS, *et al.* An administrative claims measure suitable for profiling hospital performance on the basis of 30-day all-cause readmission rates among patients with heart failure. *Circ Cardiovasc Qual Outcomes* 2008;**1**: 29-37. doi:10.1161/CIRCOUTCOMES.108.802686
9. McCormick N, Lacaille D, Bhole V, Avina-Zubieta JA. Validity of heart failure diagnoses in administrative databases: a systematic review and meta-analysis. Guo Y Ed. *PLoS ONE*. 2014;**9**: e104519. doi:10.1371/journal.pone.0104519.
10. Averbuch T, Lee SF, Mamas MA, Oz UE, Perez R, Connolly SJ, *et al.* Derivation and validation of a two-variable index to predict 30-day outcomes following heart failure hospitalization. *ESC Heart Fail* 2021;**8**:2690-2697. doi:10.1002/ehf2.13324
11. van Walraven C, Dhalla IA, Bell C, Etchells E, Stiell IG, Zarnke K, *et al.* Derivation and validation of an index to predict early death or unplanned readmission after discharge from hospital to the community. *Can Med Assoc J* 2010;**182**:551-557. doi:10.1503/cmaj.091117
12. Van Spall HG, Averbuch T, Lee SF, Oz UE, Mamas MA, Januzzi JL, *et al.* The LENT index predicts 30 day outcomes following hospitalization for heart failure. *ESC Heart Fail* 2021;**8**:518-526. doi:10.1002/ehf2.13109
13. Averbuch T, Lee SF, Zagorski B, Mebazaa A, Fonarow GC, Thabane L, *et al.* Effect of a transitional care model following hospitalization for heart failure: 3-year outcomes of the Patient-Centered Care Transitions in Heart Failure (PACT-HF) randomized controlled trial. *Eur J Heart Fail* 2024;**26**:652-660. doi:10.1002/ejhf.3134
14. Dharmarajan K, Hsieh AF, Lin Z, Bueno H, Ross JS, Horwitz LJ, *et al.* Diagnoses and timing of 30-day readmissions after hospitalization for heart failure, acute myocardial infarction, or pneumonia. *JAMA* 2013;**309**:355-363. doi:10.1001/jama.2012.216476
15. Udani K, Patel D, Mangano A. A retrospective study of admission NT-proBNP levels as a predictor of readmission rate, length of stay and mortality. *HCA Healthc J Med [Internet]* 2021;**2**: 207-214. doi:10.36518/2689-0216.1143 Available from: <https://scholarlycommons.hcahealthcare.com/hcahealthcarejournal/vol2/iss3/9>
16. Fuery MA, Leifer ES, Samsky MD, Sen S, O'Connor CM, Fiuzat M, *et al.* Prognostic impact of repeated NT-proBNP measurements in patients with heart failure with reduced ejection fraction. *JACC Heart Fail* 2024;**12**:479-487. doi:10.1016/j.jchf.2023.11.007
17. Rubio-Gracia J, Demissei BG, Ter Maaten JM, Cleland JG, O'Connor CM, Metra M, *et al.* Prevalence, predictors and clinical outcome of residual congestion in acute decompensated heart failure. *Int J Cardiol* 2018;**258**:185-191. doi:10.1016/j.ijcard.2018.01.067
18. Sposito AC, Ramires JAF, Jukema JW, Molina JC, Silva PMD, Ghadanfar MM, *et al.* Physicians' attitudes and adherence to use of risk scores for primary prevention of cardiovascular disease: cross-sectional survey in three world regions. *Curr Med Res Opin* 2009;**25**:1171-1178. doi:10.1185/03007990902846423
19. Yazdan-Ashoori P, Lee SF, Ibrahim Q, Van Spall HGC. Utility of the LACE index at the bedside in predicting 30-day readmission or death in patients hospitalized with heart failure. *Am Heart J* 2016;**179**: 51-58. doi:10.1016/j.ahj.2016.06.007
20. Gevaert AB, Tibebe S, Mamas MA, Ravindra NG, Lee SF, Ahmad T, *et al.* Clinical phenogroups are more effective than left ventricular ejection fraction categories in stratifying heart failure outcomes. *ESC Heart Fail* 2021;**8**: 2741-2754. doi:10.1002/ehf2.13344