Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Preoperative assessment of 180-day outcomes after first tibial transverse transport for severe diabetic foot ulcers

QiKai Hua

hqk100@yeah.net

The First Affiliated Hospital of Guangxi Medical University

ShiKai Lin

The First Affiliated Hospital of Guangxi Medical University

ZhiXue Chen

The First Affiliated Hospital of Guangxi Medical University

Yan Cui

Zhongshan Hospital of Xiamen University

HongJie Su

The First Affiliated Hospital of Guangxi Medical University

RuiQing Mo

The First Affiliated Hospital of Guangxi Medical University

Peilin zhou

Guangxi Medical University First Affiliated Hospital: The First Affiliated Hospital of Guangxi Medical University

XingYu Chen

The First Affiliated Hospital of Guangxi Medical University

SiYuan Wang

Zhongshan Hospital Affiliated to Xiamen University

ZhongWei Jia

Shanxi Provincial People's Hospital

Zhun Wei

The First Affiliated Hospital of Guangxi Medical University

LinHua Wang

The First Affiliated Hospital of Hunan University of Chinese Medicine

Feng Huang

Institute of Cardiovascular Diseases

Article

Keywords:

Posted Date: November 4th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11510631/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

Preoperative assessment of 180-day outcomes after first tibial transverse transport for severe diabetic foot ulcers

Shikai Lin[1†], Zhixue Chen[1†], Yan Cui[2†], Hongjie Su[1], Ruiqing Mo[1], Peilin Zhou[1], Xingyu Chen[1], Siyuan Wang[2], Qikai Hua[1*], Zhongwei Jia[3*], Zhun Wei[1*], Linhua Wang[4*], Feng Huang[1*]

1 Department of Diabetic Foot Limb Salvage and Limb Ulcer Repair, The First Affiliated Hospital of Guangxi Medical University, Nanning 530021, Guangxi, China

2 Department of Orthopaedic Trauma, Zhongshan Hospital, School of Medicine, Xiamen University, Xiamen 361004, Fujian, China

3 Department of Orthopaedics, Shanxi Provincial People's Hospital, Taiyuan 030012, Shanxi, China

4 Department of Limb and Joint Surgery, The First Affiliated Hospital of Hunan University of Chinese Medicine, Changsha 410007, Hunan, China

†These authors contributed equally: Shikai Lin, Zhixue Chen, Yan Cui.

*Corresponding authors: Qikai Hua (hqk100@yeah.net); Zhongwei Jia (doctor_jiazw@sina.com); Zhun Wei (weizhun129@163.com); Linhua Wang (155053648@qq.com); Feng Huang (huangfeng3000@126.com).

Primary corresponding author: Qikai Hua (hqk100@yeah.net). Department of Diabetic Foot Limb Salvage and Limb Ulcer Repair, The First Affiliated Hospital of Guangxi Medical University, 6 Shuangyong Road, Nanning 530021, Guangxi, China. Tel: +86 151 7712 5661.

Abstract

Patients with severe diabetic foot ulcers need preoperative estimates of recovery, survival and limb preservation. We develop a six-predictor score for healing or persistent non-healing with survival and limb preservation, or ipsilateral below-knee amputation or all-cause death, within 180 days after first tibial transverse transport. Development includes 423 consecutive patients and external validation includes 213 patients from three centres. Among 205 patients with complete inputs, the severe-outcome area under the receiver operating characteristic curve is 0.895 (95% confidence interval 0.803–0.963). Compared with a four-predictor reference, the score reduces overall three-outcome probability error but increases severe-outcome probability error. Mean predicted severe risk is 15.0%, compared with 5.9% observed. Here, we show that combined local and systemic assessment provides quantitative prognostic information across distinct recovery outcomes, while external risk overestimation establishes the need for local recalibration.

Introduction

Severe diabetic foot ulcers (DFUs) commonly involve deep infection, tissue necrosis, or ischaemia and place patients at risk of non-healing, major amputation, and death.[1,2] Multimodal limb-salvage care combines infection control, debridement, offloading, and revascularisation when indicated to promote healing and preserve foot function.[3,4,5]

Tibial transverse transport (TTT) uses transverse distraction of a tibial cortical window as part of multimodal limb-salvage care for severe, recalcitrant DFUs. Clinical studies suggest improvements in distal perfusion.[6,7,8] A multicentre cohort reported a 2-year major amputation rate of 4.9% among 1072 surviving patients with complete follow-up.[7] Persistent non-healing among patients who survive with their limb preserved requires continuing local treatment and specialist follow-up.[9] Preoperative counselling therefore needs to address recovery and the care that patients may subsequently need.

TTT studies have examined associations of inflammatory markers, foot temperature, and perfusion with postoperative healing.[10,11] Other DFU cohorts have linked heart failure, end-stage renal disease, and larger wounds to non-healing, supporting assessment of local disease and systemic health.[12] Existing ulcer classifications provide limited individual prognostic information.[13] Machine learning studies have predicted amputation level and composite outcomes comprising non-healing, ulcer recurrence, amputation, or death.[14,15,16,17] For patients undergoing first TTT, the added prognostic value of a quantitative assessment of coexisting local and systemic factors remains unclear. Its applicability across centres also requires assessment.

Here, we develop a preoperative score for 180-day outcomes in consecutive patients undergoing first TTT for Wagner grade 3–5 DFUs and validate it at three independent centres. Outcomes distinguish healing and persistent non-healing among patients alive with their limb preserved from ipsilateral below-knee amputation or all-cause death. Comparisons with ulcer classifications and a parsimonious clinical

reference identify improved overall three-outcome probability prediction, with local calibration required before clinical use. The intended use is to inform prognostic discussions, continuing wound care, and perioperative planning.

Results

Study cohorts and 180-day outcomes

Of 965 TTT/TPD records, 540 were excluded for absence of diabetes, receipt of TPD, or Wagner grade 1–2 disease, leaving 425 eligible first-TTT patients. Excluding two with indeterminate 180-day outcomes left 423 patients (median age 64 years [IQR 57–71.5]; 318 [75.2%] male). Outcomes were healing in 319 (75.4%), persistent non-healing in 61 (14.4%), and severe outcomes in 43 (10.2%). The severe category comprised 28 deaths (6.6%) and 19 below-knee amputations (4.5%), including four amputations followed by death; each patient was counted once. Previous anti-infective treatment, debridement, and revascularisation were documented in 310 (73.3%), 179 (42.3%), and 70 (16.5%) patients, respectively (Table 1; Supplementary Table 1; Fig. 1).

Death-related records documented heart/cardiorenal failure in ten patients, renal injury/failure in one, and severe infection/necrotising fasciitis in ten. These non-exclusive labels included two patients with both heart failure and infection. Median time to death was 64 days after TTT; 21 deaths (75.0%) occurred after day 30 (Supplementary Table 2).

External screening identified 219 patients. Excluding six Shanxi patients with Wagner grade 2 left 213: Shanxi 94, Xiamen 81, and Hunan 38. Outcomes were healing in 163 (76.5%), persistent non-healing in 38 (17.8%), and severe outcomes in 12 (5.6%): six ipsilateral below-knee amputations and six deaths, without overlap.

Preoperative score and predictive performance

The selected predictors represented systemic health (age, heart failure, and eGFR), wound extent (wound area), local conditions (palpated foot temperature), and infection–inflammation burden (IBG). Selection considered performance, availability, and implementation burden (Supplementary Table 17). The full development cohort yielded a 0–59-point card (Table 2).

The original post-selection five-fold out-of-fold assessment yielded a cumulative mAUC of 0.8047 (95% CI 0.7462–0.8497). Repeated nested assessment, which reselected predictor identities within a six-predictor budget, yielded a mean mAUC of 0.7813 (SD 0.0151; Supplementary Table 4).

In the subsequent matched six-predictor assessment (SF_align), one-versus-rest AUCs were 0.782, 0.683, and 0.828 for healing, persistent non-healing, and severe outcomes. Persistent non-healing was less well discriminated (AUC 0.683, 95% CI 0.618–0.746; calibration slope 1.027, 0.637–1.529; Supplementary Table 5D; Supplementary Fig. 5).

The automated workflow, which reselected predictors and configurations within training data, yielded mAUC 0.7825 (95% CI 0.7262–0.8318). Matched procedures retaining six identities performed similarly with retained or re-estimated IBG thresholds (Supplementary Tables 5–6, 16, and 27; Supplementary Fig. 8).

The same-centre temporal analysis used 302 patients treated earlier (2020–2024) for development and 121 patients treated later (2025–2026) for evaluation, retaining overall-development predictor identities and upstream rules. Later-cohort mAUC was 0.7810 (95% CI 0.6726–0.8665; Supplementary Table 15; Supplementary Fig. 3).

Prognostic information beyond ulcer classification

SINBAD and UT classifications were reviewed against source records for all 423 patients. SINBAD depth, area, and infection components were uniformly 1; site, neuropathy, and ischaemia varied. All patients had UT grade 3 disease, with 315 at stage B and 108 at stage D. The UT comparison therefore assessed ischaemia within infected deep wounds. In the subsequent matched repeated assessment, paired mAUC gains over SINBAD and UT stage were 0.1689 and 0.2236; multiclass Brier score decreased by 0.0519 and 0.0611, respectively. The 95% CIs supported these directions, and automated-workflow comparisons were consistent (Supplementary Tables 5–6; Supplementary Fig. 9A,B).

The four-predictor reference combined age, heart failure, eGFR, and wound area. In matched required-input analyses, six- versus four-predictor severe-outcome Brier scores were 0.0701 versus 0.0734 (difference −0.0032, 95% CI −0.0075 to 0.0010). Cumulative mAUC increased by 0.0424 (0.0061–0.0761), and multiclass Brier decreased by 0.0168 (difference −0.0168, −0.0283 to −0.0057; Supplementary Table 7).

In exploratory component analyses, composite severe-risk probabilities yielded death AUCs of 0.7804 versus 0.7769 for required-input six- versus four-predictor models (difference 0.0036, 95% CI −0.0491 to 0.0559). Amputation AUCs were 0.8811 versus 0.8188 (difference 0.0623, 0.0133–0.1256; Supplementary Table 11).

In matched regression models, removing foot temperature reduced mAUC by 0.0388 (difference −0.0388, 95% CI −0.0729 to −0.0039). Removing IBG increased multiclass Brier score by 0.0090 (0.0018–0.0168). These comparisons indicated contributions to discrimination and probability prediction, respectively (Supplementary Tables 12–13; Supplementary Fig. 9C,D).

Replacing IBG with graded CRP in the same integer-scoring and repeated-evaluation procedure yielded a multiclass Brier difference (CRP−IBG) of −0.0017 (95% CI −0.0075 to 0.0039). Neither version showed a clear advantage. Complete observed inputs were available in 301/423 patients (71.2%) for IBG and 396/423 (93.6%) for CRP. Nine severe outcomes occurred among the 122 patients with incomplete IBG inputs. The locked calculator returned no risk estimate when required inputs were incomplete; full-cohort

internal evaluation used within-fold imputation. Required-input comparisons gave similar results (Supplementary Tables 8–9, 14D, and 19). A secondary CRP–four-predictor comparison yielded a severe-outcome Brier difference of −0.0046 (−0.0092 to −0.0003), without multiplicity adjustment (Supplementary Table 7).

Outcome differences across risk strata

In the subsequent matched repeated assessment, training-fold low/intermediate/high boundaries were applied to the corresponding evaluation folds. Mean any-adverse-outcome rates across repetitions were 15.4%, 38.8%, and 72.0%; severe-outcome rates were 4.1%, 16.5%, and 50.5% (Fig. 2A; Supplementary Table 10).

With training-QWK boundaries used for category assignment, mean sensitivities were 80.3%, 31.8%, and 37.9% for healing, persistent non-healing, and severe outcomes, respectively. Severe-outcome specificity was 95.6% (Fig. 2B; Supplementary Table 10). Class-specific probabilities are reported in Supplementary Table 5D. Across ten repetitions, 141/423 patients (33.3%) entered at least two strata, showing that assignment varied with training partitions (Fig. 2C; Supplementary Tables 5 and 10).

Additional preoperative cardiorenal review scenarios

Prespecified severe-risk thresholds of 5%, 10%, 15%, and 20% defined scenarios for additional cardiorenal review and perioperative planning after routine assessment. At 10%, the required-input six-predictor procedure flagged 29.1 patients per 100 and identified 73.3% of severe outcomes. Corresponding values for the four-predictor reference were 26.5 and 66.3%. Incremental net benefit was 0.50 per 100 (95% CI −0.37 to 1.41; Supplementary Table 18; Supplementary Fig. 4).

Three-centre external validation

The locked IBG score was directly calculable in 205/213 external patients (96.2%), including all 12 patients with severe outcomes. The CRP availability pathway covered 211 patients (99.1%) and the four-predictor reference 212 (99.5%). All 38 Hunan patients had complete score inputs. Table 1 and Supplementary Table 20 report centre-specific characteristics and input availability; matched comparisons used 205 patients.

The locked IBG score had an AUC of 0.895 (95% CI 0.803–0.963) for severe outcomes and 0.936 (0.897–0.971) for any adverse outcome. Centre-restricted severe-outcome AUC was 0.906 (0.828–0.969). Among the same 205 patients, Brier differences versus the four-predictor reference were 0.0160 (0.0019–0.0292) for severe outcomes and −0.0353 (−0.0677 to −0.0057) across all three outcomes (Fig. 3; Supplementary Tables 21 and 28).

Mean predicted severe-outcome risk was 15.0%, compared with 5.9% observed (O/E 0.391, 95% CI 0.202–0.684). The calibration slope was 0.976 (0.574–1.478). The class-specific calibration slope for

persistent non-healing was 2.291 (1.475–3.325), indicating residual miscalibration for that category (Supplementary Tables 21–22; Supplementary Fig. 5).

Review scenarios in external patients

At 10%, IBG flagged 77/205 patients, identified 11/12 severe outcomes, and missed one. The four-predictor reference flagged 82, identified nine, and missed three. At 5%, both identified all 12, flagging 110 and 95 patients, respectively (Table 3; Supplementary Table 24).

Supplementary evaluation of baseline-risk updating

Updating only the two cumulative intercepts reduced the cross-validated severe-outcome Brier score to 0.0439. Against the similarly updated four-predictor reference, Brier differences were −0.0022 (95% CI −0.0085 to 0.0039) for severe outcomes and −0.0732 (−0.0987 to −0.0498) across all three outcomes. Calibration shifts persisted in leave-one-centre-out evaluation. Hunan had no severe outcomes, precluding centre-specific AUC estimation (Supplementary Tables 23–26; Supplementary Figs. 6–7).

Discussion

We assessed three 180-day outcomes after first TTT for severe DFUs: healing, persistent non-healing with survival and limb preservation, and amputation or death. Combining six preoperative predictors improved overall three-outcome probability prediction in development and external comparisons. The four-predictor reference provided substantial discrimination of severe outcomes. The clinical contribution is a quantitative assessment of coexisting local and systemic factors that supports discussion of recovery alongside the possibility of prolonged care.

This framework brings together wound-related and patient-related outcomes, a distinction described by Jeffcoate and colleagues.[18] A recent DFU core outcome set likewise includes healing, amputation, death, and quality of life.[19] A previous multicentre TTT study excluded follow-up deaths from healing and limb-preservation assessment.[7] Here, healing required wound closure, survival, and limb preservation, with deaths and severe events after wound closure retained within the 180-day window. Most development-cohort deaths occurred in ICUs; death-related records documented heart/cardiorenal failure and severe infection, and 75% occurred after postoperative day 30 (Supplementary Table 2). These findings support continued attention to organ function and infection management throughout recovery.

Persistent non-healing represents an ongoing need for treatment and support. For the 61 patients with unhealed wounds at 180 days, recovery expectations extend to dressings, offloading, specialist follow-up, and family assistance. Ribu and colleagues associated healing with improved social functioning and mental health, whereas social functioning declined with persistent ulceration.[20] EURODIALE documented continuing local treatment and specialist consultations, with hospitalisation a major cost driver.[9] These studies connect prolonged wound recovery with daily activity and care needs that warrant discussion before surgery.

Wound repair after TTT depends on adequate blood supply, infection control, and viable tissue.[4,5] In an early TTT study, four major amputations were associated with extensive arterial thrombosis or unreconstructed severe popliteal disease.[6] This illustrates how persistent or new vascular compromise can interrupt repair. Salvageability of tissue and weight-bearing structures also shapes treatment goals. Prognostic assessment therefore needs to consider local repair conditions together with the patient's capacity to tolerate continuing treatment.

Cardiorenal status helps characterise this capacity. EURODIALE linked older age, heart failure, end-stage renal disease, and larger wounds to non-healing.[12] Meloni and colleagues observed poorer survival and limb outcomes with heart failure and dialysis among revascularised patients with ischaemic DFUs.[21] Larger wounds increase the tissue requiring repair. Hypoperfusion in heart failure and vascular and host-defence abnormalities in renal disease could affect perfusion, infection responses, and treatment tolerance.[12,21] Patients with similar local disease severity can therefore follow different recovery trajectories. Existing TTT treatment classifications emphasise cardiorenal management; our study integrates this information with wound characteristics to estimate prognosis before first TTT.

Local ulcer classifications were concentrated in severe categories: all patients had Wagner grade 3–5 and UT grade 3 disease. SINBAD depth, area, and infection components were constant. Within these categories, wound extent and systemic illness still varied. Our comparisons assessed the prognostic information provided by this variation, including contributions from foot temperature and infection–inflammation markers.

Palpated foot temperature provides a bedside observation beyond wound extent. Liu and colleagues reported an association between measured foot temperature and transcutaneous oxygen pressure.[11] In our matched analysis, removing palpated temperature reduced discrimination. Low temperature may suggest impaired perfusion, whereas warmth may reflect relatively preserved local perfusion or accompany inflammation. Both findings should be interpreted alongside vascular assessment and wound examination. IWGDF/IDSA advises against using foot temperature to diagnose soft-tissue infection.[5] Prospective standardisation could establish how palpation grades relate to objective temperature and perfusion measurements.

Inflammatory markers also reflect interactions between local disease and systemic illness. Jeandrot and colleagues found that combining CRP and PCT improved discrimination of mild DFU infection.[22] Heart-failure studies associated CRP and low-level PCT elevations with poorer prognosis.[23,24] Raised CRP in heart failure can therefore reduce specificity for infection while retaining prognostic information. Infection diagnosis and 180-day prognosis address different questions. IBG integrates leucocyte responses, acute-phase responses, and PCT, with higher grades requiring contributions across domains. It summarises the breadth of laboratory responses; clinical findings remain necessary to determine infection site, severity, and systemic involvement.[5]

IBG and CRP offer different practical advantages. IBG integrates several laboratory domains, and its removal increased three-outcome probability error (Supplementary Table 13). CRP reduces testing requirements: complete inputs were available in 93.6% of development patients, compared with 71.2% for IBG. The external availability pathway covered 99.1% of patients (Supplementary Tables 9 and 20). The direct comparison showed no clear difference in overall three-outcome probability prediction (Supplementary Table 8).

Absolute risk accuracy is central to counselling patients at another centre. The locked score retained severe-outcome risk ranking externally but overestimated mean risk and had greater severe-outcome probability error than the four-predictor reference. Persistent non-healing also remained miscalibrated. Lower event rates, case mix, and care pathways could contribute to these differences. Because overestimation can increase review volume, implementation requires assessment of local risk levels and review capacity.[25,26]

Baseline-risk updating reduced probability error while retaining the original points. Against the similarly updated reference, multiclass Brier score improved, whereas the severe-outcome Brier difference interval included zero. Calibration shifts varied across held-out centres, so prospective evaluation of updated parameters should assess local risk levels and calibration over time.

After local calibration, the score could support preoperative counselling and perioperative planning by distinguishing the prospect of prolonged wound care from the risk of amputation or death. Anticipated non-healing can guide discussions of dressings, offloading, follow-up, and family support. Higher severe-outcome risk could prompt additional assessment of heart-failure stability, recent renal-function changes, and current treatment. This review would complement vascular assessment, infection control, and decisions about limb-salvage timing.[3,4,5] Planning should also address anaemia and nutrition; albumin and prealbumin require interpretation alongside inflammation, intake, and weight change.[27,28] Treatment goals depend on tissue salvageability, systemic health, and patient preferences.

The threshold for additional review should reflect local capacity, the consequences of missed events, and patient burden. Lower thresholds identified more severe outcomes but also flagged more patients for review.

The score was developed retrospectively at one centre. External validation included 12 severe outcomes and none in Hunan, limiting precision for incremental severe-outcome performance and stability across centres. Patients had already been selected for first TTT, sometimes after anti-infective treatment, debridement, or revascularisation. Predictions therefore describe their latest preoperative state within existing care pathways. Assessing risk changes after treatment of reversible conditions requires longitudinal data, prespecified new prediction times, and documented treatment.

Foot-temperature records retained consensus grades, so interobserver agreement requires prospective assessment. SINBAD neuropathy was retrospectively coded from clinical diagnoses. Predictors and some grading rules were determined during development, leaving potential selection-related optimism in internal evaluation. Updated parameters estimated in the full sample also require independent evaluation. Persistent non-healing was less well discriminated, and stratum assignments varied across training partitions. For counselling, calibrated continuous probabilities would complement subsequent wound reassessment. Men predominated in the cohorts; performance across sex and other sociodemographic groups was not assessed.

Actual additional reviews beyond routine care and patient care burden were not measured. Prospective implementation studies should record new or earlier reviews, management changes, and 180-day outcomes. They should also assess function, quality of life, and care burden.[19]

Methods

Study design

This retrospective study comprised single-centre model development and three-centre external validation. The development cohort included consecutive patients at The First Affiliated Hospital of Guangxi Medical University from January 1, 2020, to April 1, 2026. The hospital's Medical Records Department supplied original records for 965 departmental patients treated with TTT or tibial periosteal distraction (TPD). Prediction was anchored after the decision to undertake first TTT, when the latest preoperative information was available but before surgery (decision-to-TTT). Outcomes were assessed for 180 days after first TTT. The study was approved by the Medical Ethics Committee of The First Affiliated Hospital of Guangxi Medical University on September 24, 2026 (2026-E0867). Informed consent for secondary use of the development-cohort records was waived. The study was not registered. Patients and the public were not involved in research design, conduct, or interpretation.

Participants

We included consecutive patients undergoing first TTT for Wagner grade 3–5 DFUs. Exclusions were absence of diabetes, receipt of TPD, Wagner grade 1–2 ulcers, or indeterminate 180-day outcomes. Each patient contributed their first TTT procedure and was the unit of analysis. Index-wound and amputation outcomes are defined below.

Consecutive patients undergoing first TTT were identified at Shanxi Provincial People’s Hospital, Zhongshan Hospital, Xiamen University, and The First Affiliated Hospital of Hunan University of Chinese Medicine. The respective ascertainment periods were January 1, 2020–April 1, 2026; January 14, 2024–April 1, 2026; and December 8, 2018–April 1, 2026. Eligibility required Wagner grade 3–5 ulcers, with the same three 180-day outcomes and latest preoperative measurement rule as in development. Each centre verified first-TTT dates and outcomes. All six amputations were ipsilateral below-knee

amputations within 180 days, with no subsequent death within that window (Supplementary Tables 20–26).

Procedures

TTT was delivered within multimodal limb-salvage care, including infection control, debridement and wound care, offloading, and metabolic management. Revascularisation was guided by vascular assessment and clinical indications.[6,7,8] The clinical team individualised treatment for each patient.

The 48 candidate variables covered demographics, medical history, cardiorenal function, blood counts and inflammation, and wound and vascular characteristics (Supplementary Table 1; Supplementary Data candidate-variable dictionary). For each test or clinical assessment, we selected the latest completed preoperative result using its original measurement date. Medical history was extracted from preoperative records. Sex was extracted as recorded in the clinical record.

Heart failure was classified from the preoperative cardiology diagnosis and latest preoperative echocardiogram as 0, none; 1, chronic; or 2, acute. Diagnosis combined clinical findings with objective structural or functional cardiac abnormalities.[29] Two treating physicians agreed on the latest preoperative foot-temperature palpation grade: cold, cool, warm, or warmer. The database retained the final consensus grade.

Renal function was classified using laboratory-reported estimated glomerular filtration rate (eGFR), standardised to 1.73 m^2. Grades 1–5 corresponded to ≥ 90, $60 \leq eGFR < 90$, $30 \leq eGFR < 60$, $15 \leq eGFR < 30$, and $eGFR < 15$ mL/min/1.73 m^2 or documented uraemia. Wound area was estimated as maximum length × maximum width and summed across wounds. The four area (A) categories were $A < 5$, $5 \leq A \leq 20$, $20 < A \leq 100$, and $A > 100$ cm^2 (Supplementary Table 19).

Infection–inflammation burden grade (IBG) combined white blood cell count (WBC), neutrophil-to-lymphocyte ratio (NLR), erythrocyte sedimentation rate (ESR), C-reactive protein (CRP), and procalcitonin (PCT). Each marker received 0–2 component points, combined as A=max(WBC,NLR), B=max(ESR,CRP), and C=PCT. Totals of 0–2, 3–4, 5, and 6 corresponded to IBG grades 0–3. The main framework retained the thresholds in Supplementary Table 3; sensitivity analyses re-estimated NLR and ESR thresholds within training folds. Automated selection treated IBG and its individual components as mutually exclusive.

The neuropathy component of the site, ischaemia, neuropathy, bacterial infection, area, and depth (SINBAD) classification used documented type 2 diabetic peripheral neuropathy as a retrospective proxy for loss of protective sensation. University of Texas (UT) grade 3 denoted extension to bone or joint; stage reflected recorded infection and ischaemia. Following source review, stages were B (infection without ischaemia) or D (both). SINBAD discrimination was assessed using the raw total; an ordinal

logistic probability map was fitted within training folds. UT used a stage D indicator with B as reference, and Wagner grade was coded numerically. All benchmarks used the same patients and outer folds.

Outcomes

The 180-day categories were Y=0: complete epithelialisation of all ipsilateral index wounds by day 180, sustained for ≥1 week, with survival and no ipsilateral below-knee amputation; Y=1: healing criteria unmet, with survival and no such amputation; and Y=2: ipsilateral below-knee amputation or all-cause death, including after wound closure. Patients with both severe events counted once as Y=2. Follow-up after day 180 could confirm 1-week persistence of late healing but did not extend the amputation/death window. Minor amputations were classified by wound healing; new or recurrent ulcers after first confirmed healing did not redefine persistent non-healing. Records, outpatient notes, photographs, and telephone follow-up were reviewed without reference to the score; three healed outcomes were confirmed by telephone alone. Final follow-up verification was completed on October 1, 2026. Y≥1 denoted any adverse outcome; Y=2 denoted severe outcomes.

Most development-cohort deaths occurred in intensive care units (ICUs). ICU physicians determined the principal clinical reasons for death contemporaneously; diabetic-foot specialists and ICU physicians jointly reviewed these determinations using original hospital records. Supplementary Table 2 separately reports non-exclusive labels for coexisting heart/cardiorenal failure, renal injury, severe infection/necrotising fasciitis, and other conditions recorded in the source workbook's special-reasons field. Amputation and death dates were extracted separately and measured from first TTT, preserving event sequence within the 180-day window.

Statistical analysis

Reporting followed STROBE and TRIPOD+AI.[30,31,32] Supplementary Methods provide full development procedures, sensitivity analyses, and methodological references.

Missing data and score development

Twenty imputations were generated within training folds using IterativeImputer with BayesianRidge and posterior sampling. Training imputations could use training outcomes; evaluation folds used a separate outcome-blind imputer fitted to training data. Imputed values were constrained to valid training ranges and categories before reconstructing IBG. Supplementary Methods describe missingness exclusions and auxiliary inputs; Supplementary Table 14 reports PCT-testing patterns.

Within the AutoScore-Ordinal framework, random-forest ranking and one-versus-rest L1-penalised screening informed selection of the six predictors. Selection considered discrimination, complexity, redundancy, missingness, and collection burden.

The automated workflow selected configurations and predictors within inner folds; matched six-predictor analyses retained predictor identities (Supplementary Table 4).

Proportional-odds cumulative logistic coefficients were scaled, rounded to integer points, and constrained by clinical direction. The six-predictor specification used age ≥65 years and a 60-point target. One full-cohort imputation (M=1) yielded the 0–59-point card; internal evaluation refitted points within training folds and pooled M=20 predictions. Supplementary Table 16 assesses imputation sensitivity.

Cumulative logistic maps converted scores into three outcome probabilities, with alignment estimated from training data only. Training-fold stratum boundaries maximised quadratic weighted kappa (QWK). The frozen boundaries of 24 and 34 defined scores of 0–23, 24–33, and 34–59, separately from probability-threshold review scenarios. External validation applied the locked card and maps fitted in 301 complete-input development records; the four-predictor reference maps used 405. Calculators returned no estimate when required inputs were missing (Supplementary Tables 10 and 19).

Internal validation and clinical comparators

The original post-selection five-fold out-of-fold (OOF) evaluation fixed predictor identities and upstream specifications while refitting points and maps in training folds. Repeated nested assessment reselected identities within a six-predictor budget. Subsequent matched analyses used ten repetitions of stratified five-fold outer evaluation, with five-fold inner training. Imputation, points, and maps were fitted within training partitions.

The mapping analysis compared aligned with original maps using identical patients, folds, and imputations. Its prespecified primary contrast was the multiclass Brier difference with training-re-estimated IBG thresholds (Supplementary Tables 4–6).

The IBG–CRP comparison retained the other five predictors and scoring rules and used identical folds and M=20 imputations. CRP grades were 0 for <10, 1 for 10 to <100, and 2 for ≥100 mg/L. Its primary contrast was CRP minus IBG multiclass Brier under common-auxiliary imputation; required-input sensitivity analyses restricted each imputer to that version's inputs (Supplementary Tables 8–9).

A same-centre temporal sensitivity analysis used 302 patients treated in 2020–2024 for development and 121 patients treated in 2025–2026 for evaluation. Points, imputers, maps, and boundaries were fitted in the earlier period; predictor identities and upstream specifications originated from overall development.

Matched foot-temperature removal used identical folds, M=20 imputations, categorical coding, and joint ordinal likelihood. Temperature remained an imputation auxiliary but was omitted from the prediction equation (Supplementary Table 12).

External validation and supplementary updating

External evaluation applied the locked IBG and four-predictor specifications to the same complete-input patients, with severe-outcome Brier difference as the primary contrast. The supplementary CRP version was fitted only in development data and locked after inspection of earlier two-centre external results. The

availability pathway used IBG when complete and CRP otherwise when its required inputs were complete. Supplementary Methods describe detection-limit and uraemia coding; a matched sensitivity analysis required numerical eGFR.

Pooled and centre-specific evaluation assessed areas under the receiver operating characteristic curves (AUCs), probability error, calibration, and threshold-specific net benefit. Centre-stratified patient bootstrap resampling (2000 draws) estimated 95% confidence intervals (CIs) for AUCs, Brier scores, and paired differences. Event rates, threshold sensitivity, and positive predictive value (PPV) used exact binomial intervals. AUC and calibration slope were not estimable without events. Centre-restricted AUC used only within-centre event/non-event pairs to assess case-mix effects.[25] All reported 95% confidence intervals were two-sided.

Supplementary updating estimated two cumulative intercept offsets while retaining points and slopes and enforcing coherent probabilities. Identical partitions supported 20 repetitions of five-fold patient cross-validation and leave-one-centre-out evaluation; 500 bootstrap refits assessed optimism. Original locked-model validation, held-out evaluation of updating, and full-sample candidate parameters were reported separately.

Performance estimation and clinical review scenarios

Cumulative mean AUC (mAUC) averaged the AUCs for $Y \geq 1$ and $Y = 2$. Multiclass Brier was the patient-level mean of summed squared errors across three outcome probabilities; severe-outcome Brier evaluated $Y = 2$.[26] Class-specific analyses assessed one-versus-rest AUC, binary Brier, calibration, and observed-to-expected ratios (O/E). Repeated estimates averaged ten complete OOF evaluations; standard deviation (SD) described variation across repetitions. Paired patient-cluster bootstrap resampling (2000 draws) retained each patient's ten predictions together; intervals were conditional on fixed OOF predictions.

Additional matched analyses assessed shared slopes, ejection-fraction substitutions, IBG removal/replacement, and component outcomes (Supplementary Methods). The four-predictor reference used age, heart failure, eGFR, and wound area, with the same folds, $M = 20$ imputations, and scoring procedure. Required-input comparisons restricted imputation to each model's inputs and primarily assessed severe-outcome Brier.

Exploratory component analyses used saved composite severe-risk probabilities to rank deaths and ipsilateral below-knee amputations separately; four patients contributed to both targets. AUC and event identification at 10% used 2000 paired patient-cluster resamples, without multiplicity adjustment (Supplementary Table 11).

Decision curves evaluated additional preoperative cardiorenal review targeting 180-day amputation or death. Thresholds spanned 5–40% in 2.5% increments, with prespecified scenarios at 5%, 10%, 15%, and 20%. Net benefit was $TP/n - FP/n \times p_t/(1-p_t)$, where p_t is the threshold and TP and FP denote flagged

patients with and without severe outcomes. Counts were averaged across repetitions; resampling retained repeated predictions together. Threshold-positive counts represented potential review volume, because actual additional consultations were not recorded.

Sample sizes reflected all eligible consecutive cases identified during each collection period; no prospective model sample-size calculation was performed. Precision was assessed using event counts, confidence intervals, and resampling variability.

Analyses were implemented in Python; the verified reproduction environment used scikit-learn 1.9.0, SciPy 1.18.0, and statsmodels 0.14.6. Stage-specific dependencies are archived with the code. OpenAI Codex assisted with code preparation and computational checks.

Software and manuscript preparation

OpenAI Codex assisted with manuscript structure, language, and figure and table layout during manuscript preparation. The authors are responsible for source-data verification, interpretation, and the manuscript's content.

Data availability

Pseudonymised individual participant data are not publicly shared because of privacy and confidentiality requirements. Requests for access may be directed to the primary corresponding author, Qikai Hua (hqk100@yeah.net), for institutional review under applicable ethical and data-protection requirements. Aggregate data supporting the findings accompany the manuscript and Supplementary Information. Source data are provided with this paper in Source_Data.zip.

Code availability

Selected code for the frozen integer point card is available at https://github.com/JSJ189/ttt-180day-score. The complete analysis code will be released in this repository upon acceptance. Analysis scripts, locked model parameters and a synthetic scoring demonstration are supplied for peer review as Supplementary Software 1. The full archived code and dated analytical protocols can be provided to editors and reviewers during peer review; access to confidential individual participant data remains subject to the conditions in Data availability. The archive covers preprocessing, development, internal evaluation, external validation, baseline-risk updating and sensitivity analyses. Stage-specific dependencies and software versions accompany the code.

References

1. Armstrong, D. G., Tan, T. W., Boulton, A. J. M. & Bus, S. A. Diabetic foot ulcers: a review. *JAMA* **330,** 62–75 (2023). https://doi.org/10.1001/jama.2023.10578

2. Chen, L., Sun, S., Gao, Y. & Ran, X. Global mortality of diabetic foot ulcer: a systematic review and meta-analysis of observational studies. *Diabetes Obes. Metab.* **25,** 36–45 (2023). https://doi.org/10.1111/dom.14840

3. Schaper, N. C. et al. Practical guidelines on the prevention and management of diabetes-related foot disease (IWGDF 2023 update). *Diabetes Metab. Res. Rev.* **40,** e3657 (2024). https://doi.org/10.1002/dmrr.3657

4. Fitridge, R. et al. The intersocietal IWGDF, ESVS, SVS guidelines on peripheral artery disease in people with diabetes and a foot ulcer. *Diabetes Metab. Res. Rev.* **40,** e3686 (2024). https://doi.org/10.1002/dmrr.3686

5. Senneville, É. et al. IWGDF/IDSA guidelines on the diagnosis and treatment of diabetes-related foot infections (IWGDF/IDSA 2023). *Diabetes Metab. Res. Rev.* **40,** e3687 (2024). https://doi.org/10.1002/dmrr.3687

6. Chen, Y. et al. Proximal tibial cortex transverse distraction facilitating healing and limb salvage in severe and recalcitrant diabetic foot ulcers. *Clin. Orthop. Relat. Res.* **478,** 836–851 (2020). https://doi.org/10.1097/CORR.0000000000001075

7. Chen, Y. et al. Effect of tibial cortex transverse transport in patients with recalcitrant diabetic foot ulcers: a prospective multicenter cohort study. *J. Orthop. Translat.* **36,** 194–204 (2022). https://doi.org/10.1016/j.jot.2022.09.002

8. Hua, Q. et al. Chinese Association of Orthopaedic Surgeons (CAOS) clinical guideline for the treatment of diabetic foot ulcers using tibial cortex transverse transport technique (version 2020). *J. Orthop. Translat.* **25,** 11–16 (2020). https://doi.org/10.1016/j.jot.2020.05.003

9. Prompers, L. et al. Resource utilisation and costs associated with the treatment of diabetic foot ulcers. Prospective data from the Eurodiale Study. *Diabetologia* **51,** 1826–1834 (2008). https://doi.org/10.1007/s00125-008-1089-6

10. Zhao, S. et al. The association of the perioperative neutrophil-to-lymphocyte ratio with wound healing in patients with Wagner grade 3 and 4 diabetic foot ulcers after tibial cortex transverse transport surgery: a prospective observational cohort study. *Front. Endocrinol.* **15,** 1420232 (2024). https://doi.org/10.3389/fendo.2024.1420232

11. Liu, N. et al. Application of transverse tibial bone transfer in diabetic foot patients: a comprehensive evaluation of postoperative prognostic factors and risk factors. *Orthop. Surg.* **17,** 2708–2716 (2025). https://doi.org/10.1111/os.70124

12. Prompers, L. et al. Prediction of outcome in individuals with diabetic foot ulcers: focus on the differences between individuals with and without peripheral arterial disease. The EURODIALE Study. *Diabetologia* **51,** 747–755 (2008). https://doi.org/10.1007/s00125-008-0940-0

13. Monteiro-Soares, M. et al. Guidelines on the classification of foot ulcers in people with diabetes (IWGDF 2023 update). *Diabetes Metab. Res. Rev.* **40,** e3648 (2024). https://doi.org/10.1002/dmrr.3648

14. Huang, J. et al. Prediction models for amputation after diabetic foot: systematic review and critical appraisal. *Diabetol. Metab. Syndr.* **16,** 126 (2024). https://doi.org/10.1186/s13098-024-01360-6

15. Xie, P. et al. An explainable machine learning model for predicting in-hospital amputation rate of patients with diabetic foot ulcer. *Int. Wound J.* **19,** 910–918 (2022). https://doi.org/10.1111/iwj.13691

16. Yilmaz, K. B. et al. Machine learning-based prediction of major amputation risk after initial limb-preserving surgery in diabetic foot. *Front. Endocrinol.* **17,** 1889329 (2026). https://doi.org/10.3389/fendo.2026.1889329

17. Pauliukienė, R., Šutienė, K., Čemerkaitė, A. & Čeponis, J. Predictors of unfavorable outcomes in diabetic foot ulcers. *Diagnostics* **15,** 3070 (2025). https://doi.org/10.3390/diagnostics15233070

18. Jeffcoate, W. J., Chipchase, S. Y., Ince, P. & Game, F. L. Assessing the outcome of the management of diabetic foot ulcers using ulcer-related and person-related measures. *Diabetes Care* **29,** 1784–1787 (2006). https://doi.org/10.2337/dc06-0306

19. Staniszewska, A. et al. Development of a core outcome set for studies assessing interventions for diabetes-related foot ulceration. *Diabetes Care* **47,** 1958–1968 (2024). https://doi.org/10.2337/dc24-1112

20. Ribu, L., Birkeland, K., Hanestad, B. R., Moum, T. & Rustoen, T. A longitudinal study of patients with diabetes and foot ulcers and their health-related quality of life: wound healing and quality-of-life changes. *J. Diabetes Complicat.* **22,** 400–407 (2008). https://doi.org/10.1016/j.jdiacomp.2007.06.006

21. Meloni, M. et al. Impact of heart failure and dialysis in the prognosis of diabetic patients with ischemic foot ulcers. *J. Clin. Transl. Endocrinol.* **11,** 31–35 (2018). https://doi.org/10.1016/j.jcte.2018.01.002

22. Jeandrot, A. et al. Serum procalcitonin and C-reactive protein concentrations to distinguish mildly infected from non-infected diabetic foot ulcers: a pilot study. *Diabetologia* **51,** 347–352 (2008). https://doi.org/10.1007/s00125-007-0840-8

23. Anand, I. S. et al. C-reactive protein in heart failure: prognostic value and the effect of valsartan. *Circulation* **112,** 1428–1434 (2005). https://doi.org/10.1161/CIRCULATIONAHA.104.508465

24. Darche, F. F. et al. Low-level elevations of procalcitonin are associated with increased mortality in acute heart failure patients, independent of concomitant infection. *Life* **11,** 1429 (2021). https://doi.org/10.3390/life11121429

25. Riley, R. D. et al. Evaluation of clinical prediction models (part 2): how to undertake an external validation study. *BMJ* **384,** e074820 (2024). https://doi.org/10.1136/bmj-2023-074820

26. Van Calster, B. et al. Calibration: the Achilles heel of predictive analytics. *BMC Med.* **17,** 230 (2019). https://doi.org/10.1186/s12916-019-1466-7

27. Evans, D. C. et al. The use of visceral proteins as nutrition markers: an ASPEN position paper. *Nutr. Clin. Pract.* **36,** 22–28 (2021). https://doi.org/10.1002/ncp.10588

28. Evans, D. C. et al. Corrigendum to The use of visceral proteins as nutrition markers: an ASPEN position paper. *Nutr. Clin. Pract.* **36,** 909 (2021). https://doi.org/10.1002/ncp.10746

29. Chinese Society of Cardiology, Chinese Medical Association, Chinese College of Cardiovascular Physician, Chinese Heart Failure Association of Chinese Medical Doctor Association, Editorial Board of Chinese Journal of Cardiology. Chinese guidelines for the diagnosis and treatment of heart failure 2024. *Zhonghua Xin Xue Guan Bing Za Zhi* **52,** 235–275 (2024). https://doi.org/10.3760/cma.j.cn112148-20231101-00405

30. von Elm, E. et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *PLoS Med.* **4,** e296 (2007). https://doi.org/10.1371/journal.pmed.0040296

31. Collins, G. S. et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ* **385,** e078378 (2024). https://doi.org/10.1136/bmj-2023-078378

32. Collins, G. S. et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ* **385,** q902 (2024). https://doi.org/10.1136/bmj.q902

Acknowledgements

This study was supported by the National Natural Science Foundation of China (82260448 and 82660523), Guangxi Natural Science Foundation (2024GXNSFDA010024), 2024 Xiamen Medical and Health Key Project (3502Z20244004), and Shanxi Provincial Key Medical Research Project (2023XM049). The funders had no role in study design, data collection, analysis, interpretation, supervision, manuscript preparation, or the decision to submit for publication.

Author contributions

S.L. contributed to study conception, data collection and analysis, manuscript writing, and figure and table preparation. Z.C. contributed to model development and data analysis. H.S., R.M., P.Z., X.C., and S.W. contributed to manuscript writing and revision. Y.C., Q.H., Z.J., Z.W., L.W., and F.H. contributed to study conception, supervision, and manuscript writing and revision. S.L. and Z.C. directly accessed and verified the underlying study data.

Competing interests

The authors declare no competing interests.

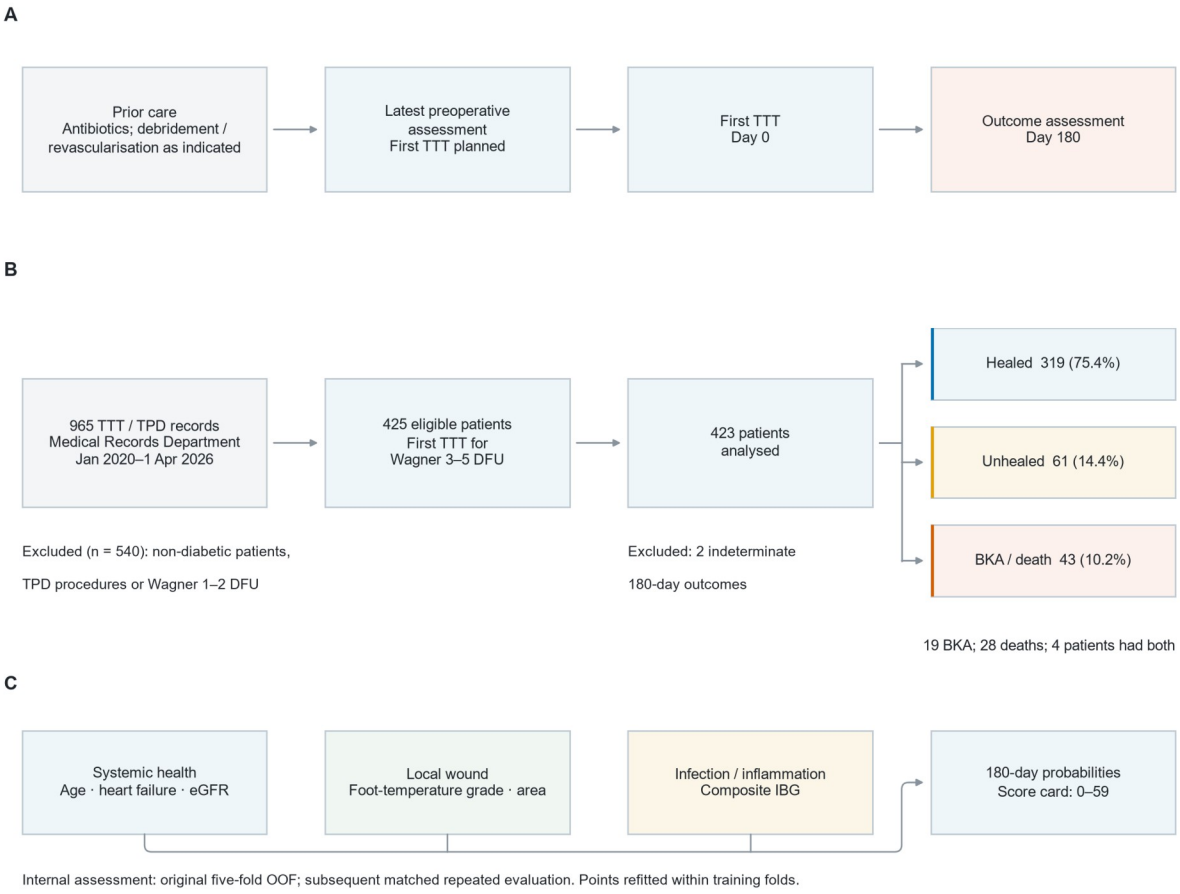
Figure legends

Fig. 1 Clinical timeline and six-predictor risk assessment. (A) Previous treatment, latest preoperative assessment, first TTT, and the 180-day outcome window. (B) Screening of 965 TTT/TPD records excluded 540 patients without diabetes, receiving TPD, or with Wagner grade 1–2 DFUs, leaving 425 candidates. Two patients with indeterminate 180-day outcomes were excluded, leaving 423. Severe outcomes comprised 19 below-knee amputations and 28 deaths, with four amputations followed by death, yielding 43 patients. (C) Six preoperative predictors and risk estimation. The 0–59-point card was derived from the full development cohort; the original five-fold OOF and subsequent matched repeated evaluations re-estimated points within training folds. TTT, tibial transverse transport; TPD, tibial periosteal distraction; DFU, diabetic foot ulcer; BKA, below-knee amputation; eGFR, estimated glomerular filtration rate; IBG, infection–inflammation burden grade; OOF, out-of-fold.

Fig. 2 Risk gradients and category assignment in subsequent matched repeated internal evaluation. (A) Mean observed outcome proportions across ten repetitions within low, intermediate, and high strata defined by training-fold boundaries; labels show mean stratum sizes. (B) Rows show observed outcomes and columns assigned strata; cells show mean within-outcome assignment proportions across repetitions. (C) Number of distinct strata assigned to each patient. Boundaries maximised training-set quadratic weighted kappa (QWK), separately from the frozen development-derived 24/34 point boundaries. Each repetition included the same 423 patients; repeated predictions were not treated as independent patients (Supplementary Table 10). BKA, below-knee amputation.

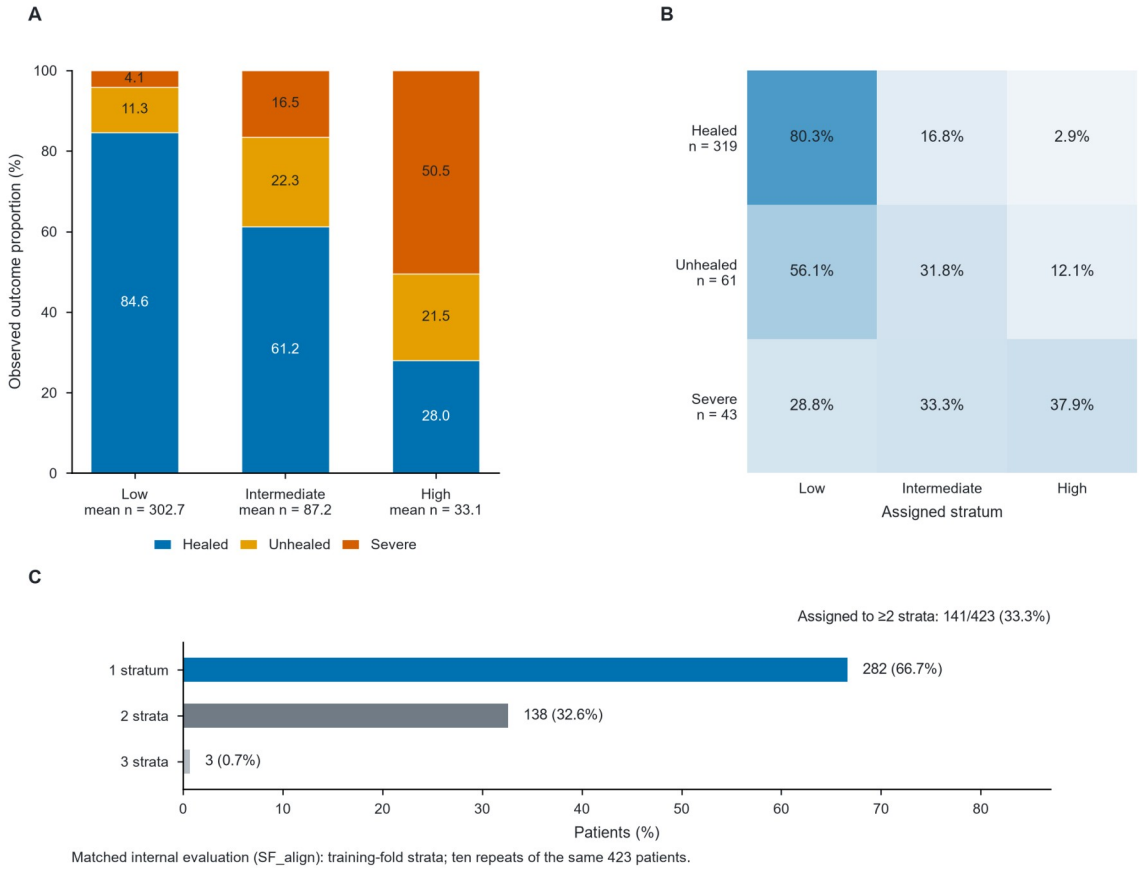
Fig. 3 Three-centre external validation of the locked score. (A) Severe-outcome AUCs in the same patients, with centre-stratified patient-bootstrap 95% CIs; parentheses show patient/event counts. (B) Calibration of locked IBG severe-risk predictions: mean predicted and observed proportions in five groups, with Wilson 95% CIs and group sizes. (C) Paired IBG-minus-four-predictor differences in severe-outcome and multiclass Brier scores among the same 205 patients, with 95% CIs; negative values favour IBG. (D) Mean predicted and observed severe risks by centre among patients with complete IBG inputs, with exact binomial 95% CIs for observed proportions. SX, XM, and HN denote Shanxi, Xiamen, and Hunan. AUC was not estimable in Hunan, where none of 38 patients had a severe outcome. AUC, area under the receiver operating characteristic curve; CI, confidence interval; IBG, infection–inflammation burden grade.

576 **Figure 1**



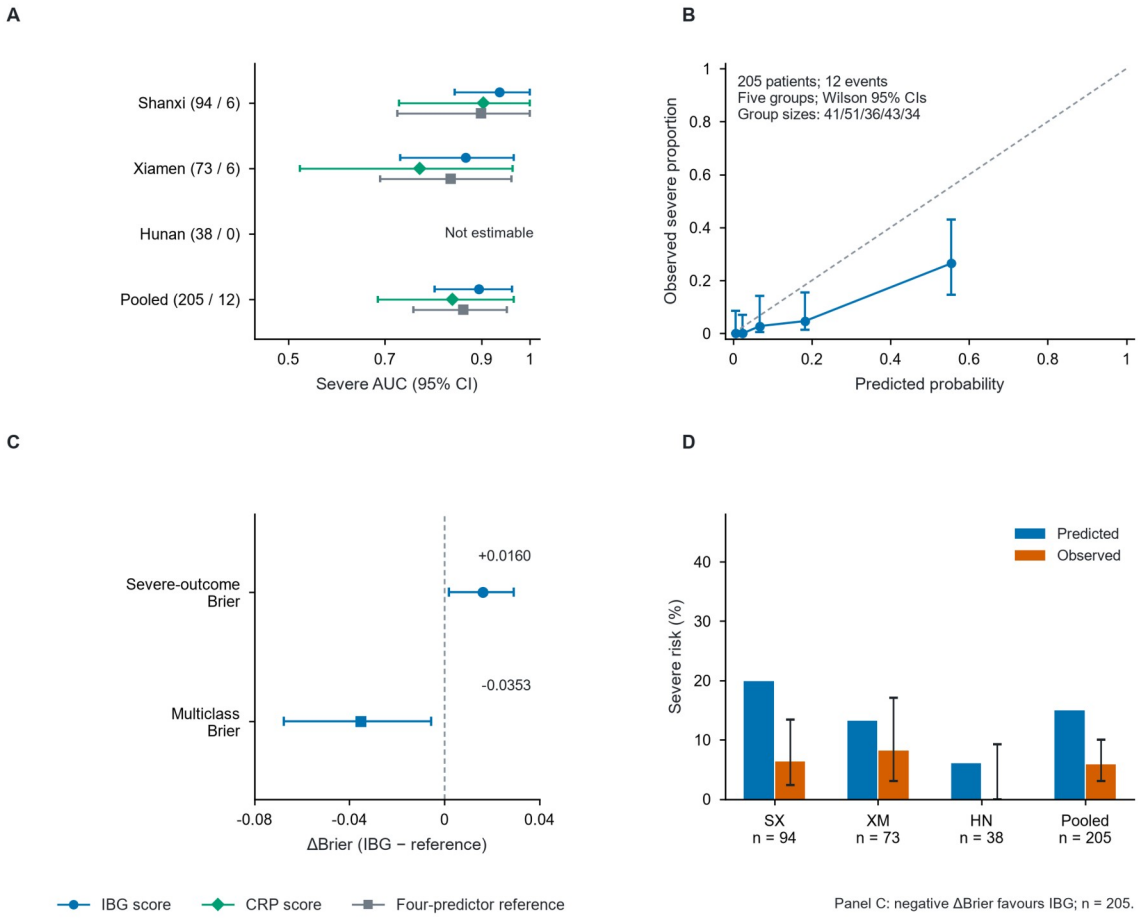
577

Figure 2



Matched internal evaluation (SF_align): training-fold strata; ten repeats of the same 423 patients.

580 **Figure 3**



581

Table 1 Preoperative characteristics and 180-day outcomes in the development cohort and three external centres

Characteristic	Development (n=423)	Shanxi (n=94)	Xiamen (n=81)	Hunan (n=38)
Age, years, median [IQR]	64 [57, 71.5]	68 [64, 72]	62 [55, 70]	67 [59.25, 72]
Male sex, n/N (%)	318/423 (75.2%)	61/94 (64.9%)	59/81 (72.8%)	25/38 (65.8%)
Chronic or acute heart failure	95/405 (23.5%)	10/94 (10.6%)	10/81 (12.3%)	1/38 (2.6%)
Acute heart failure	28/405 (6.9%)	0/94 (0.0%)	5/81 (6.2%)	0/38 (0.0%)
Renal function: grades 1–2	194/423 (45.9%)	65/94 (69.1%)	48/81 (59.3%)	24/38 (63.2%)
Grades 3–4	182/423 (43.0%)	19/94 (20.2%)	20/81 (24.7%)	7/38 (18.4%)
Grade 5	47/423 (11.1%)	10/94 (10.6%)	13/81 (16.0%)	7/38 (18.4%)
Wound area: <5 cm²	65/423 (15.4%)	9/94 (9.6%)	14/81 (17.3%)	7/38 (18.4%)
5–20 cm²	128/423 (30.3%)	28/94 (29.8%)	22/81 (27.2%)	15/38 (39.5%)
20<A≤100 cm²	135/423 (31.9%)	29/94 (30.9%)	22/81 (27.2%)	16/38 (42.1%)
>100 cm²	95/423 (22.5%)	28/94 (29.8%)	23/81 (28.4%)	0/38 (0.0%)
Palpated foot temperature: cold or cool	132/423 (31.2%)	78/94 (83.0%)	31/81 (38.3%)	18/38 (47.4%)
Warm	166/423 (39.2%)	16/94 (17.0%)	19/81 (23.5%)	9/38 (23.7%)
Warmer	125/423 (29.6%)	0/94 (0.0%)	31/81 (38.3%)	11/38 (28.9%)
IBG: grade 0	88/307 (28.7%)	46/94 (48.9%)	19/74 (25.7%)	19/38 (50.0%)
Grade 1	107/307 (34.9%)	27/94 (28.7%)	22/74 (29.7%)	13/38 (34.2%)
Grades 2–3	112/307 (36.5%)	21/94 (22.3%)	33/74 (44.6%)	6/38 (15.8%)
Wagner: grade 3	93/423 (22.0%)	65/94 (69.1%)	27/81 (33.3%)	2/38 (5.3%)
Grade 4	296/423 (70.0%)	26/94 (27.7%)	46/81 (56.8%)	36/38 (94.7%)
Grade 5	34/423 (8.0%)	3/94 (3.2%)	8/81 (9.9%)	0/38 (0.0%)
180-day outcome: healed	319/423 (75.4%)	62/94 (66.0%)	63/81 (77.8%)	38/38 (100.0%)
Persistently unhealed	61/423 (14.4%)	26/94 (27.7%)	12/81 (14.8%)	0/38 (0.0%)
Below-knee amputation or death	43/423 (10.2%)	6/94 (6.4%)	6/81 (7.4%)	0/38 (0.0%)
Below-knee amputation	19/423 (4.5%)	5/94 (5.3%)	1/81 (1.2%)	0/38 (0.0%)
All-cause death	28/423 (6.6%)	1/94 (1.1%)	5/81 (6.2%)	0/38 (0.0%)
Complete IBG score inputs	301/423 (71.2%)	94/94 (100.0%)	73/81 (90.1%)	38/38 (100.0%)
Complete CRP score inputs	396/423 (93.6%)	94/94 (100.0%)	79/81 (97.5%)	38/38 (100.0%)

Data are observed values before imputation, presented as median [IQR] for age and n/N (%) otherwise; N denotes non-missing observations for that variable. IBG availability alone differs from completeness of all score inputs. A score is available only when all inputs required by that version are present. Four development patients underwent amputation and subsequently died; each counted once in the severe composite. External deaths and below-knee amputations did not overlap. Clinical definitions are provided in Methods and Supplementary Table 19.

Table 2 The 0–59-point card derived from the six-predictor framework

Predictor	Category	Points
Age	<65 years	0
	≥65 years	3
Heart failure	Grade 0	0
	Grade 1	3
	Grade 2	16
Palpated foot temperature	Grade 1 (cold)	16
	Grade 2 (cool)	13
	Grade 3 (warm)	8
	Grade 4 (warmer)	0
eGFR-based renal function	Grades 1–2	0
	Grades 3–4	1
	Grade 5	8
Wound area	<5 cm²	0
	5–20 cm²	3
	20<A≤100 cm²	9
	>100 cm²	10
Infection–inflammation burden grade (IBG)	Grade 0	0
	Grade 1	3
	Grades 2–3	6

Heart failure grades 0/1/2 denote none/chronic/acute heart failure. Palpated foot-temperature grades 1–4 denote cold/cool/warm/warmer. Renal function comprises five grades based on standardised eGFR and documented uraemia; wound area comprises four categories. Definitions are provided in Supplementary Table 19 and IBG rules in Supplementary Table 3. The card was derived from one full-cohort imputation (M=1). Internal evaluation re-estimated points and probability maps within training folds and pooled predictions across 20 imputations.

Table 3 External review scenarios and event identification at prespecified severe-risk thresholds

Model	Risk threshold	Patients flagged for review/205	Events identified/12	Events missed	Sensitivity, % (95% CI)	PPV, % (95% CI)	Net benefit/100 (95% CI)
IBG score	5%	110	12	0	100.0 (73.5–100.0)	10.9 (5.8–18.3)	3.34 (0.28–7.14)
IBG score	10%	77	11	1	91.7 (61.5–99.8)	14.3 (7.4–24.1)	1.79 (-1.14–5.42)
IBG score	15%	60	11	1	91.7 (61.5–99.8)	18.3 (9.5–30.4)	1.15 (-1.92–4.85)
IBG score	20%	51	11	1	91.7 (61.5–99.8)	21.6 (11.3–35.3)	0.49 (-2.80–4.39)
Four-predictor reference	5%	95	12	0	100.0 (73.5–100.0)	12.6 (6.7–21.0)	3.72 (0.67–7.50)
Four-predictor reference	10%	82	9	3	75.0 (42.8–94.5)	11.0 (5.1–19.8)	0.43 (-2.22–3.85)
Four-predictor reference	15%	32	8	4	66.7 (34.9–90.1)	25.0 (11.5–43.4)	1.84 (-0.69–4.85)
Four-predictor reference	20%	18	7	5	58.3 (27.7–84.8)	38.9 (17.3–64.3)	2.07 (-0.24–5.00)

All comparisons use the same 205 patients, including 12 severe outcomes. Sensitivity is identified events/12; positive predictive value (PPV) is identified events/patients flagged. Both use Clopper–Pearson exact binomial 95% CIs. Net-benefit intervals use patient bootstrap resampling within centres. Counts represent potential review volume at four prespecified thresholds. Net benefit weights true- and false-positive classifications by the threshold and does not estimate events prevented.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryInformation.pdf