

Using Big Data to Improve Bacterial Sepsis Risk Stratification Among Immunocompromised
Cancer Patients

Margaret L Lind

A dissertation

submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2021

Reading Committee:

Amanda Phipps, Chair

Stephen Mooney

Steven Pergam

Program Authorized to Offer Degree:

Department of Epidemiology

©Copyright 2021
Margaret L Lind

University of Washington

Abstract

Using Big Data to Improve Bacterial Sepsis Risk Stratification Among Immunocompromised Cancer Patients

Margaret L Lind

Chair of the Supervisory Committee:

Amanda Phipps

Department of Epidemiology

Allogeneic hematopoietic stem cell transplant (HCT) recipients are an immunocompromised population that is disproportionately affected by sepsis, a life-threatening dysregulated immunologic response to an infection. While it is well established that early detection and treatment of sepsis with fluids and broad-spectrum antibiotics reduce the risk of mortality, recent data suggests early broad-spectrum antibiotic use in allogeneic HCT recipients may have microbiota-mediated detrimental effects on morbidity and mortality. Because of the risks associated with both missed and falsely identified sepsis events among allogeneic HCT recipients, early and accurate sepsis diagnosis is crucial. However, sepsis is generally challenging to diagnose and is made more complicated in allogeneic HCT recipients by the fact that sepsis presents differently following transplantation and common complications of the transplant procedure present like sepsis. In previous work, we demonstrated that current sepsis clinical prognostic tools have limited predictive validity among allogeneic HCT recipients and concluded that population specific prediction tools are needed. In the following papers, we outline the development and validation of two automated population-specific prognostic tools for

bacterial sepsis (a full predictor and a clinical-factor specific tool) and two bedside tools (a tree and a score) specifically for use in data or computationally limited scenarios (ie. outside of intensive care units - ICUs). Each tool was developed to provide bacterial sepsis risk stratification at a crucial time in patient management (blood culture collection) in both inpatient and outpatient settings. Our performance-based validation suggests that our full predictor automated tool (super HCT bacterial sepsis learner - SHBSL) had superior predictive validity among allogeneic HCT recipients and that, compared with existing tools, our bedside tools (HCT Bacteremia Sepsis Score - HBS2 and HCT Bacteremia Sepsis Tree - HBST) had superior predictive validity among patients with blood cultures collected outside of ICUs. Through our novel health outcomes-based tool validation, we showed that SHBSL led to meaningfully fewer deaths and little additional antibiotic use compared with existing tools among a simulated cohort of 2000 allogeneic HCT recipients followed for 100 days post transplantation and that HBST and HBS2 provided the best alternative to SHBSL in scenarios for which the use of an automated tool may be improbable. Taken together, these findings support the use of SHBSL when possible and HBST or HBS2 in situations for which computationally intensive tools are impractical.

Dedication

*This dissertation is dedicated to my loving and supportive parents, Maureen Laflin and Douglas
Lind.*

Acknowledgments

First, I would like to thank my dissertation committee members: Amanda Phipps, Steven Pergam, Stephen Mooney, Marco Carone, and Benjamin Althouse. Steve and Amanda, I am especially grateful for your mentorship, guidance, compassion, and understanding. I am scared to think of what this project would have been without you.

Second, I want to thank Maria Paleologos for helping me keep organized, tracking down signatures, and in general helping me keep my head on straight. Additionally, I want to thank Ryan Basom, Chris Davis, Gary Schoch, and John Klaassen for helping collect and clean the data.

Third, I would like to thank my parents (Maureen and Douglas), my brother (Brendan), my aunt (Molly) and my partner (Pete). Each of you provided a much-needed moment of support, kindness, and love over the past four years and I am incredibly thankful to have you in my life.

Fourth, I want to thank Judy and David who graciously let me live with them and were always interested in hearing about my day or my opinion on a new book.

Five, I would like to thank my grant funders. The support and training I was received as a ITHS TL1 and NHBLI F31 Fellow far exceeded my expectation and I am very thankful for the protected research time the Fellowships allotted me.

Finally, I want to thank my friends for helping me get through this bear of a process. Between walks around Green Lake to Tuesday night Bingo, you have made these last few years active, fun, and spontaneous, thank you.

Table of Contents

Primary Manuscripts	9
Paper 1: Immunocompromised cancer patient specific rapid clinical tools for bacteremia with high risk for sepsis.....	9
Abstract	9
Introduction	11
Methods	12
Results	16
Discussion	18
Conclusion	23
References	24
Tables	26
Figure Legends	29
Figures	31
Paper 2: Development and Validation of a Machine Learning Model to Predict Bacterial Sepsis among Immunocompromised Stem Cell Transplant Recipients	36
Abstract	36
Introduction	38
Methods	39
Results	43
Discussion	45
Conclusion	48
References	49
Tables	52
Figure Legends	55
Figures	56
Paper 3: Worth the switch? Modeling the impact of a novel bacterial sepsis prediction tool on mortality among allogeneic HCT recipients	58
Abstract	58
Introduction	60
Materials and Methods	61
Results	69
Discussion	71
Conclusion	74
References	75
Tables	77
Figure Legends	82
Figures	85
Supplement Material.....	91
Supplemental Paper 1: Immunocompromised cancer patient specific rapid clinical tools for bacteremia with high risk for sepsis	91
Missing Data	91
Validation versus Modeling Dataset	94
Detailed Methods	95
Calibration Plots for Developed and Existing Prediction Tools	100
Sensitivity Analyses Overview	102
Sensitivity Analyses Results	107
Window of Time for Data Inclusion:	110
Alternative Examination Measurement:	111
Alternative Culture Collection Inclusion:	112
Surveillance Culture Exclusion:	113

High and Low Risk Bacteremia Evaluation:	114
Supplemental Paper 2: Development and Validation of a Machine Learning Model to Predict Bacterial Sepsis among Immunocompromised Stem Cell Transplant Recipients	115
Potential Predictors and Missingness	117
Imputation Approach	120
Detailed Description of Super Learner Library	121
Net Reclassification	123
Ensemble Selection Rational	134
Sensitivity Analyses	138
References	149
Supplemental Paper 3: Worth the switch? The impact of a novel bacterial sepsis prediction tool on mortality among allogeneic HCT recipients with potential infections.....	150
Model Validation	150
Modeling Assumptions	152
Observed Mortality Proportions.....	153
Sensitivity Analyses	159

Primary Manuscripts

Paper 1: Immunocompromised cancer patient specific rapid clinical tools for bacteremia with high risk for sepsis

Abstract

Background: Diagnosing sepsis remains challenging and existing clinical tools have limited predictive value among immunocompromised patients, including allogeneic hematopoietic transplant (HCT) recipients. With the goal of improving sepsis risk stratification at the time of blood culture collection, we developed two easily calculable bedside clinical tools (tree and score) that target high (sepsis) risk bacteremia for allogeneic HCT recipients with potential bloodstream infections (PBIs) identified outside of intensive care units (ICUs).

Methods: Using data-driven approaches and bedside assessable predictors, we developed our tools among all non-ICU PBIs from the first 100 days following adult allogeneic HCT recipients from June 2010-2019. We compared our tool's predictive validity in relation to high-risk bacteremia against existing tools [Systemic Inflammatory Response Syndrome (SIRS), quick Sequential Organ Failure Assessment (qSOFA), and National Early Warning Score (NEWS)] using a held-out validation dataset.

Findings: Between June 2010-2019, 1563 recipients contributed 7253 non-ICU PBIs and 174 high-risk bacteremia events. Our score contains one temperature, heart rate, respiratory rate, and diastolic blood pressure criteria and our tree contains one respiratory rate, two temperature, and two age decision points. The score and tree had AUCs of 0.78 (95% Confidence Interval (CI): 0.72-0.84) and 0.75 (CI: 0.68-0.81), respectively. SIRS had the highest AUC of existing tools – 0.65 (CI: 0.58-0.73). At upper-left method selected optimal cut-points, the score and tree had sensitivities of 72.4% (CI: 58.4-83.1%) and 62.1% (CI: 46.6-75.4%), respectively, and specificities of 70.6% (CI: 67.9-73.2%) and 78.3% (CI: 75.1-81.1%), respectively. SIRS 2+ had

a sensitivity of 51.7% (CI: 38.1-65.1%) and specificity of 76.2 (CI: 74.0-78.3%).

Conclusion: Our population-specific score had superior predictive validity in relation to bacteremia with high-risk for sepsis among allogeneic HCT recipients. If implemented, our score may lead to fewer missed sepsis events relative to the same number of falsely identified events than existing clinical tools.

Introduction

Early diagnosis of sepsis remains challenging, especially among immunocompromised patients, such as allogeneic hematopoietic stem cell transplant (HCT) recipients.¹ In the absence of a gold-standard diagnostic test, clinical scoring tools have been developed to help identify patients at high risk of sepsis among those already suspected of having an infection. Available scores were developed using expert opinion of common symptomology and/or data from general population of hospitalized or at-risk adult patients and were built to estimate risk in relation to downstream sepsis endpoints (e.g. short-term mortality and intensive care unit [ICU] admission).²

Unfortunately, following transplantation, sepsis often presents atypically, with common transplant-related side effects, such as engraftment syndrome and anemia, presenting with sepsis-like symptoms.^{3,4} It is therefore unsurprising that existing tools have limited predictive validity among transplant patients, and recognized limitations may be exacerbated among allogeneic HCT recipients.⁵ Moreover, the endpoints used to estimate sepsis risk in existing tools (short-term mortality and/or ICU admission) suffer from a major conceptual misclassification – they are impacted by both infection and symptom-based treatment that are received by the vast majority of allogeneic HCT recipients at the time of or following blood culture collection. It follows that tools built against these endpoints may not properly identify patients who would have died or been admitted to the ICU *in the absence of treatment*, which is ultimately the clinical goal of a risk stratification tool.

This conceptual limitation can be avoided by developing tools that predict risk of an upstream target of sepsis treatment. Current sepsis treatments, though ultimately aiming to prevent mortality and ICU admission, directly target symptom presentation (e.g. intravenous fluids for hypovolemia) or the underlying infection (empiric, broad spectrum antibiotics).⁶ As the

primary causal target in sepsis treatment bundles, the underlying infection provides a prediction outcome that is less prone to the biases highlighted above and leads to actionable treatment information.

Updating bedside tools to reflect the unique sepsis presentation experience by this population are needed. In order to address our aim of rapid risk stratification and antibiotic treatment decision support at the time of blood culture collection, we developed two HCT recipient specific bedside clinical tools (score - HCT Bacteremia Sepsis Score HBS2 and decision tree - HCT Bacteremia Sepsis Tree HBST) for bacteremia with high risk for sepsis. Our tools, similar to quick sequential organ failure assessment (qSOFA), were built to be used in data and/or computationally limited scenarios (i.e. outside of the ICU), are easy to calculate manually, and include only factors easily assessed from the bedside. To test the performance of our developed clinical tools among allogeneic HCT recipients with potential bloodstream infections (PBIs), we compared their predictive validity to those of available clinical scores: qSOFA, Systemic Inflammatory Response Syndrome (SIRS), and National Early Warning Score (NEWS).

Methods

Population and Cohort Development:

We retrospectively identified all adult (> 18 years) allogeneic HCT recipients who received their first transplant at the Seattle Cancer Care Alliance/Fred Hutchinson Cancer Research Center (SCCA/FHCRC) and agreed to research participation between June 2010 and June 2019 and experienced at least one PBI outside of the ICU. For each included patient, we examined data from the 100 days immediately following their first transplantation and included all PBIs that occurred outside of the ICU in our cohort. We defined PBI as a blood culture that occurred in the absence of recent (last three days) administration or prescription of an antibiotic other than the

first-line neutropenic fever prophylaxis (levofloxacin). To reduce the chance of capturing the same PBI twice, we combined the results of positive cultures collected within one hour of each other and excluded follow-up cultures, or continuously collected cultures (within 36 hours of prior culture) for the 14 days following a positive result.⁵ The study was approved by the FHCRC Institutional Review Board and follows the TRIPOD reporting guidelines.

Outcome and Predictor Definitions:

With the goal of developing practical risk stratification tools for allogeneic HCT recipients at the time of blood culture collection, we built our tools to predict common bloodstream bacterial causes of sepsis (gram-negative species, *Staphylococcus aureus*, or *Streptococcus* species bacteremia – referred to as high-risk bacteremia)⁷ and limited our list of potential predictors to factors easily obtained from the bedside. Specifically, we included clinical examination factors [systolic blood pressure (SBP), diastolic blood pressure (DBP), respiratory rate, heart rate, temperature, oxygen saturation (SpO₂), supplemental oxygen use, Glasgow Coma Scale score (GCS)], body mass index (BMI), and demographic factors (age and sex) as potential predictors. White blood cell count (WBC) was used to calculate SIRS but, as WBC is not calculable at the bedside, it was not included in model development.

Data Collection and Cleaning:

Data from the 24 hours prior to and the 2 hours following PBI onset (culture collection) were queried from a transplant database prospectively maintained by the FHCRC and directly abstracted from electronic medical records (EMR). Each predictor was standardized to a uniform measure (i.e., temperature in Celsius), and greater than or lesser than values were converted to the first included number of the statement (e.g. if $\leq 34^{\circ}\text{C}$, 34°C was used). Non-oral temperatures were standardized to oral temperatures by subtracting 0.2°C from arterial and adding 0.45°C to axillary temperatures.⁸ Binary forms of each predictor were generated for

HBS2 (see *Supplement: Detailed Methods*).

We addressed the large amount of structurally ambiguous missingness in our retrospective data by removing predictors with more than 35% missingness during the measurement window (24 hours prior and 2 hours following culture collection; e.Table1). We accounted for remaining missingness using a single, normal value imputation approach (See *Supplement: Missing Data*, e.Figure 1/e.Table 2).^{2,5}

From the cleaned and imputed dataset, we selected the factor measurement recorded closest in time to culture collection for analytic use and randomly assigned recipients to either the modeling or validation group at a ratio of 7:3. This split provided us with a large enough validation sample to reliably estimate tool performance and modeling dataset to perform our model selection process. Each patient's PBIs were then added to a modeling or validation dataset based on the patient's grouping.

Culture Collection Guidelines:

Center guidelines recommended daily culture collection among patients with fevers and surveillance cultures drawn bi-weekly while inpatient and weekly while outpatient for patients treated with high-dose glucocorticoids.⁹ However, treating teams had final say over all culture collection. A detailed description of the center's guidelines during the study period are published elsewhere.⁵

Score and Decision Tree Development:

In the absence of a clear clinical presentation of high-risk bacteremia among our population at the time of culture collection and the desire to develop practical clinical tools, we developed our tools using data-driven predictor selection approaches (decision tree: best subset, score: forward stepwise selection).¹⁰ Briefly, the optimal model for each tool was selected based on an assessment of its cross-validated area under the curve (cv-AUC) relative to its flexibility (HBST:

number of terminal nodes; HBS2: number of predictors). HBST was modeled using pruned classification trees and the number of included predictors and cost complexity factor (CCF) used in tree pruning dictated model flexibility. HBS2 was modeled using logistic regression and model flexibility was determined by the number of included binary form predictors (e.Figure 2). The criterion scores for HBS2 were calculated using a beta integer weighting approach.¹¹ A detailed description of our model development methods can be found in the *Supplement: Detailed Methods*.

Tool Evaluation (Calibration and Discrimination):

Using the validation dataset, we compared the predictive validity of HBST, HBS2 and three existing clinical scores (SIRS, qSOFA, and NEWS) in relation to our defined high-risk bacteremia endpoint among all identified PBIs (primary analysis) and among PBIs stratified by patient location (inpatient versus outpatient) at the time of culture collection (secondary analysis). We assessed model calibration by comparing predicted versus observed outcome probabilities and calibration plots.^{12,13} Model discrimination was evaluated using AUCs and accuracy metrics (sensitivity, specificity and negative and positive predictive value) for qSOFA 2+, SIRS 2+, NEWS 4+ and 7+; for our tools, metrics were estimated for each cut-point (primary analysis) and for Upper-Left method selected optimal cut-points (secondary, sensitivity analyses). We accounted for potential within-person correlation in the accuracy metric uncertainty estimates using generalized estimating equations with robust standard errors. Predictive value and clinical usefulness were visually assessed using a stepwise Receiver Operating Characteristic (ROC) and decision curve, respectively. All analyses were performed using R 3.6.2 and statistical packages rms, SuperLearner, fbrc, geepack, and OptimalCutoffs.^{14–19}

Sensitivity Analysis:

We tested our evaluation results' robustness to varying missing data structures, data collection periods, measurement selections, and analyzed cultures. Additionally, we estimated the ability of tools to discriminate between high-risk and low-risk (*defined as Coagulase-Negative Staphylococci*) bacteremia events. All sensitivity analyses were conducted in the validation dataset. A detailed description of these analyses is presented in *Supplement: Sensitivity Analyses*.

Results

Study Population and Cohort:

Between June 2010 and June 2019, a total of 1943 individuals received their first allogeneic HCT at the SCCA/FHCRC and agreed to participate in research. Among them, 1563 (80.4%) experienced at least one PBI outside of the ICU and contributed a total of 7253 non-ICU PBIs to our cohort (Figure 1). The median age was 54 (Interquartile Range [IQR]: 42 - 63) years and most of the population was non-Hispanic white (82.4%). Patients who experienced at least one high-risk bacteremia event were demographically similar to and had similar transplant characteristics as those who did not experience an event. However, patients who experienced a high-risk bacteremia event were more likely to have died during follow-up (13.4%) than those who did not experience an event (7.8%, Table 1).

Criteria and Decision Tree Development:

GCS and SpO2 were dropped due to excess missingness (> 35% - e.Table 1). Of the 7253 examined PBIs, 174 (2.4%) were defined as high-risk bacteremia (e.Table 2). Following individual-level modeling/validation randomization, the modeling and validation datasets contained 5061 and 2192 PBIs, respectively, and 116 (2.3%) and 58 (2.6%) high-risk bacteremia events, respectively. Patients randomized to the modeling dataset were demographically and clinically similar to individuals randomized to the validation dataset (e.Table 3). The majority of PBIs occurred while patients were in outpatient settings (77.0%) but a smaller percent of

outpatient PBIs were defined as high-risk bacteremia (outpatient 1.0%, inpatient 7.2%).

The optimal decision tree had a cost complexity pruning of 0.02 and predictor count of three (e.Figure 3). The best performing (highest cv-AUC) model with these parameters contained temperature, respiratory rate, and age at culture collection. The resulting tree (HBST) is comprised of 5 decision points and 6 terminal nodes. We identified the optimal score as one containing four binary predictors (Figure 2) and the best performing score of that size contained one temperature, heart rate, respiratory rate, and DBP criteria (Figure 2). An integer weight of one was found to be the lowest integral weight that led to a similar cv-AUC as unconstrained beta results and used in the final score development (e.Figure 4). The final score (HBS2) has 4 criteria and 8 cut-points ranging in value from 0-7 (Figure 2).

Tool Evaluation (Calibration and Discrimination):

The probability of high-risk bacteremia was well predicted by each (developed or existing) tool (e.Table 4). The calibration plots (e.Figure 5) demonstrate that all tools had moderate calibration and that HBS2 over-estimated the probability of high-risk (observed probability > 13%) individuals. Our developed tools had moderate to high AUC in relation to high-risk bacteremia among all patients and among patients stratified by location of PBIs: All: HBS2: 0.78 (95% Confidence Interval (CI): 0.72-0.84), HBST: 0.75 (CI: 0.68-0.81); Inpatient: HBS2: 0.70 (CI: 0.61-0.78), HBST: 0.66 (CI: 0.57-0.75); Outpatient: HBS2: 0.73 (CI: 0.63-0.83), HBST: 0.70 (CI: 0.60-0.81; Table 2, Figure 3). Of the examined existing tools, SIRS had the highest AUC among all PBIs (0.65, CI: 0.58-0.73) and qSOFA had the lowest (0.51, CIs: 0.46-0.57; Table 2, Figure 3). Among PBIs identified in inpatient settings, NEWS had the highest AUC of existing tools (0.64) and among PBIs identified in outpatient settings, SIRS had the highest AUC of existing tools (0.53; Table 2, Figure 3).

At their Upper-Left Method selected cut-point, HBS2 (2+) and HBST (3.3%) had

moderate sensitivity (HBS2: 72.4%, HBST: 62.1%) and specificity (HBS2: 70.6%, HBST: 78.3%, Table 2, Figure 3). At these cut-points, HBS2 missed seven (23%) of the gram-negative, zero of the *Staphylococcus aureus*, and nine (33%) of the *Streptococcus species* bacteremia events. SIRS 2+, the tool with the highest sensitivity, missed 15 (50%) gram-negative, zero *Staphylococcus aureus*, and 13 (48%) *Streptococcus species* bacteremia events (e.Table 5).

All the evaluated prediction tools had low positive predictive value (PPV) and high (> 93.0%) negative predictive value (NPV). Among existing tools, qSOFA (2+) had the highest PPV (22.2%, CI: 5.6-57.9%) and NEWS (4+) had the lowest (3.9%, CI: 2.5-6.2%). SIRS (2+), the only existing tool with a sensitivity above 50%, had a PPV of 5.6% (3.9-7.9%). At cut-points with similar sensitivity as SIRS 2+ (51.7%; CI: 38.1-65.1%), HBS2 and HBST had PPVs of 11.5% (CI: 8.1-16.0%) and 9.1% (CI: 6.6-12.3%, Table 2), respectively.

Sensitivity Analyses:

The predictive validity of each tool (according to AUC) varied under the examined scenarios.

Among non-surveillance cultures, HBS2 had an AUC five units higher than that of SIRS, the existing tool with the highest AUC and at a cut-point with the closest specificity to that of SIRS 2+ (SIRS: 54%, HBS2: 59%), HSB2 had a sensitivity seven percentage points higher than that of SIRS 2+ (e.Table 11). When PBIs were limited to high and low-risk bacteremia events, HBS2 and HSBT had AUCs 10 and seven units higher than that of NEWS (e.Table12). In general, the AUCs of HBS2 and HBST estimated in the sensitivity analyses were lower than the AUCs estimated in the primary analysis, however, they were still higher than for existing tools under all examined scenarios (e.Tables 6-13, e.Figure 6).

Discussion

We developed two population-specific, rapid clinical tools (a decision tree-HBST and score-HBS2) for high-risk bacteremia among allogeneic HCT recipients with PBIs identified outside of

the ICU. Our tools were built to provide immediate risk stratification and treatment decision support at the time of culture collection in data limited and/or computationally limited settings and are comprised of a limited number of predictors easily assessed at the bedside. In relation to high-risk bacteremia (gram-negative species, *Staphylococcus aureus*, and *Streptococcus* species bacteremia), our developed tools had moderate to high AUC (HBS2: 0.78; HBST: 0.75) and provided cut-points with moderate to high specificity (70s) and moderate sensitivity (60s to 70s).

We built our tools to provide information regarding the relative likelihood of high-risk bacteremia at a common antibiotic initiation decision point (culture collection). In order to evaluate if our tools resulted in more accurate treatment related information than existing clinical tools, we examined if our tools had overall higher predictive validity (higher AUC) and if they provided clinical cut-points with better balance between capture (sensitivity) and over-capture (1-specificity). When compared to existing bedside tools and our tree (HBST), our score (HBS2) had a higher AUC and more practical clinical cut-points (i.e., higher PPV for any given sensitivity). SIRS (2+), the existing criteria with the highest AUC, had an inadequately low sensitivity (51.7%) given the high risk of mortality among allogeneic HCT recipients with culture positive sepsis and a moderate PPV (5.6%).¹ While low PPVs were observed for all tools at all examined cut-points and were expected given the rarity of our outcome, our tools offer cut-points with similar PPV and specificity as existing tools but with meaningfully higher sensitivity.²⁰⁻²⁴ For example, HBS2 2+ had a PPV and specificity similar to that of SIRS 2+ (PPV: HBS2: 6%, SIRS: 6%; specificity: HBS2: 71%, SIRS: 76%) and a sensitivity 20 percentage points higher than that of SIRS (2+). This suggests that among 100 allogeneic HCT recipients with high-risk bacteremia and 100 without, HBS2 would capture approximately 20 more events than SIRS (2+) relative to an equal false positive rate and less than six additional

falsely identified events (Figure 4).

The superior nature of our tools, namely HBS2, held under all examined scenarios. Of particular interest was their performance when surveillance cultures were removed (e.Table12) and when discrimination between high and low-risk bacteremia²⁵ events were examined (e.Table13). Though our tools had lower validity under both scenarios than among the primary analysis, they still provided higher AUCs and HBS2 provided cut-points with better sensitivity to specificity ratios than existing tools. However, the optimal cut-point for HBS2 under these scenarios would be three instead of two. The fact that HBS2's superior validity held under these scenarios suggests that (1) if surveillance culture collection were to be discontinued or HBS2 were to be used only on non-surveillance cultures, it would still provide meaningfully more accurate risk estimates than existing tools and (2) HBS2's ability to discriminate between high and low-risk events includes an ability to select between high and low-risk bacteremia events.

The difference in predictive ability between SIRS and HBS2 may appear surprising given the similarities of these scores (both include temperature, heart rate, and respiratory rate). However, when we examine ridgeline density plots (Figure 5) by presence of high-risk bacteremia, we see that HBS2 is a better reflection of the presentation of high-risk bacteremia within this population than SIRS. For example, at a measurement of 90 heartbeats per minute (the cut-off in SIRS), the density is higher among patients without high-risk bacteremia than with high-risk bacteremia (Figure 5). Conversely, at 100 heartbeats per minute (the cut-off in HBS2), the density is higher among patients with high-risk bacteremia than without. This suggests that the heart rate cut-off of SIRS is overly sensitive and likely a reason for the tool's lower PPV relative to sensitivity compared with HBS2. The plots also show that, while a high white blood cell count is likely predictive of high-risk bacteremia among allogeneic HCT recipients with

PBIs, a white blood cell count cut-off of $< 4/\text{cm}^3$ (the low-end cut-off in SIRS) is too sensitive and, as with the overly inclusive heart rate cut-off, may partially explain the limited predictive ability of SIRS within this population (Figure 5).

We observed lower AUC, sensitivity, and PPV for all tools among PBIs identified in outpatient settings than we did among all PBIs. The lower observed predictive validity may be driven, in part, by the rarity of high-risk bacteremia among PBIs identified in outpatient settings ($\sim 1.0\%$). However, it may also reflect differences in underlying high-risk bacteremia presentations by patient location not specifically captured by our tools. Nevertheless, our tools had higher AUC and higher sensitivity for any given PPV than existing tools among location stratified PBIs. This suggests that, even if presentation varies by patient location, our tools better predict high-bacteremia risk among allogeneic HCT recipients (at the time of culture collection) in both inpatient and outpatient settings than existing tools.

Despite our tools having higher AUC and providing cut-points with higher sensitivity for any given PPV than existing tools, the predictive value of our tools does not compare to that of automated (machine learning) sepsis prediction tools.²⁶ However, this does not negate their potential importance or utility. Unlike automated tools, our tools were built to provide decision support with minimal computational requirements and in limited data scenarios, such as directly following hospital admission or within the emergency room. Because automated tools require integration into EMR systems, a level of computer literacy, and data pre-processing, their predictions under limited resource scenarios may be delayed or computationally impossible. Further, our tools provide population-specific decision support for clinicians without the requirement of additional computerized alerts and the risk of furthering clinical alert fatigue.²⁷

Strengths and Limitations:

Our study has several limitations common to retrospective analyses. First, though our population

is representative of our center's transplant population, it may not be representative of the larger US or international transplant population. Second, our tools were developed and evaluated among patients with blood cultures, the collection of which may vary by location. Third, our study relied on retrospectively collected data, which suffers from missingness and measurement error. These limitations may impact the generalizability of our evaluation results and, while we performed numerous sensitivity analyses trying to assess their impacts, external validation should be conducted prior to tool implementation. Furthermore, due to our desire to include only clinical examination factors as predictors, we did not examine laboratory factors, such as white blood cell count (which our ridgeline density plots suggest may be predictive) and/or lactate levels. Lactate collection following a positive tool event may provide additional predictive value and future analysis should examine the benefit of a two-tiered system. Finally, as with existing tools, our tools were built to predict a sepsis associated endpoint (not sepsis itself) among patients with potential/suspected infections. This means that their usefulness as sepsis risk stratification tools among all allogeneic HCT recipients (regardless of PBI status) and in relation to non-bacteremia associated sepsis events has not been tested and may be limited.

Our study has many strengths. First, our study includes a large population of eligible adult allogeneic HCT recipients who received their first allogeneic transplant over a 10-year period and were observed regularly for the first 100 days following transplantation. The large sample resulting from this follow-up (>7000 PBIs) allowed us to feel comfortable with the reliability of the predictive values we estimated in the validation dataset. Second, our tools were developed using data-driven methods that allowed for population specific categorization to occur and the unique presentation of high-risk bacteremia within this population to be captured. Third, we showed that our predictive tool evaluation was robust to numerous assumptions made

including those regarding missing data and culture inclusion. This confirmed that our data cleaning and modeling assumptions did not drive the superiority of our developed tools in comparison to existing clinical tools. Fourth, while package limitations prevented us from accounting for potential within-person correlation in our AUC uncertainty estimates, we were able to account for this correlation in our accuracy metric uncertainty estimates using generalized estimating equations. Finally, our tools were developed using easily assessed bedside examination factors which allows them to be of practical use in a variety of clinical settings.

Conclusion

We developed two rapid bedside clinical tools for high-risk bacteremia among allogeneic HCT recipients already considered high risk for an infection. Our tools were built to supply bacterial sepsis risk stratification and to support treatment decisions at the time of culture collection among recipients in data limited and/or computationally limited settings. Among PBI events identified outside of the ICU, our developed tools (specifically HBS2) had higher predictive validity than existing bedside tools. These findings support the use of HBS2 as an alternative sepsis consideration prompt for this high-risk population and suggest that, if used, HBS2 may lead to improved early detection of bacterial sepsis among allogeneic HCT recipients with PBIs.

References

1. Kumar G, Ahmad S, Taneja A, et al. Severe sepsis in hematopoietic stem cell transplant recipients*. *Crit. Care Med.* 2015;43(2):411–421.
2. Seymour CW, Liu VX, Iwashyna TJ, et al. Assessment of Clinical Criteria for Sepsis: For the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA.* 2016;315(8):762–774.
3. Dadwal S, Eftekharai Z, Tushondra T, Johnson D. A Machine-Learning Based Prediction Model for Sepsis in Patients Undergoing Hematopoietic Stem Cell Transplantation. *ASH Annual Meeting.* 2018;the Systematic Inflammatory Response Syndrome (SIRS) and quick-Sequential Organ Failure Assessment (qSOFA) criteria:
4. Kochanek M, Schalk E, von Bergwelt-Baildon M, et al. Management of sepsis in neutropenic cancer patients: 2018 guidelines from the Infectious Diseases Working Party (AGIHO) and Intensive Care Working Party (iCHOP) of the German Society of Hematology and Medical Oncology (DGHO). *Ann Hematol.* 2019;98(5):1051–1069.
5. Lind ML, Phipps AI, Mooney S, et al. Predictive value of three clinical criteria for sepsis (qSOFA, SIRS, and NEWS) with respect to short-term mortality in allogeneic hematopoietic cell transplant recipients with suspected infections. *Clin. Infect. Dis.* 2020;
6. Dellinger RP, Levy MM, Carlet JM, et al. Surviving Sepsis Campaign: International guidelines for management of severe sepsis and septic shock: 2008. *Intensive Care Med.* 2008;34(1):17–60.
7. Hämäläinen S, Kuittinen T, Matinlahti I, et al. Severe sepsis in autologous stem cell transplant recipients: Microbiological aetiology, risk factors and outcome. *Scandinavian Journal of Infectious Diseases.* 2009;41(1):14–20.
8. Healthwise Staff. Fever Temperatures: Accuracy and Comparison. 2018;
9. Stohs E, Chow VA, Liu C, et al. Limited Utility of Outpatient Surveillance Blood Cultures in Hematopoietic Cell Transplant Recipients on High-Dose Steroids for Treatment of Acute Graft-versus-Host-Disease. *Biology of Blood and Marrow Transplantation.* 2019;25(6):1247–1252.
10. James G, Witten D, Hastie T, Tibshirani R. An Introduction to Statistical Learning: with Applications in R. New York: Springer; 2017.
11. Mehta HB, Mehta V, Girman CJ, Adhikari D, Johnson ML. Regression coefficient-based scoring system should be used to assign weights to the risk index. *J Clin Epidemiol.* 2016;79:22–28.
12. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for some traditional and novel measures. *Epidemiology.* 2010;21(1):128–138.
13. Cox DR. Two further applications of a model for binary regression. *Biometrika.* 1958;45(3–4):562–565.
14. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria: 2019.
15. López-Ratón M, Rodríguez-Álvarez MX, Cadarso-Suárez C, Gude-Sampedro F. OptimalCutpoints: An R Package for Selecting Optimal Cutpoints in Diagnostic Tests. *Journal of Statistical Software.* 2014;61(1):1–36.
16. Polley E, LeDell E, Kennedy C, van der Laan M. SuperLearner: Super Learner Prediction. 2018.

17. Jr FEH. rms: Regression Modeling Strategies. 2020.
18. Peter E. fbroc: Fast Algorithms to Bootstrap Receiver Operating Characteristics Curves. 2019.
19. Højsgaard S, Halekoh U, Yan J. The R Package geepack for Generalized Estimating Equations. *Journal of Statistical Software*. 2005;15(1):1–11.
20. Weber D, Jenq RR, Peled JU, et al. Microbiota Disruption Induced by Early Use of Broad-Spectrum Antibiotics Is an Independent Risk Factor of Outcome after Allogeneic Stem Cell Transplantation. *Biology of Blood and Marrow Transplantation*. 2017;23(5):845–852.
21. Golob JL, Pergam SA, Srinivasan S, et al. Stool Microbiota at Neutrophil Recovery Is Predictive for Severe Acute Graft vs Host Disease After Hematopoietic Cell Transplantation. *Clin. Infect. Dis*. 2017;65(12):1984–1991.
22. Holler E, Butzhammer P, Schmid K, et al. Metagenomic Analysis of the Stool Microbiome in Patients Receiving Allogeneic Stem Cell Transplantation: Loss of Diversity Is Associated with Use of Systemic Antibiotics and More Pronounced in Gastrointestinal Graft-versus-Host Disease. *Biology of Blood and Marrow Transplantation*. 2014;20(5):640–645.
23. Ogimi C, Krantz EM, Golob JL, et al. Antibiotic Exposure Prior to Respiratory Viral Infection Is Associated with Progression to Lower Respiratory Tract Disease in Allogeneic Hematopoietic Cell Transplant Recipients. *Biol. Blood Marrow Transplant*. 2018;
24. Galloway-Peña JR, Jenq RR, Shelburne SA. Can Consideration of the Microbiome Improve Antimicrobial Utilization and Treatment Outcomes in the Oncology Patient? *Clin Cancer Res*. 2017;23(13):3263–3268.
25. Al-Mazroea AH. Incidence and Clinical Significance of Coagulase Negative Staphylococci in Blood. *Journal of Taibah University Medical Sciences*. 2009;4(2):137–147.
26. Fleuren LM, Klausch TLT, Zwager CL, et al. Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. *Intensive Care Med*. 2020;46(3):383–400.
27. Co Z, Holmgren AJ, Classen DC, et al. The tradeoffs between safety and alert fatigue: Data from a national evaluation of hospital medication-related clinical decision support. *J Am Med Inform Assoc*. 2020;27(8):1252–1258.

Tables

Table 1: Baseline Characteristics by Occurrence of Sepsis Prediction Tool Event

	Full Population	Developed High-Risk Bacteremia ^a	Did Not Developed High-Risk Bacteremia ^a
	N = 1563	N = 157	N = 1406
Calendar Year at Transplant (Median, IQR)	2014 (2012, 2016)	2013 (2011, 2016)	2014 (2012, 2016)
Age at Transplant (Median, IQR)	54 (43, 63)	53 (41, 61)	55 (43, 63)
Gender (Count, % Male)	889 (56.9)	88 (56.1)	801 (57.0)
Race (Count, %)			
American Indian or Alaskan Native	26 (1.7)	2 (1.3)	24 (1.7)
Asian	116 (7.4)	13 (8.3)	103 (7.3)
Black	27 (1.7)	5 (3.2)	22 (1.6)
Multiple	23 (1.5)	3 (1.9)	20 (1.4)
Native Hawaiian or other Pacific Islander	25 (1.6)	5 (3.2)	20 (1.4)
Unknown	58 (3.7)	5 (3.2)	53 (3.8)
White	1288 (82.4)	124 (79.0)	1164 (82.8)
HCT Cell Donor Type			
Bone marrow	124 (7.9)	20 (12.7)	104 (7.4)
Bone marrow, peripheral blood stem cells	3 (0.2)	0 (0.0)	3 (0.2)
Umbilical cord blood	174 (11.1)	25 (15.9)	149 (10.6)
Peripheral blood stem cells	1262 (80.7)	112 (71.3)	1150 (81.8)
Donor Relation (Count, %)			
Child	36 (2.3)	6 (3.8)	30 (2.1)
Half Sibling	5 (0.3)	0 (0.0)	5 (0.4)
Not Related	1085 (69.4)	107 (68.2)	978 (69.6)
Parent	17 (1.1)	4 (2.5)	13 (0.9)
Relative	1 (0.1)	0 (0.0)	1 (0.1)
Full Sibling	419 (26.8)	40 (25.5)	379 (27.0)
Acute GVHD (Grade > 1) (Count, %) ^a	1166 (74.6)	123 (78.3)	1043 (74.2)
Died (Count, %) ^a	130 (8.3)	21 (13.4)	109 (7.8)

^a High-risk bacteremia defined as any gram-negative bacteremia, *Staphylococcus aureus* or *Streptococcal* bacteremia

^b Experiences during follow-up

Table 2: Discrimination - Predictive Accuracy of Sepsis Criteria in Relation to High-risk Bacteremia^a

	AUC ^b	Sensitivity ^c	Specificity ^c	Positive Predictive Value ^c	Negative Predictive Value ^c
All Potential Bloodstream Infections					
HBS2 - Bedside Score	0.78 (0.72, 0.84)				
Cut-point 0	-	100 (100, 100%)	0.0 (0.0, 0.2%) ^e	-	-
Cut-point 1	-	93.1 (82.8, 97.4%)	33.9 (31.3, 36.6%)	3.7 (2.7, 5.0%)	99.5 (98.6, 99.8%)
Cut-point 2	-	72.4 (58.4, 83.1%)	70.6 (67.9, 73.2%)	6.3 (4.6, 8.5%)	98.9 (98.1, 99.4%)
Cut-point 3	-	63.8 (47.0, 77.8%)	82.1 (80.0, 84.0%)	8.8 (6.4, 12.0%)	98.8 (97.9, 99.3%)
Cut-point 4	-	48.3 (34.4, 62.4%)	89.9 (88.1, 91.4%)	11.5 (8.1, 16.0%)	98.5 (97.6, 99.0%)
Cut-point 5	-	31.0 (20.1, 44.6%)	94.7 (93.5, 95.7%)	13.7 (9.0, 20.4%)	98.1 (97.2, 98.7%)
Cut-point 6	-	6.9 (2.6, 17.2%)	98.2 (97.5, 98.7%)	9.5 (3.8, 21.7%)	97.5 (96.6, 98.2%)
Cut-point 7	-	1.7 (0.2, 11.3%)	99.2 (98.6, 99.5%)	5.3 (0.7, 29.5%)	97.4 (96.5, 98.1%)
HBST - Bedside Decision Tree	0.75 (0.68, 0.81)				
Cut-point 0.83%	-	100 (100, 100%)	0.0 (0.0, 0.2%) ^e	-	-
Cut-point 0.95%	-	94.8 (85.3, 98.3%)	23.4 (19.1, 28.3%)	3.3 (2.4, 4.4%)	99.4 (98.2, 99.8%)
Cut-point 3.3%	-	62.1 (46.6, 75.4%)	78.3 (75.1, 81.1%)	7.2 (5.2, 9.8%)	98.7 (97.8, 99.3%)
Cut-point 3.4%	-	58.6 (43.6, 72.2%)	84.1 (82.2, 85.8%)	9.1 (6.6, 12.3%)	98.7 (97.8, 99.2%)
Cut-point 3.7%	-	39.7 (27.2, 53.6%)	90.7 (89.2, 92.0%)	10.4 (7.1, 15.0%)	98.2 (97.3, 98.8%)
Cut-point 3.8%	-	36.2 (24.3, 50.1%)	92.3 (90.8, 93.5%)	11.3 (7.6, 16.5%)	98.2 (97.3, 98.8%)
SIRS (2+)	0.65 (0.58, 0.73)	51.7 (38.1, 65.1%)	76.2 (74.0, 78.3%)	5.6 (3.9, 7.9%)	98.3 (97.4, 98.9%)
qSOFA (2+)	0.51 (0.46, 0.57)	3.4 (0.9, 12.9%)	99.7 (99.3, 99.8%)	22.2 (5.6, 57.9%)	97.4 (96.5, 98.1%)
NEWS	0.60 (0.52, 0.67)				
Cut-point of 4	-	29.3 (18.4, 43.3%)	80.6 (78.4, 82.6%)	3.9 (2.5, 6.2%)	97.7 (96.6, 98.4%)
Cut-point of 7	-	6.9 (2.6, 17.2%)	98.5 (97.2, 99.2%)	11.4 (3.9, 29.1%)	97.5 (96.6, 98.2%)
Inpatient Potential Bloodstream Infections					
HBS2 - Bedside Score (2+) ^d	0.70 (0.61, 0.78)	88.2 (72.5, 95.5%)	37.9 (32.0, 44.2%)	9.5 (6.7, 13.4%)	97.8 (94.2, 99.2%)
HBST - Bedside Decision Tree (3.3% +) ^d	0.66 (0.57, 0.75)	73.5 (57.0, 85.3%)	50.1 (44.0, 56.2%)	9.8 (6.7, 14.2%)	96.2 (93.0, 98.0%)
SIRS (2+)	0.61 (0.52, 0.70)	73.5 (57.0, 85.3%)	49.9 (44.3, 55.5%)	9.8 (6.7, 14.0%)	96.2 (93.0, 98.0%)
qSOFA (2+)	0.54 (0.46, 0.62)	5.9 (1.5, 20.7%)	98.9 (97.4, 99.5%)	28.6 (7.2, 67.3%)	93.4 (90.8, 95.3%)
NEWS (7+)	0.64 (0.55, 0.73)	11.8 (4.5, 27.5%)	94.3 (88.6, 97.3%)	13.3 (4.4, 34.1%)	93.5 (90.8, 95.5%)
Outpatient Potential Bloodstream Infections					
HBS2 - Bedside Score (2+) ^d	0.73 (0.63, 0.83)	50.0 (29.6, 70.4%)	79.6 (77.0, 81.9%)	3.4 (1.9, 6.1%)	99.1 (98.2, 99.6%)
HBST - Bedside Decision Tree (3.3% +) ^d	0.70 (0.60, 0.81)	45.8 (24.6, 68.7%)	86.0 (82.6, 88.8%)	4.5 (2.5, 7.8%)	99.1 (98.1, 99.6%)
SIRS (2+)	0.53 (0.42, 0.65)	20.8 (7.4, 46.5%)	83.4 (81.2, 85.4%)	1.8 (0.6, 4.8%)	98.7 (97.6, 99.2%)
qSOFA (2+)	0.47 (0.40, 0.54)	0.0 (0.0, 14.2%) ^e	99.9 (99.5, 100%)	0.0 (0.0, 84.2%) ^e	98.6 (97.7, 99.1%)
NEWS (7+)	0.48 (0.37, 0.59)	0.0 (0.0, 14.2%) ^e	99.7 (99.3, 99.9%)	0.0 (0.0, 52.2%) ^e	98.6 (97.7, 99.1%)

^a High-risk bacteremia defined as gram-negative, *Streptococcus* species, or *Staphylococcus aureus* culture confirmed bacteremia

^b 95% Confidence Interval estimated using 2000 stratified bootstrapped replicates

^c 95% Confidence Intervals estimated using generalized estimating equations with robust standard errors

^d Upper-Left Method Selected Optimal Cut-Point

^e 95% Confidence Intervals estimated using Clopper Pearson methods (zeros in either the denominator or numerator prevented reliable generalized estimating equations uncertainty)

estimation)

Figure Legends

Figure 1: Flowchart of hematopoietic cell transplant (HCT) recipients included in the analysis sample and the potential bloodstream infections (PBIs) included in the cohort. PBI was defined as a blood culture that occurred in the absence of recent antibiotic (last 3 days) administration or prescription.

Figure 2: The final high (sepsis) risk bacteremia (A) Decision Tree (HBST) and (B) Score (HBS2). High-risk bacteremia was defined as blood culture confirmed gram-negative, *Staphylococcus aureus*, or *Streptococcus* species bacteremia. (A) The coloring of the terminal nodes refers to the probability of a sepsis event where teal indicates a probability greater than the prevalence of high-risk bacteremia within the cohort (2.4%) and purple indicates a probability less than the high-risk bacteremia prevalence. (B) The Score Range provides all numeric possibilities resulting from the score.

Figure 3: High (sepsis) risk bacteremia Receiver Operating Characteristic (ROC) curves for our novel bedside tools: HCT Bacteremia Sepsis Score (HBS2) and HCT Bacteremia Sepsis Tree (HBST) and existing bedside criteria: Systemic Inflammatory Response Syndrome (SIRS), quick Sequential Organ Failure Assessment (qSOFA), and National Early Warning Score (NEWS) estimated among (A) all examined potential bloodstream infections (PBIs), (B) PBIs identified in inpatient settings, and (C) PBIs identified in outpatient settings. High-risk bacteremia was defined as blood culture confirmed gram-negative, *Staphylococcus aureus*, or *Streptococcus* species bacteremia.

Figure 4: Diverging stacked bar chart of number of true positive, false positive, true negative, and false negative patients out of 100 patients with high-risk bacteremia and 100 without for SIRS 2+, qSOFA 2+, NEWS 7+, HBST 3.3%+, and HBS2 2+. High-risk bacteremia was defined

as blood culture confirmed gram-negative, *Staphylococcus aureus*, or *Streptococcus* species bacteremia.

Figure 5: Ridgeline density plots of the continuous predictors from all examined tools by presence or absence of high-risk bacteremia. Predictor measurements for all examined potential bloodstream infections (PBIs) were included and, in alignment with the primary analysis, the measurement collected closest in time to culture collection was retained. To improve visual discrimination, the following low and high measurement tails were dropped: GCS (< 10), heart rate (< 40 & > 160 beats per minute), SpO2 (< 80%), DBP (< 40 & > 105 mmHg), SBP (< 90 & > 190 mmHg), and respiratory rate (< 10 & > 40 breaths per minute). High-risk bacteremia was defined as blood culture confirmed gram-negative, *Staphylococcus aureus*, or *Streptococcus* species bacteremia and PBI was defined as a blood culture that occurred in the absence of recent antibiotic (last 3 days) administration or prescription.

Figures

Figure 1: Flowchart of Hematopoietic Cell Transplant Recipient (HCT) Study Population and Potential Infection Cohort

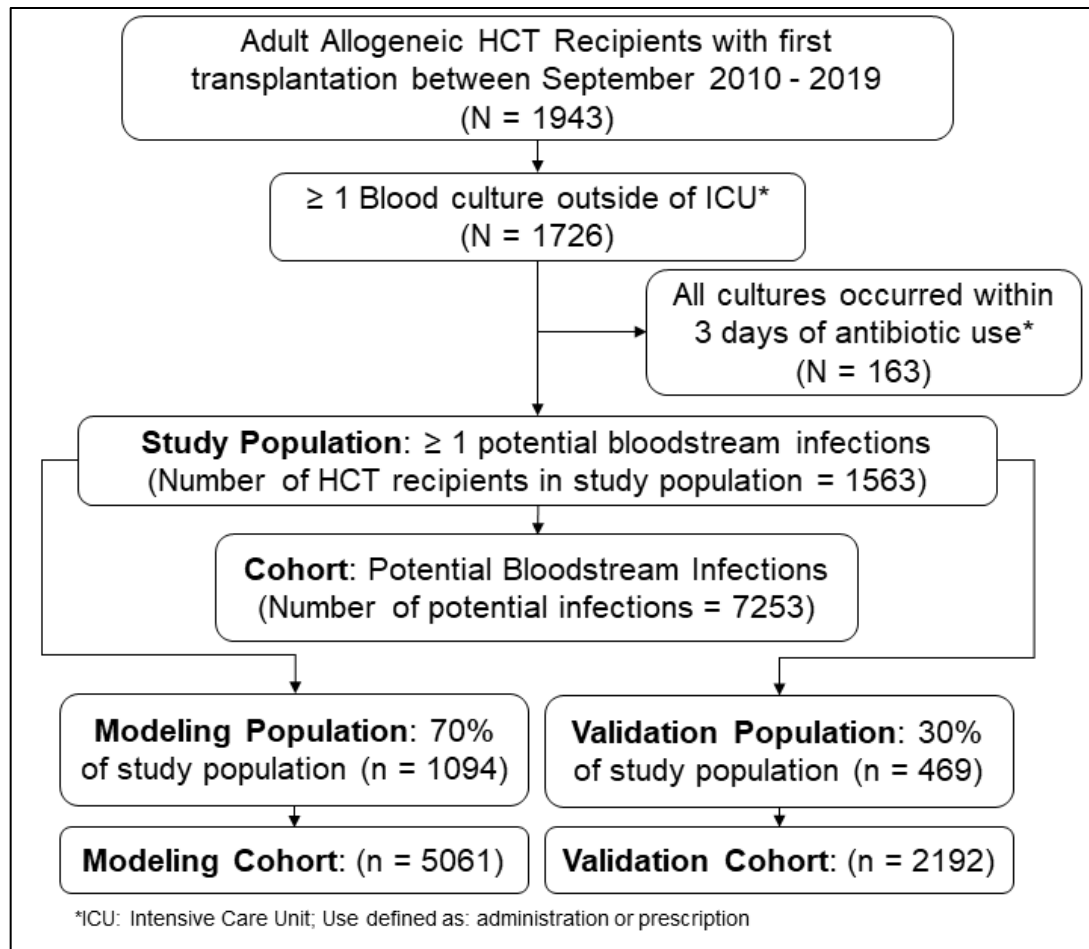


Figure 2: Allogeneic Hematopoietic Stem Cell Transplant Recipient Specific Clinical Prediction Tools (Decision Tree [B], Score [A]) for Bacteremia with High Risk for Sepsis

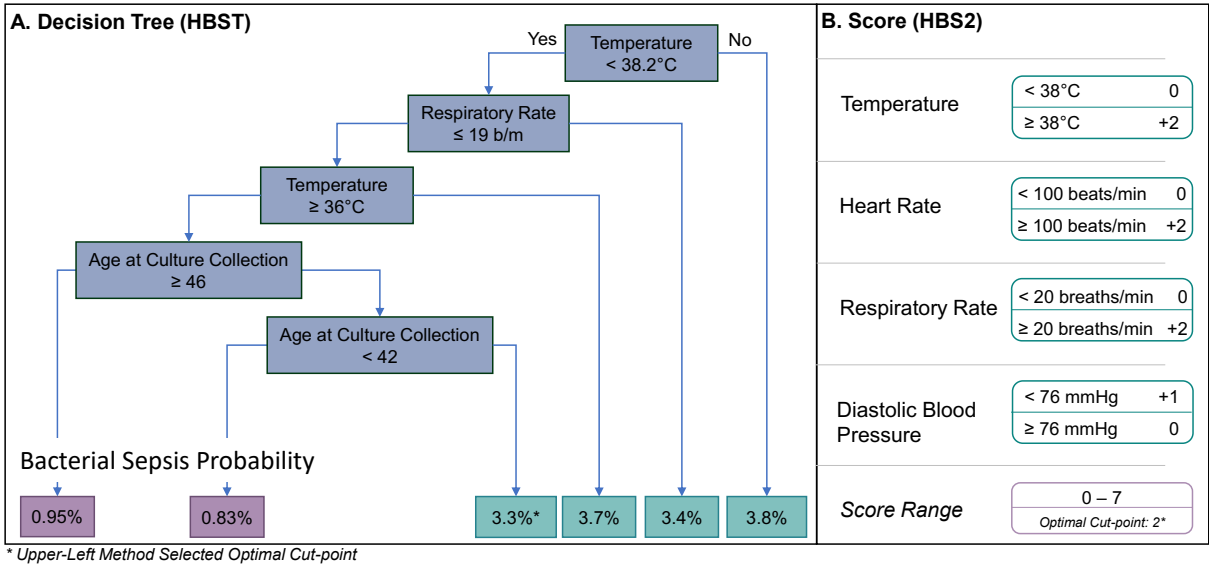


Figure 3: Receiver Operating Characteristic (ROC) Curve for Each Evaluated Prediction Tool in Relation to High-Risk Bacteremia Among Allogeneic Hematopoietic Stem Cell Transplant Recipients with Potential Bloodstream Infections

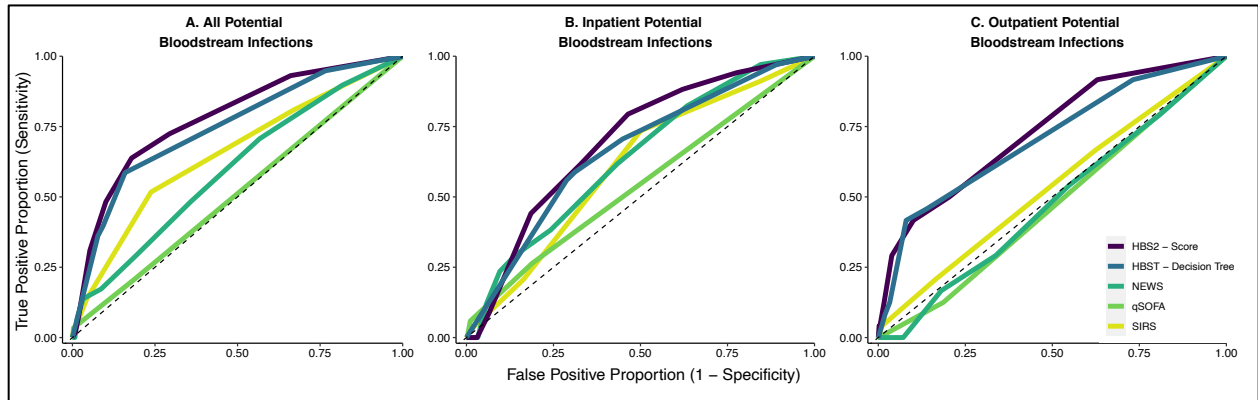


Figure 4: Bedside Tool Results Diverging Stacked Bar Chart Among 100 Allogeneic HCT Recipients with High-Risk Bacteremia and 100 Without

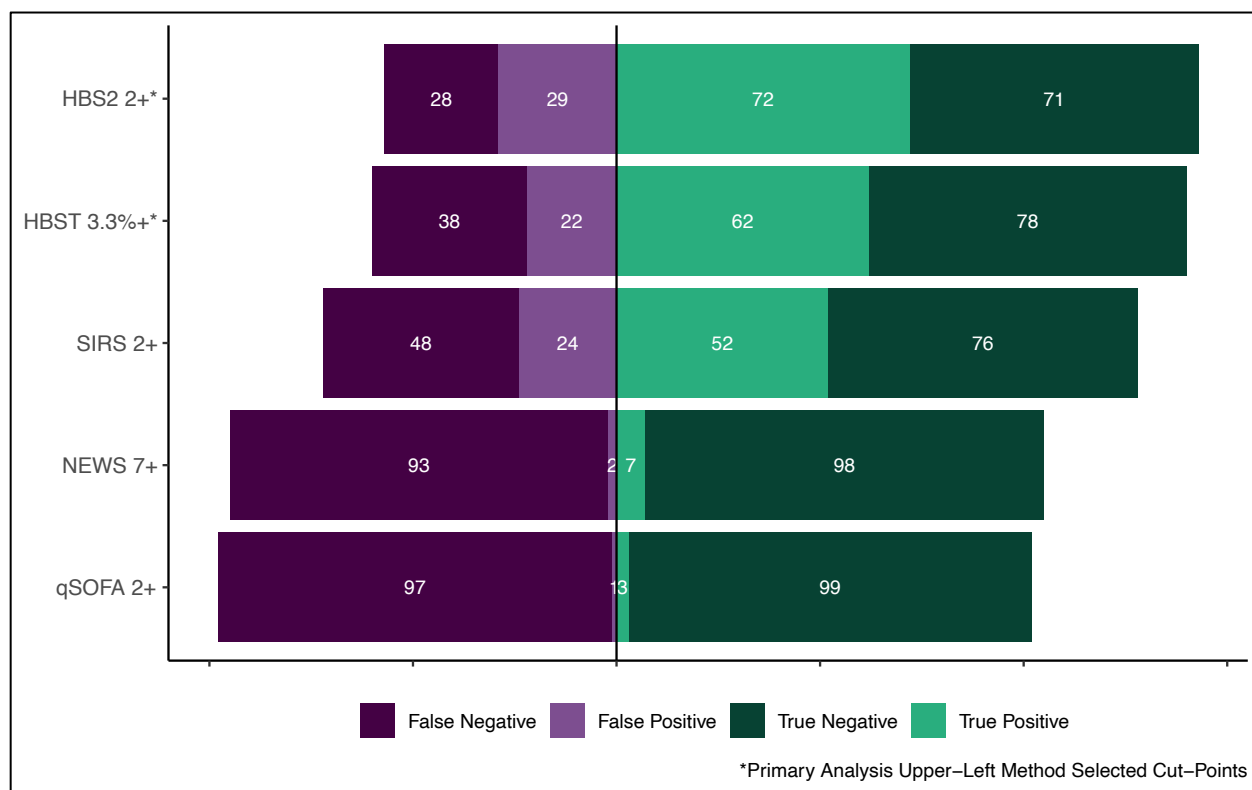
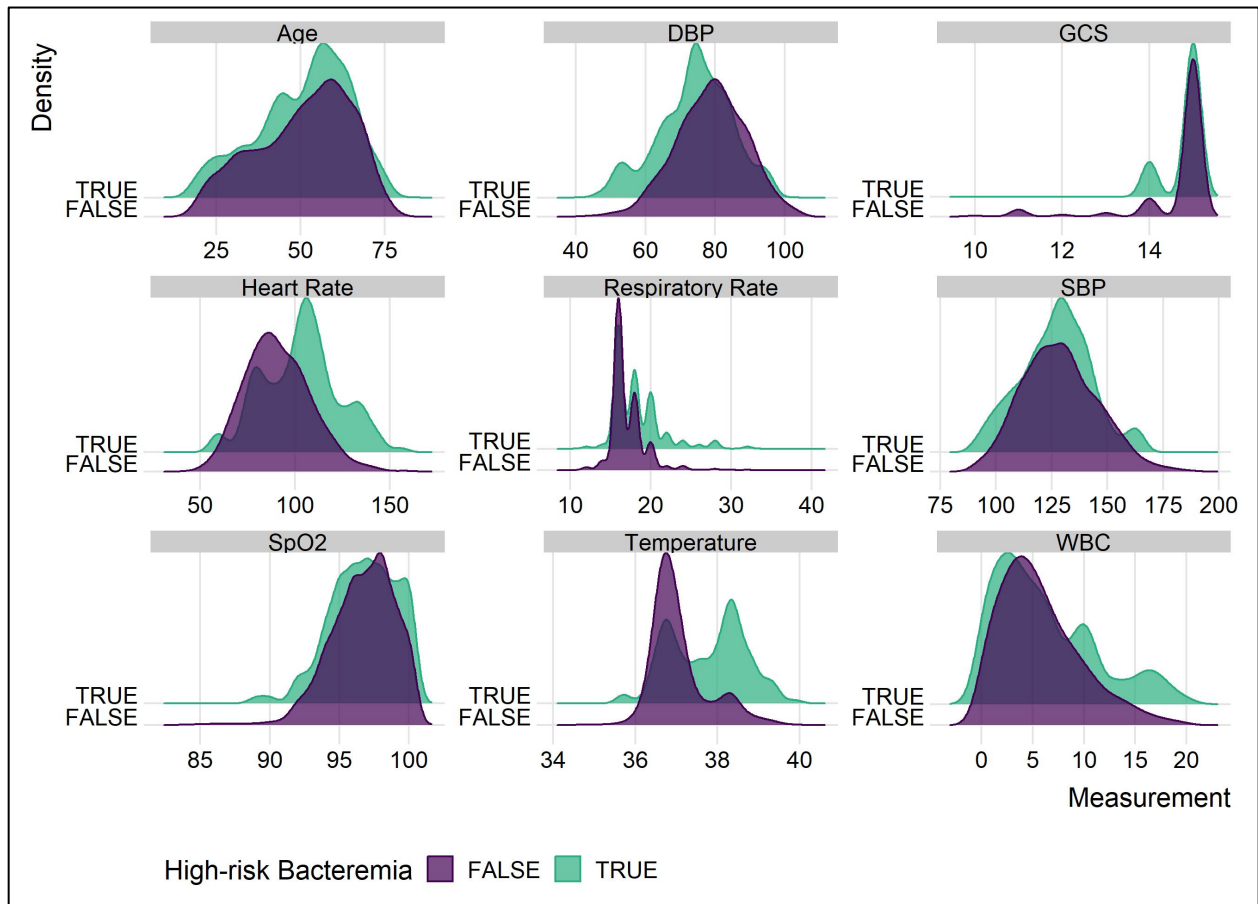


Figure 5: Ridgeline Density Plots of the Continuous Predictors from All Examined Tools by Presence or Absence of High-risk Bacteremia



Paper 2: Development and Validation of a Machine Learning Model to Predict Bacterial Sepsis among Immunocompromised Stem Cell Transplant Recipients

Abstract

Importance: Sepsis disproportionately affects allogeneic hematopoietic cell transplant (allo-HCT) recipients and accurate detection is crucial. However, the atypical presentation of sepsis within this population makes detection challenging and existing clinical sepsis tools have limited predictive value among this high-risk population.

Objective: Develop a full predictor (demographic, transplant, clinical, and laboratory factor) and clinical factor-specific automated bacterial sepsis decision support tool for allo-HCT recipients with potential bloodstream infections (PBIs).

Design, Setting, and Participants: Data from adult allo-HCT recipients transplanted at the Fred Hutchinson Cancer Research Center between June 2010-2019 were randomly divided into 70% modeling and 30% validation datasets. Tools were developed using the area under the curve (AUC) optimized Super Learner and their predictive performance was compared to existing clinical sepsis tools [National Early Warning Score (NEWS), quick Sequential Organ Failure Assessment (qSOFA), and Systemic Inflammatory Response Syndrome (SIRS)] using the validation dataset.

Main Outcomes and Measures: Our primary outcome was high sepsis-risk bacteremia (blood culture confirmed gram-negative species, *Staphylococcus aureus*, or *Streptococcus* species bacteremia) and our secondary outcomes were 10 and 28-day mortality. Tool discrimination and calibration were examined using accuracy metrics and expected versus observed probabilities.

Results: Between June 2010-2019, a majority male (57.2%) and non-Hispanic white (77.9%) population of 1943 allo-HCT recipients received their first transplantation and 1594 experienced ≥ 1 PBI. Of the 8131 observed PBIs, 238 (2.9%) were high sepsis-risk bacteremia. In

relation to high sepsis-risk bacteremia, our full predictor tool had the highest AUROC (0.85, 95% confidence intervals (CI): 0.81-0.89), followed by our clinical factor-specific tool (0.72, CI: 0.66-0.78). SIRS had the highest AUROC of existing tools (0.64, CI: 0.57-0.71). The full predictor tool had the highest AUROCs for PBIs identified in inpatient (0.82) and outpatient (0.82) settings and in relation to 10-day (0.85) and 28-day mortality (0.80).

Conclusion and Relevance: Compared with existing and our clinical factor-specific tools, our full predictor tool had superior predictive accuracy for our primary (high sepsis-risk bacteremia) and secondary (short-term mortality) outcomes for both inpatient and outpatient settings. If used at the time of culture collection, our full predictor tool may inform more timely sepsis detection among allo-HCT recipients.

Introduction

Sepsis disproportionately affects immunocompromised populations, including allogeneic hematopoietic cell transplant (allo-HCT) recipients.¹ Early empiric broad-spectrum antibiotic use among HCT recipients with bacterial sepsis has been shown to reduce the risk of 30-day mortality by three-fold.² However, antibiotic use is associated with increased antibiotic resistance, acquisition of hospital acquired infections and, in HCT recipients, may lead to detrimental microbiota-mediated effects on mortality.³⁻⁸ To minimize these effects, early and accurate sepsis detection among allo-HCT recipients is crucial.

Unfortunately, sepsis is challenging to diagnose among allo-HCT recipients. In immunosuppressed hosts, sepsis has less common presentations, subtle clinical findings and progresses more rapidly than in immunocompetent populations.⁹ Further, common complications of HCT such as engraftment syndrome, anemia, and drug side effects can present with similar signs and symptoms as sepsis.^{9,10} Unsurprisingly, current clinical sepsis tools, which were built among general population patients, have limited predictive value among allo-HCT recipients and their use may lead to missed sepsis events and antibiotic misuse.¹¹

With the goal of improving the accuracy of bacterial sepsis risk prediction at the time of clinical infection consideration (blood culture collection) in both inpatient and outpatient settings, we developed two automated bacterial sepsis decision support tools for allo-HCT recipients with potential bloodstream infections (PBIs) using a state-of-the-art stacked ensembling technique, the Super Learner.¹² Our tools – (1) super HCT bacterial sepsis learner (SHBSL) and (2) clinical factor-specific super HCT bacterial sepsis learner (C-SHBSL) – have the potential to be integrated into electronic medical record (EMR) systems and to provide treatment decision support in the form of sepsis risk probability estimates for each patient or

flags for high-risk patients at the time of blood culture collection (e.Figure1). Through retrospective evaluation, we tested if our population specific tools led to more accurate risk prediction among allo-HCT recipients than commonly used existing tools.

Methods

Study Overview, Population, and Data Collection

We identified adult (≥ 18 years) allo-HCT recipients who agreed to research participation and received their first transplant at the Fred Hutchinson Cancer Research Center / Seattle Cancer Care Alliance (FHCRC/SCCA) between June 2010 and June 2019. Data from the initial 100 days post-transplant – a period of close patient monitoring by inpatient and outpatient transplant teams - or until patient death, if prior to day 100, were retrospectively collected.

Data were queried from a prospectively maintained transplant database and EMRs. We defined outliers as measurements outside of biologically plausible ranges and, following a chart review verification of a randomly selected 10%, dropped outliers prior to missing value imputation. This study was approved by the FHCRC Institutional Review Board and followed the TRIPOD diagnostic/prognostic reporting guidelines.

Cohort Development

We developed and evaluated the tools among all PBIs that occurred within the study period. We defined PBI as any blood culture, including surveillance cultures, collected in the absence of recent (last 3 days) antibiotic administration or prescription other than the center recommended neutropenia prophylaxis (levofloxacin). We considered multiple cultures collected within one hour to be the same event (culture time defined as time of first non-surveillance culture) and dropped follow-up cultures, defined as cultures collected continuously (within 36 hours of prior) for the 14 days following a positive culture or until distinct organisms were identified, whichever came first.

Patients for whom any PBIs were identified were randomly divided into a modeling (70%) or validation (30%) group and their PBIs were assigned to the modeling or validation dataset based on the patient's grouping.

Prophylaxis and Culture Collection Guidelines

Center guidelines during the study period have been described elsewhere.¹¹ Briefly, cultures were recommended for patients with fevers and surveillance cultures were recommended weekly or bi-weekly for patients on high dose glucocorticoids (>0.5mg/kg). Levofloxacin was used as the primary agent for antibiotic prophylaxis during neutropenic periods while patients with neutropenic fever received empiric therapy with ceftazidime, cefepime, or meropenem.

Outcome and Predictors

The atypical presentation of sepsis among HCT recipients prevented us from using a symptom based primary endpoint.^{13,14} Instead, we selected bacteremia with high risk for sepsis - blood culture with gram-negative species, *Staphylococcus aureus*, or *Streptococcus* species bacteremia – based on evidence that sepsis is significantly more likely among HCT recipients with PBIs and the selected bloodstream infections than without (See Supplement: Outcome Justification).¹⁵

Because our endpoint excludes blood culture negative sepsis events,^{16,17} we feared our tools may capture bacteremia but not sepsis risk. To test our tools ability to capture risk in general, not just bacteremia risk, we examined two short-term mortality secondary outcomes (10-day and 28-day mortality) under the assumption that mortality was more common among patients with sepsis than without.¹⁸ However, unlike Seymour et al., we chose against a primary mortality endpoint because we were unable to account for the impact antimicrobial and supportive treatment undergone by the vast majority of our retrospectively collected sample have on these outcomes.^{19–21}

We considered demographic and clinical factors easily assessed at the bedside as potential predictors in C-SHBSL. For SHBSL, we additionally included a wide array of comorbidities, transplant factors, biomarkers, vitals, and recent infection data (e.Table1). Additionally, in SHBSL, we included binary indicators of measurement missingness and, to account for trends in numerical predictors during the window of measurement collection, linear change-over-time predictors developed using univariable linear regressions.

When applicable, predictors were standardized to a uniform unit (i.e. weight in kg) and lesser or greater than values were converted to the first number included in the statement (e.g. 35°C for $\leq 35^{\circ}\text{C}$). Temperatures were standardized to oral temperatures by adding 0.45°C to axillary and subtracting 0.2°C from arterial measurements.²² Indicator (0/1) variables were developed from categorical predictors, all of which were nominal. Each predictor was standardized to a mean of zero and standard deviation of one.

Predictor data for the 24 hours prior to and 2 hours following culture collection were extracted and the measurement taken closest in time to collection was retained. Potential predictors with more than 35% missingness were excluded. Binary missingness and change-over-time predictors were included if the static measurement was included. Any missing observations for frequently captured predictors were filled using a single normal value imputation approach, a common approach in sepsis prediction (See Supplement: Imputation Approach, e.Figure2/e.Table2).^{18,23–25}

Tool Development

We developed our tools in the modeling dataset using the Super Learner with 10-fold cross-validation, optimized with respect to area under the receiver operating characteristic curve (AUROC). The Super Learner is a statistical learning technique that uses cross-validation to construct an optimal ensemble (weighted average) of user-supplied candidate algorithms.²⁶ The

resulting model, the *Super Learner*, has been proven (in large samples) to perform as well as and shown (in small samples) to perform essentially as well as the best performing ensemble of specified algorithms.^{26–28} To optimize the resulting *Super Learners*, our library included a broad selection of parametric, nonparametric, linear and non-linear algorithms. Additionally, our library included hyperparameter and screening algorithm variants (See Supplement: Detailed Description of Super Learner Library). As with the modeling/validation randomization, observations were partitioned into cross-validation folds on the person level.

Tool Evaluation (Calibration and Discrimination)

Using the validation dataset, we compared the calibration and discrimination of our tools to three existing clinical sepsis tools available for use in outpatient settings (National Early Warning Score [NEWS], quick Sequential Organ Failure Assessment [qSOFA], and Systemic Inflammatory Response Syndrome [SIRS]). We assessed calibration by comparing the mean predicted versus observed outcome probabilities and examined calibration plots.^{29–32} To evaluate discrimination, we compared AUROCs, area under the precision/recall curve (pr-AUC), categorical net reclassification index (NRI; See Supplement: Net Reclassification), and predictive accuracy metrics (sensitivities, specificities, and negative [NPV] and positive predictive values [PPV]) for all cut-points (primary analysis) and qSOFA/SIRS (2+), NEWS (4 and 7+), and SHBSL/C-SHBSL (primary analysis selected optimal [Upper-Left] cut-points) for secondary and sensitivity analyses. Optimal cut-points (Upper-Left) were identified for the primary and secondary analyses. To visually compare tool discrimination, we generated ROC and Precision/Recall curves.

We performed secondary analyses looking at tool performance by inpatient status at time of culture collection and in relation to short-term (10 and 28-day) mortality.^{18,33} Inpatient status was defined for each calendar day and patients were considered inpatient for the full day of

hospital admission, discharge and during intensive care unit (ICU) stays. Additionally, we estimated the cross-validated AUROCs (cvAUCs) of our tools using 10-fold nested cross-validation and the bootstrapped AUROCs (btAUC) of existing tools using 1000 modeling/validation split variants. All analyses were performed using R 3.6.2 base functions and functions from the rms, fbroc, OptimalCutpoints, and SuperLearner packages.^{12,32,34–36}

Sensitivity Analyses

We performed numerous sensitivity analyses to examine the generalizability of our tools. Specifically, we explored whether our evaluation results were robust to varying demographic distributions, missing data, recent antibiotic use, data collection windows, and culture collection and inclusion definitions. The sensitivity analyses were performed using the validation dataset and are described in the *Supplement: Sensitivity Analyses*.

Results

Study Population and Cohort

Between June 2010 and June 2019, a majority male (57.2%) and non-Hispanic white (77.9%) population of 1943 individuals received their first allo-HCT at the SCCA/FHCRC and 1594 experienced at least one PBI (Table1, e.Figure3). During follow-up, these patients experienced 8131 PBIs, 6046 (74.4%) were identified in outpatient and 2085 (25.6%) in inpatient settings (e.Figure4). Patients who developed high-risk bacteremia were more likely to have experienced grade 4 acute GVHD and died during follow-up (aGVHD: 9.8%, death: 17.1%) than patients who did not (aGVHD: 2.7%, death: 7.5%, Table1).

Of all PBIs, 238/8131 (2.9%) were defined as high-risk bacteremia and 258 (3.2%) and 74 (0.9%) occurred within 28 and 10 days of death, respectively. The modeling dataset contained 5845 PBIs, of which, 168 (2.9%) were high-risk bacteremia. Of the 2286 PBIs in the validation dataset, 70 (3.1%) were high-risk bacteremia. Patients randomized to the modeling and

validation groups were demographically and medically similar (e.Table3).

Predictor Selection

Of the 35 vitals/biomarkers examined, 17 were discarded due to excessive missingness (>35%; e.Table1).

Tool Evaluation (Calibration and Discrimination)

In relation to high-risk bacteremia, SHBSL had the highest AUROC and pr-AUC [AUC: 0.85, 95% confidence intervals (CI): 0.81-0.89; pr-AUC: 0.20] and C-SHBSL had the second highest (AUC: 0.72, CI: 0.66-0.78; pr-AUC: 0.13). qSOFA had the lowest AUROC (0.51, CI: 0.46-0.56) and all three existing tools had pr-AUCs of 0.07 (Table2; Figure1A/B). At Upper-Left method selected cut-points, SHBSL and C-SHBSL had sensitivities of 80.0% (CI: 68.7-88.6%) and 65.7% (CI: 53.4-76.7%), respectively, and specificities of 72.8% (CI: 70.9-74.6%) and 66.9% (CI: 64.9-68.9%), respectively (Table2). At cut-points with sensitivities of $\geq 90\%$, SHBSL had PPVs of $\leq 7.3\%$ and C-SHBSL of $\leq 4\%$. All tools had high NPVs (range: 96.9-100%, e.Table4). Compared with the other tools, SHBSL had significantly better risk classification (NRI 0.31-0.52, Pvalues ≤ 0.001 ; e.Table5).

SHBSL had the highest AUROCs for both secondary analyses. Among both location stratified cohorts, SHBSL had AUROCs of 0.82. Inpatient PBIs had higher predictive scores than outpatient PBIs (e.Figure5) and the Upper-Left method selected cut-points for SHBSL, C-SHBSL, and NEWS varied by location (Table2). In relation to 10-day and 28-day mortality, SHBSL had AUROCs of 0.85 and 0.80, respectively. When compared to the primary analysis (all PBIs), existing tools had lower outpatient but higher short-term mortality AUROCs (Table3, Figure2). SHBSL and C-SHBSL had cvAUCs of 0.83 (CI: 0.70-0.91) and 0.74 (CI: 0.59-0.86), respectively (e.Figure6; e.Table6). SIRS had the highest btAUC 0.64 (CI: 0.52-0.77) (e.Table6).

The mean predicted high-risk bacteremia probability of each tool matched the observed

proportion in the validation dataset (3.1%; e.Table7) The calibration plots show that SHBSL had the best overall calibration (the predicted vs. observed probability line falls closest to the ideal [diagonal] calibration line; Figure1.C).

Sensitivity Analyses

SHBSL had the highest AUROC under all examined scenarios (range: 0.81-0.94) and higher pr-AUCs than existing tools under all examined scenarios (range: 0.11-0.45; e.Table 8-12) and C-SHBSL for all but the non-Hispanic white only scenario (SHBSL: 0.15, C-SHBSL: 0.17; e.Table8). SHBSL and C-SHBSL had lower AUROCs (0.02 and 0.01 points) among the surveillance cultures removed cohort than in the primary analysis. Under varying culture collection (e.Table11), racial makeup (e.Table9), and missingness (e.Table8) scenarios, SHBSL had AUROCs between 0.15 and 0.31 points higher than that of existing tools and primary analysis selected optimal cut-point sensitivities between 70.0% and 94.1% and specificities between 72.3% and 75.8%.

Discussion

We built two automated prediction tools (full predictor: SHBSL; clinical factor-specific: C-SHBSL) to provide sepsis related treatment decision support at the time of blood culture collection among allo-HCT recipients in either inpatient or outpatient settings. We developed these tools using a state-of-the-art ensemble modeling technique (the Super Learner) and included demographic and vital predictors in both tools as well as a wide array of demographic, clinical, and transplant predictors in SHBSL. In comparison to existing clinical sepsis tools, these tools had higher estimated discrimination and calibration in relation to high sepsis-risk bacteremia and short-term mortality among allo-HCT recipients with PBIs.

Regardless of cohort, sepsis proxy endpoint, or data scenario, SHBSL and C-SHBSL had moderate to high (0.80-0.94) and moderate (0.60-0.86) AUROCs, respectively, and existing tools

had low to moderate (0.45-0.71) AUROCs. In addition, SHBSL had higher PPVs relative to all sensitivities than existing tools and to all sensitivities >20% than C-SHBSL (pr-AUC, precision/recall curve). These results suggests that (1) the additional data cleaning and processing required to run SHBSL instead of C-SHBSL is justified, (2) SHBSL can be run on complete or incomplete data (e.Table7), (3) the superior observed performance of SHBSL is likely generalizable to other transplant centers (Sensitivity Analyses), (4) SBHSL can be run in inpatient or outpatient settings but location specific cut-points are needed (Table2, e.Figure5) and (5) SHBSL would likely lead to fewer false flags relative to any sensitivities than existing tools (Table2 – prAUC; Precision/Recall Curve – Figure1.B). Finally, the fact SHBSL had superior predictive performance irrespective of sepsis related endpoint suggests that, compared to existing tools, SHBSL is better able to predict sepsis risk, not just high-risk bacteremia, among allo-HCT recipients with PBIs.

We observed lower than expected sensitivities for NEWS and SIRS in relation to short-term mortality. Previous 10-day mortality sensitivity estimates for NEWS7+ and SIRS2+ among allo-HCT recipients were 63.9 and 69.4 percentage points higher than our estimates respectively.¹¹ This difference is likely driven by our decision to evaluate potential (all blood cultures) instead of the commonly used suspected (body fluid cultures collected within a specific antibiotic time epoch) infections (SIs) and missingness frequency differences (SI study temperature missingness 6%; ours: 35%).^{11,18} While examining PBIs likely contributed to our large amount of missingness and means we included some cultures in the absence of clinical suspicion of infection (surveillance cultures), we evaluated PBIs rather than SIs because our primary endpoint required blood culture confirmation and we wanted our tools to inform antibiotic treatment use rather than predict disease escalation among patients already receiving

antibiotic treatment.^{33,37} To ensure our tools were not biased towards surveillance cultures, we performed a sensitivity analysis excluding surveillance cultures and found that neither C-SHBSL nor SHBSL performed meaningfully different (e.Table11).

Unlike most automated sepsis prediction tools – which were built to improve early detection of clinically identified sepsis through real-time prediction - we built SBHSL/C-SHBSL to improve the accuracy of the sepsis-risk prediction itself. This decision was driven by the atypical presentation of sepsis within this population, the desire to build tools functional in both inpatient and outpatient settings (limited continuous data), the limited accuracy of existing clinical scores, and reliance on outcomes that require cultures (absence may not equal negative). While this prevents direct comparison between SBHSL/C-SHBSL and real-time tools, the performances of real-time tools provide a reference to what constitutes as “good.” According to a 2020 systematic review, the AUROCs of real-time tools and SHBSL are similar (inpatient/ICU: SHBSL: 0.82, Real-Time: 0.68-0.99; ED/Outpatient: SHBSL: 0.82, Real-Time: 0.87-0.97).³⁸ This comparison demonstrates that, not only does SHBSL have superior performance compared to existing clinical tools, its performance is numerically comparable to state-of-the-art real-time prediction tools.

Our study has many strengths. First, the 100-day observation of our large sample allowed us to develop bacterial sepsis prediction tools for recipients throughout their most immunocompromised periods regardless of their inpatient status. Second, our data-driven modeling approach allowed the unique presentation of bacterial sepsis within this population to be captured. Third, we showed that our predictive tool evaluation was robust to numerous assumptions made including those regarding missing data and culture inclusion. Finally, through secondary analyses using alternative endpoints, we were able to show that SHBSL captured risk,

regardless of proxy endpoint, with similarly high AUROCs.

Study Limitations

Our study has several limitations. First, our sample is reflective of the northwestern US transplant community but may not reflect the larger transplant population. Second, our tools rely on culture collection and are potentially biased toward our center's practices. Third, our analyses relied on retrospective data and suffered from data collection and cleaning related biases. Such limitations are not unique to our study and stem from our inability to evaluate our tools externally or prospectively at the time of this study. Though we performed numerous sensitivity analyses testing our tools' generalizability, future analyses should validate them among prospectively collected and external data. Additionally, while it is possible to estimate predictor importance from the *Super Learner*, the inclusion of correlated predictors (ex. weight and height) prevented reliable estimation.^{39,40} Further, despite rapid advancement in EMR systems, EMR integration of automated tools – including ours – would be costly and in the absence of a well maintained and organized system, may be impossible. Finally, the Super Learner ensembles had lower (<1%) cvAUCs than a supplied algorithm, however, we presented the ensembles as our final models because they provide the least biased options for future updates (See Supplement: Ensemble Selection Rationale).

Conclusion

We built two automated bacterial sepsis decision support tools for allo-HCT recipients with PBIs in either inpatient or outpatient settings. When compared to existing prediction tools and C-SHBSL, we found SHBSL to have superior performance regardless of patient location at time of culture collection and examined sepsis related endpoint. Our results suggest that, if SHBSL were to be run at the time of blood culture collection, it would provide clinically relevant information regarding bacterial sepsis risk and antibiotic initiation needs of allo-HCT recipients

References

1. Kumar G, Ahmad S, Taneja A, et al. Severe sepsis in hematopoietic stem cell transplant recipients*. *Crit Care Med*. 2015;43(2):411-421. doi:10.1097/CCM.0000000000000714
2. Mikulska M, Del Bono V, Bruzzi P, et al. Mortality after bloodstream infections in allogeneic haematopoietic stem cell transplant (HSCT) recipients. *Infection*. 2012;40(3):271-278. doi:10.1007/s15010-011-0229-y
3. Weber D, Jenq RR, Peled JU, et al. Microbiota Disruption Induced by Early Use of Broad-Spectrum Antibiotics Is an Independent Risk Factor of Outcome after Allogeneic Stem Cell Transplantation. *Biology of Blood and Marrow Transplantation*. 2017;23(5):845-852. doi:10.1016/j.bbmt.2017.02.006
4. Golob JL, Pergam SA, Srinivasan S, et al. Stool Microbiota at Neutrophil Recovery Is Predictive for Severe Acute Graft vs Host Disease After Hematopoietic Cell Transplantation. *Clin Infect Dis*. 2017;65(12):1984-1991. doi:10.1093/cid/cix699
5. Holler E, Butzhammer P, Schmid K, et al. Metagenomic Analysis of the Stool Microbiome in Patients Receiving Allogeneic Stem Cell Transplantation: Loss of Diversity Is Associated with Use of Systemic Antibiotics and More Pronounced in Gastrointestinal Graft-versus-Host Disease. *Biology of Blood and Marrow Transplantation*. 2014;20(5):640-645. doi:10.1016/j.bbmt.2014.01.030
6. Ogimi C, Krantz EM, Golob JL, et al. Antibiotic Exposure Prior to Respiratory Viral Infection Is Associated with Progression to Lower Respiratory Tract Disease in Allogeneic Hematopoietic Cell Transplant Recipients. *Biol Blood Marrow Transplant*. Published online May 16, 2018. doi:10.1016/j.bbmt.2018.05.016
7. Galloway-Peña JR, Jenq RR, Shelburne SA. Can Consideration of the Microbiome Improve Antimicrobial Utilization and Treatment Outcomes in the Oncology Patient? *Clin Cancer Res*. 2017;23(13):3263-3268. doi:10.1158/1078-0432.CCR-16-3173
8. Gouliouris T, Warne B, Cartwright EJP, et al. Duration of exposure to multiple antibiotics is associated with increased risk of VRE bacteraemia: a nested case-control study. *J Antimicrob Chemother*. 2018;73(6):1692-1699. doi:10.1093/jac/dky075
9. Dadwal S, Eftekhari Z, Tushondra T, Johnson D. A Machine-Learning Based Prediction Model for Sepsis in Patients Undergoing Hematopoietic Stem Cell Transplantation. *ASH Annual Meeting*. 2018;the Systemic Inflammatory Response Syndrome (SIRS) and quick-Sequential Organ Failure Assessment (qSOFA) criteria. Accessed March 2, 2019. <https://ash.confex.com/ash/2018/webprogram/Paper117002.html>
10. Kochanek M, Schalk E, von Bergwelt-Baildon M, et al. Management of sepsis in neutropenic cancer patients: 2018 guidelines from the Infectious Diseases Working Party (AGIHO) and Intensive Care Working Party (iCHOP) of the German Society of Hematology and Medical Oncology (DGHO). *Ann Hematol*. 2019;98(5):1051-1069. doi:10.1007/s00277-019-03622-0
11. Lind ML, Phipps AI, Mooney S, et al. Predictive value of three clinical criteria for sepsis (qSOFA, SIRS, and NEWS) with respect to short-term mortality in allogeneic hematopoietic cell transplant recipients with suspected infections. *Clin Infect Dis*. Published online March 4, 2020. doi:10.1093/cid/ciaa214
12. Polley E, LeDell E, Kennedy C, Lendle S, Laan M van der. *SuperLearner: Super Learner Prediction*.; 2018. Accessed April 26, 2019. <https://CRAN.R-project.org/package=SuperLearner>

13. Kam HJ, Kim HY. Learning representations for the early detection of sepsis with deep neural networks. *Comput Biol Med.* 2017;89:248-255. doi:10.1016/j.compbiomed.2017.08.015
14. Calvert JS, Price DA, Chettipally UK, et al. A computational approach to early sepsis detection. *Comput Biol Med.* 2016;74:69-73. doi:10.1016/j.compbiomed.2016.05.003
15. Hämäläinen S, Kuittinen T, Matinlauri I, Nousiainen T, Koivula I, Jantunen E. Severe sepsis in autologous stem cell transplant recipients: Microbiological aetiology, risk factors and outcome. *Scandinavian Journal of Infectious Diseases.* 2009;41(1):14-20. doi:10.1080/00365540802454706
16. Kethireddy S, Bilgili B, Sees A, et al. Culture-Negative Septic Shock Compared With Culture-Positive Septic Shock: A Retrospective Cohort Study. *Crit Care Med.* 2018;46(4):506-512. doi:10.1097/CCM.0000000000002924
17. Gupta S, Sakhuja A, Kumar G, McGrath E, Nanchal RS, Kashani KB. Culture-Negative Severe Sepsis: Nationwide Trends and Outcomes. *Chest.* 2016;150(6):1251-1259. doi:10.1016/j.chest.2016.08.1460
18. Seymour CW, Liu VX, Iwashyna TJ, et al. Assessment of Clinical Criteria for Sepsis: For the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA.* 2016;315(8):762-774. doi:10.1001/jama.2016.0288
19. Abe R, Oda S, Sadahiro T, et al. Gram-negative bacteremia induces greater magnitude of inflammatory response than Gram-positive bacteremia. *Crit Care.* 2010;14(2):R27. doi:10.1186/cc8898
20. Alexandraki I, Palacio C. Gram-negative versus Gram-positive bacteremia: what is more alarmin(g)? *Crit Care.* 2010;14(3):161. doi:10.1186/cc9013
21. Kumar A, Roberts D, Wood KE, et al. Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. *Crit Care Med.* 2006;34(6):1589-1596. doi:10.1097/01.CCM.0000217961.75225.E9
22. Healthwise Staff. Fever Temperatures: Accuracy and Comparison. Published online September 23, 2018. Accessed May 15, 2019. <https://www.mottchildren.org/health-library/tw9223>
23. Knaus WA, Zimmerman JE, Wagner DP, Draper EA, Lawrence DE. APACHE-acute physiology and chronic health evaluation: a physiologically based classification system. *Crit Care Med.* 1981;9(8):591-597. doi:10.1097/00003246-198108000-00008
24. Seymour CW, Kahn JM, Cooke CR, Watkins TR, Heckbert SR, Rea TD. Prediction of critical illness during out-of-hospital emergency care. *JAMA.* 2010;304(7):747-754. doi:10.1001/jama.2010.1140
25. Vincent JL, Moreno R, Takala J, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. *Intensive Care Med.* 1996;22(7):707-710. doi:10.1007/bf01709751
26. Laan MJ van der, Rose S. *Targeted Learning: Causal Inference for Observational and Experimental Data.* 2011 edition. Springer; 2011.
27. Benitez AE, Musinguzi N, Bangsberg DR, et al. Super learner analysis of real-time electronically monitored adherence to antiretroviral therapy under constrained optimization and comparison to non-differentiated care approaches for persons living with HIV in rural Uganda. *J Int AIDS Soc.* 2020;23(3). doi:10.1002/jia2.25467

28. Houssaini A, Assoumou L, Marcelin AG, Molina JM, Calvez V, Flandre P. Investigation of Super Learner Methodology on HIV-1 Small Sample: Application on Jaguar Trial Data. *AIDS Res Treat*. 2012;2012. doi:10.1155/2012/478467
29. Pirracchio R, Petersen ML, Carone M, Rigon MR, Chevret S, van der LAAN MJ. Mortality prediction in the ICU: can we do better? Results from the Super ICU Learner Algorithm (SICULA) project, a population-based study. *Lancet Respir Med*. 2015;3(1):42-52. doi:10.1016/S2213-2600(14)70239-5
30. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for some traditional and novel measures. *Epidemiology*. 2010;21(1):128-138. doi:10.1097/EDE.0b013e3181c30fb2
31. Cox DR. Two further applications of a model for binary regression. *Biometrika*. 1958;45(3-4):562-565. doi:10.1093/biomet/45.3-4.562
32. Jr FEH. *Rms: Regression Modeling Strategies*.; 2020. Accessed October 23, 2020. <https://CRAN.R-project.org/package=rms>
33. Singer M, Deutschman CS, Seymour CW, et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA*. 2016;315(8):801-810. doi:10.1001/jama.2016.0287
34. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.; 2019. Accessed August 30, 2019. <https://www.R-project.org/>
35. Peter E. *Fbrc: Fast Algorithms to Bootstrap Receiver Operating Characteristics Curves*.; 2019. <https://CRAN.R-project.org/package=fbrc>
36. López-Ratón M, Rodríguez-Álvarez MX, Cadarso-Suárez C, Gude-Sampedro F. OptimalCutpoints: An R Package for Selecting Optimal Cutpoints in Diagnostic Tests. *Journal of Statistical Software*. 2014;61(1):1-36. doi:10.18637/jss.v061.i08
37. Bone RC, Grodzin CJ, Balk RA. Sepsis: A New Hypothesis for Pathogenesis of the Disease Process. *Chest*. 1997;112(1):235-243. doi:10.1378/chest.112.1.235
38. Fleuren LM, Klausch TLT, Zwager CL, et al. Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. *Intensive Care Med*. 2020;46(3):383-400. doi:10.1007/s00134-019-05872-y
39. Greenwell BM, Boehmke BC, McCarthy AJ. A Simple and Effective Model-Based Variable Importance Measure. *arXiv:180504755 [cs, stat]*. Published online May 12, 2018. Accessed January 28, 2021. <http://arxiv.org/abs/1805.04755>
40. Lieberman M, Morris J. The Precise Effect of Multicollinearity on Classification Prediction. *Multiple Linear Regression Viewpoint*. 2014;40(1):5-10.

Tables

Table1: Baseline Characteristics of Participants by Occurrence of High-Risk Bacteremia

	Full Population	Developed High-Risk Bacteremia	Did Not Develop High-Risk Bacteremia
	N = 1594	N = 193	N = 1401
Calendar Year at Tx (Median, IQR) ^a	2014 (2012, 2016)	2013 (2011, 2016)	2014 (2012, 2016)
Age at Tx (Median, IQR) ^a	54 (43, 63)	53 (41, 62)	55 (43, 63)
Gender (Count, % Male)	911 (57.2)	107 (55.4)	804 (57.4)
Ethnicity/Race (Count, %)			
American Indian or Alaskan Native ^b	24 (1.5)	2 (1.0)	22 (1.6)
Asian ^b	119 (7.5)	16 (8.3)	103 (7.4)
Black ^b	25 (1.6)	5 (2.6)	20 (1.4)
Multiple ^b	22 (1.4)	5 (2.6)	17 (1.2)
Native Hawaiian or other Pacific Islander ^b	25 (1.6)	5 (2.6)	20 (1.4)
Unknown ^b	54 (3.4)	5 (2.6)	49 (3.5)
White ^b	1242 (77.9)	141 (73.1)	1101 (78.6)
LatinX	83 (5.2)	14 (7.3)	69 (4.9)
Underlying Disease			
Acute Lymphoblastic Leukemia	212 (13.3)	34 (17.6)	178 (12.7)
Acute Nonlymphocytic Leukemia	580 (36.4)	64 (33.2)	516 (36.8)
Myelodysplastic Syndromes	410 (25.7)	48 (24.9)	362 (25.8)
Other	392 (24.6)	47 (24.4)	345 (24.6)
HCT Cell Donor Type			
Bone marrow	126 (7.9)	20 (10.4)	106 (7.6)
Bone marrow, peripheral blood stem cells	3 (0.2)	1 (0.5)	2 (0.1)
Umbilical cord blood	179 (11.2)	31 (16.1)	148 (10.6)
Peripheral blood stem cells	1286 (80.7)	141 (73.1)	1145 (81.7)
Donor Relation (Count, %)			
Child	38 (2.4)	6 (3.1)	32 (2.3)
Half Sibling	5 (0.3)	0 (0)	5 (0.4)
Not Related	1107 (69.4)	135 (69.9)	972 (69.4)
Parent	17 (1.1)	4 (2.1)	13 (0.9)
Relative	1 (0.1)	0 (0)	1 (0.1)
Sibling	426 (26.7)	48 (24.9)	378 (27.0)
Acute GVHD (Count, %) ^c			
Grade 0	328 (20.6)	32 (16.6)	296 (21.1)
Grade 1	78 (4.9)	7 (3.6)	71 (5.1)
Grade 2	952 (59.7)	97 (50.3)	855 (61.0)
Grade 3	179 (11.2)	38 (19.7)	141 (10.1)
Grade 4	57 (3.6)	19 (9.8)	38 (2.7)
High-Risk Bacteremia ^d (Count, %) ^e			
Gram-Negative species	104 (6.5)	104 (53.9)	0 (0.0)
<i>Staphylococcus aureus</i>	8 (0.5)	8 (4.1)	0 (0.0)
<i>Streptococcus</i> species	87 (5.5)	87 (45.1)	0 (0.0)
Neutropenia Prophylaxis (Count, %) ^e	1439 (90.3)	178 (92.2)	1261 (90.0)
Died (Count, %) ^e	138 (8.7)	33 (17.1)	105 (7.5)

^a IQR: Interquartile range, Tx: transplant

^b Non-Hispanic

^c Experiences during follow-up

^d Six patients experienced more than one type of high-risk bacteremia event and appear in two rows

^e Defined as administration or prescription of levofloxacin during follow-up

Table 2: Discrimination - Predictive Accuracy of Sepsis Criteria in Relation to High Sepsis-risk Bacteremia Among Allogeneic HCT Recipients with Potential Bloodstream Infections

	AUROC ^{b,c}	PR-AUC ^b	Sensitivity ^d	Specificity ^d	Positive Predictive Value ^d	Negative Predictive Value ^d	Upper-Left Cut-point ^e
Clinical Factor Specific - C-SHBSL ^a	0.72 (0.66, 0.78)	0.13	65.7% (53.4, 76.7%)	66.9% (64.9, 68.9%)	5.9% (4.4, 7.8%)	98.4% (97.6, 99.0%)	9.3%
Full Predictor-SHBSL ^a	0.85 (0.81, 0.89)	0.20	80.0% (68.7, 88.6%)	72.8% (70.9, 74.6%)	8.5% (6.5, 10.9%)	99.1% (98.6, 99.5%)	5.0%
SIRS (2+)	0.64 (0.57, 0.71)	0.07	54.3% (41.9, 66.3%)	75.9% (74.0, 77.6%)	6.6% (4.7, 9.0%)	98.1% (97.4, 98.7%)	2
qSOFA (2+)	0.51 (0.46, 0.56)	0.07	7.1% (2.4, 15.9%)	99.5% (99.1, 99.7%)	29.4% (10.3, 56.0%)	97.1% (96.4, 97.8%)	1
NEWS	0.6 (0.52, 0.67)	0.07					3
Cut-point of 4+	-	-	31.4% (20.9, 43.6%)	81.9% (80.2, 83.5%)	5.2% (3.3, 7.8%)	97.4% (96.6, 98.1%)	-
Cut-point of 7+	-	-	8.6% (3.2, 17.7%)	98.7% (98.1, 99.1%)	17.1% (6.6, 33.6%)	97.2% (96.4, 97.8%)	-

^a Based on Upper-Left Selected Cut-points from primary validation analysis (C-SHBSL: 9.3%; SHBSL: 5.0%)

^b AUROC: area under the receiver operating characteristic curve, PR-AUC: area under the precision recall curve

^c 95% Confidence Interval estimated using 2000 stratified bootstrapped replicates

^d 95% Confidence Intervals estimated using Clopper Pearson methods

^e Risk-probability cut-point closed to the 01 (upper left) corner of the receiver operating characteristic curve

Table3: Predictive Value of Sepsis Prediction Tools in Relation to High Sepsis-risk Bacteremia Stratified by Patient Location at Time of Culture Collection and in Relation to Short-term Mortality Among Allogeneic HCT Recipients with Potential Bloodstream Infections

	AUROC ^b	Sensitivity ^c	Specificity ^c	Upper-Left Cut-Point ^d
High Sepsis-risk Bacteremia: Patient Location Stratified PBIs				
<i>Inpatient</i>				
Clinical factor-specific-C-SHBSL ^a	0.64 (0.54, 0.74)	97.5% (86.8, 99.9%)	7.6% (5.4, 10.2%)	17.5%
Full Predictor-SHBSL ^a	0.82 (0.76, 0.89)	97.5% (86.8, 99.9%)	29.9% (25.9, 34.1%)	9.1%
SIRS (2+)	0.57 (0.48, 0.66)	67.5% (50.9, 81.4%)	53.0% (48.5, 57.4%)	2
qSOFA (2+)	0.53 (0.46, 0.60)	7.5% (1.6, 20.4%)	98.0% (96.4, 99.0%)	1
NEWS	0.60 (0.50, 0.70)			3
Cut-point of 4+	-	40.0% (24.9, 56.7%)	77.5% (73.6, 81.1%)	
Cut-point of 7+	-	10.0% (2.8, 23.7%)	95.4% (93.2, 97.1%)	
<i>Outpatient</i>				
Clinical factor-specific-C-SHBSL ^a	0.60 (0.50, 0.69)	23.3% (9.9, 42.3%)	84.3% (82.5, 86%)	5.0%
Full Predictor-SHBSL ^a	0.82 (0.75, 0.89)	56.7% (37.4, 74.5%)	85.4% (83.6, 87%)	3.8%
SIRS (2+)	0.57 (0.46, 0.68)	36.7% (19.9, 56.1%)	82.6% (80.7, 84.3%)	2
qSOFA (2+)	0.48 (0.41, 0.55)	6.7% (0.8, 22.1%)	99.9% (99.6, 100%)	1
NEWS	0.56 (0.46, 0.65)			2
Cut-point of 4+	-	20.0% (7.7, 38.6%)	83.2% (81.3, 84.9%)	
Cut-point of 7+	-	6.7% (0.8, 22.1%)	99.6% (99.2, 99.9%)	
Short-term Mortality				
<i>10-Day Mortality</i>				
Clinical factor-specific-C-SHBSL ^a	0.76 (0.68, 0.83)	80.6% (62.5, 92.5%)	66.6% (64.6, 68.5%)	10.0%
Full Predictor-SHBSL ^a	0.85 (0.79, 0.91)	90.3% (74.2, 98.0%)	72.0% (70.1, 73.9%)	5.0%
SIRS (2+)	0.63 (0.53, 0.73)	45.2% (27.3, 64.0%)	75.2% (73.4, 77.0%)	3
qSOFA (2+)	0.64 (0.54, 0.74)	22.6% (9.6, 41.1%)	99.6% (99.2, 99.8%)	1
NEWS	0.71 (0.61, 0.81)			3
Cut-point of 4+	-	41.9% (24.5, 60.9%)	81.8% (80.2, 83.4%)	
Cut-point of 7+	-	16.1% (5.5, 33.7%)	98.7% (98.1, 99.1%)	
<i>28-Day Mortality</i>				
Clinical factor-specific-C-SHBSL ^a	0.71 (0.66, 0.77)	71.1% (60.6, 80.2%)	67.4% (65.4, 69.4%)	9.9%
Full Predictor-SHBSL ^a	0.80 (0.75, 0.84)	78.9% (69.0, 86.8%)	73.2% (71.3, 75.1%)	5.0%
SIRS (2+)	0.65 (0.60, 0.70)	44.4% (34.0, 55.3%)	75.7% (73.9, 77.5%)	2
qSOFA (2+)	0.58 (0.53, 0.63)	8.9% (3.9, 16.8%)	99.6% (99.2, 99.8%)	1
NEWS	0.64 (0.58, 0.70)			3
Cut-point of 4+	-	37.8% (27.8, 48.6%)	82.3% (80.6, 83.9%)	-
Cut-point of 7+	-	8.9% (3.9, 16.8%)	98.8% (98.2, 99.2%)	-

^aBased on Upper-Left Selected Cut-points from primary validation analysis (C-SHBSL: 9.3%; SHBSL: 5.0%)

^b AUROC: area under the receiver operating characteristic curve, 95% Confidence Interval estimated using 2000 stratified bootstrapped replicates

^c 95% Confidence Intervals estimated using Clopper Pearson methods

^dRisk-probability cut-point closed to the 01 (upper left) corner of the receiver operating characteristic curve

Figure Legends

Figure 1: Receiver Operating Characteristic (ROC) curve (A), Precision and Recall curve (B), and Calibration Plots (C) for our novel sepsis prediction tools: clinical factor-specific - C-SHBSL and full predictor - SHBSL and existing tools: Systemic Inflammatory Response Syndrome (SIRS), quick Sequential Organ Failure Assessment (qSOFA), and National Early Warning Score (NEWS) in relation to high sepsis-risk bacteremia among hematopoietic cell transplant (HCT) recipients with potential bloodstream infections. High sepsis-risk bacteremia was defined as gram-negative, *Staphylococcus aureus*, or *Streptococcus* bacteremia.

Figure 2: Receiver Operating Characteristic (ROC) curves for our novel sepsis prediction tools: clinical factor specific - C-SHBSL and full predictor - SHBSL and existing tools: Systemic Inflammatory Response Syndrome (SIRS), quick Sequential Organ Failure Assessment (qSOFA), and National Early Warning Score (NEWS) in relation to high sepsis-risk bacteremia among (A) PBIs identified in the inpatient setting and (B) potential bloodstream infections identified in an outpatient setting and in relation to (C) 10-day mortality and (D) 28-day mortality among all potential bloodstream infections. High sepsis-risk bacteremia was defined as gram-negative, *Staphylococcus aureus*, or *Streptococcus* bacteremia. 10-day and 28-day mortality were defined as death within 10 or 28 days of culture collection.

Figures

Figure 1: Receiver Operating Characteristic Curves, Precision Recall Curve, and Calibration Plot in Relation to High Sepsis-Risk Bacteremia for All Potential Bloodstream Infections Identified Among Adult Allogeneic HCT Recipients

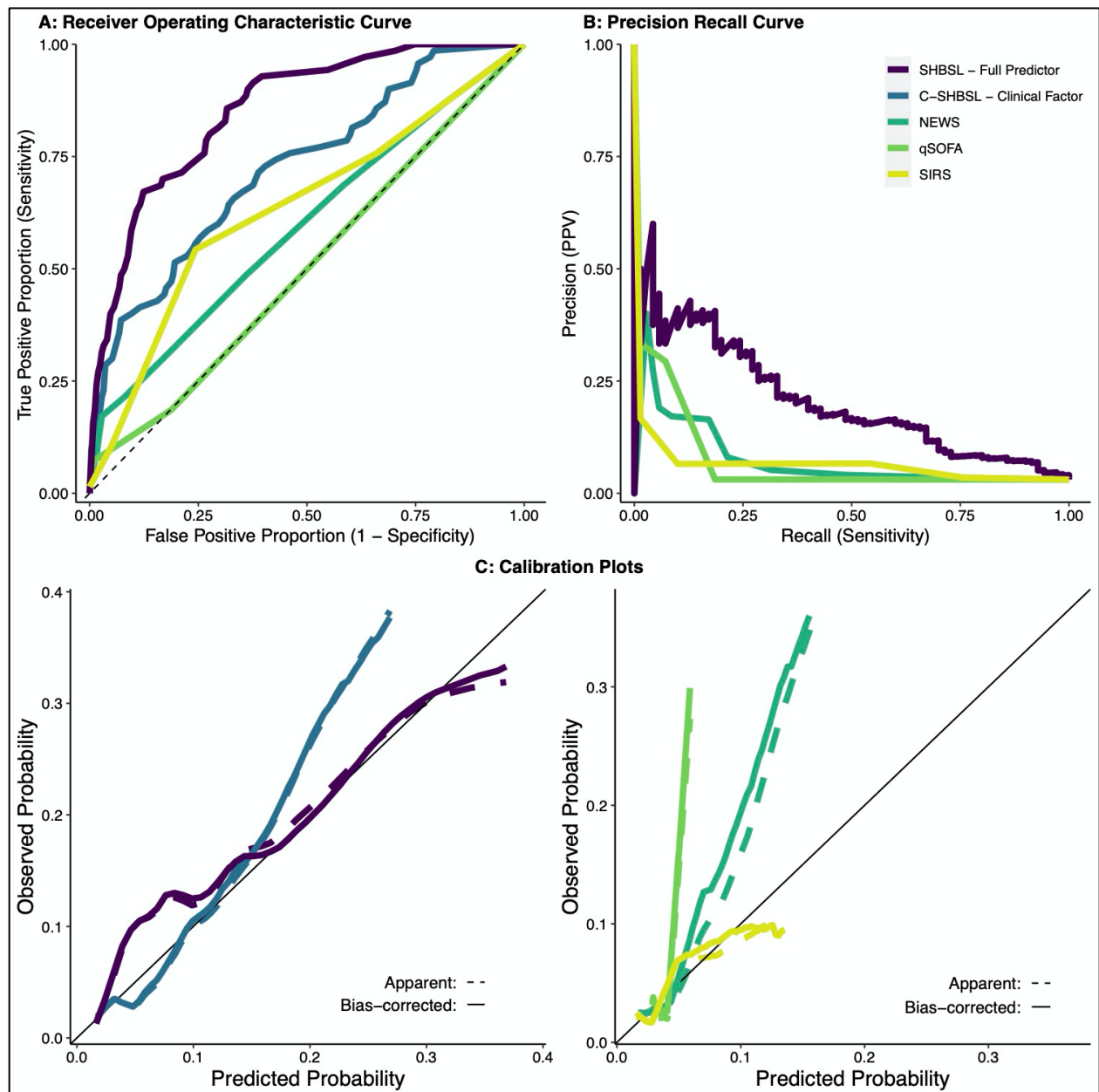
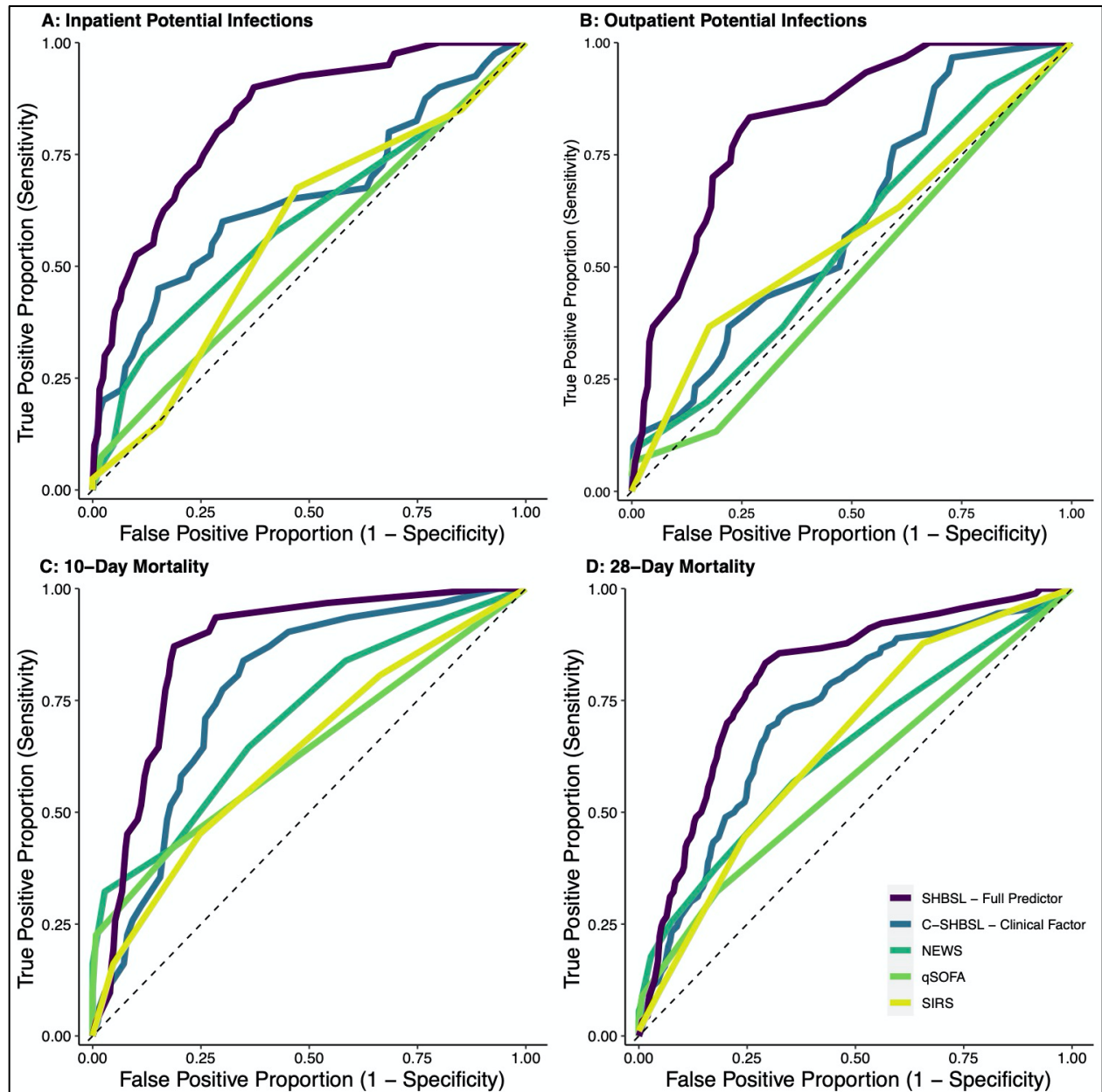


Figure 2: Receiver Operating Characteristic Curves for High Sepsis-Risk Bacteremia Among Inpatient and Outpatient Identified Potential Bloodstream Infections and for 10-Day and 28-Day Mortality Among All Potential Bloodstream



Paper 3: Worth the switch? Modeling the impact of a novel bacterial sepsis prediction tool on mortality among allogeneic HCT recipients

Abstract

Background: Recently developed population-specific sepsis prognostic tools [automated: Super HCT Bacterial Sepsis Learner (SHBSL), Clinical Factor-Specific Super HCT Bacterial Sepsis Learner (C-SHBSL); Bedside: HCT Bacteremia Sepsis Score (HBS2), HCT Bacteremia Sepsis Tree (HBST)] were shown to have greater predictive validity in relation to high-sepsis-risk bacteremia than existing tools (Systemic Inflammatory Response Syndrome-SIRS, quick Sequential Organ Failure System-qSOFA, and National Early Warning Score-NEWS) among allogeneic hematopoietic cell transplant (aHCT) recipients.

Methods: To estimate benefits of tool implementation, we simulated the first 100 days post-transplantation for patients at risk for potential bloodstream infection (PBI; primary analysis) and patients at risk for PBIs identified outside of intensive care units (ICUs, secondary analysis) under set clinical and biological conditions and varied the prognostic tool used. We accounted for prognostic accuracy metric uncertainty through 1000 iterations with random draws from each metric's distribution and summarized the number of deaths and antibiotic days by means and 95% prediction intervals (2.5-97.5 percentile, PI). Through standardization to a base scenario (no tool), we compared the number of prevented deaths by tool specified antibiotics.

Findings: SHBSL led to fewer deaths on average than all examined tools and for all sensitivity/specificity combinations for all but HBS2. Compared with HBS2, SHBSL led to 1 (PI: -1, 4) fewer deaths for the primary and 1 (PI: 0, 3) fewer death for the secondary analysis. SHBSL resulted in more efficient antibiotic use (2000 days of tool specified antibiotic use prevented more deaths) than all but qSOFA for both analyses. HBS2 and HBST resulted in more efficient antibiotic use than SIRS and NEWS for the secondary analysis.

Conclusion: These results suggest that, compared with other examined tools, SHBSL led to fewer simulated deaths with relatively few additional days of antibiotics. Within data or computationally limiting settings, HBS2 and HBST provided the best alternative to SHBSL.

Introduction

Following transplantation, allogeneic hematopoietic cell transplant (HCT) recipients are at high risk of developing and dying of bacterial sepsis.[1,2] Because sepsis commonly presents atypically during this period, commonly used existing sepsis prognostic tools, which rely on symptom-based criteria developed among general patient populations, perform poorly.[1,3–5] With the goal of providing additional decision support at the time of blood culture collection, Lind et al. developed two population-specific automated bacterial sepsis decision support tools for allogeneic HCT recipients with potential bloodstream infections (PBIs) and two population-specific bedside bacterial sepsis decision support tools for recipients with PBIs identified outside of intensive care units (ICUs).(Paper 1, 2)

Through the validation of these tools, Lind et al. found that, in relation to high-risk bacteremia (culture confirmed gram-negative, *Streptococcus* species, or *Staphylococcus aureus* bacteremia) the population-specific tools had better predictive validity among all allogeneic HCT recipients with PBIs than existing prognostic tools (Systemic Inflammatory Response Syndrome [SIRS], quick Sequential Organ Failure Assessment [qSOFA], and National Early Warning Score [NEWS]). Specifically, the area under the receiver-operator characteristic curve (AUC) of SIRS – the existing tool with the highest AUC – was 21 percentage points lower than the Super HCT Bacterial Sepsis Learner (SHBSL - automated), eight lower than Clinical Factor-specific Super HCT Bacterial Sepsis Learner (C-SHBSL - automated), 11 lower than that of HCT Bacteremia Sepsis Score (HBS2 - bedside), and six lower than that of HCT Bacteremia Sepsis Tree (HBST - bedside).(Paper 1, 2) Similarly, among allogeneic HCT recipients with PBIs identified outside the ICU (the bedside tools' target population), HBS2 and HBST had AUCs 13 and 10 percentage points higher than that of SIRS. Regarding clinical practicality, each of these population-specific tools provided cut-points with higher sensitivity relative to their positive

predictive value and specificity than the examined existing tools.(Paper 1, 2)

While Lind et al. previously demonstrated that the population-specific tools had superior predictive validity when compared with three commonly used existing tools, whether their superior validity justifies their implementation into clinical settings is unknown. To examine this question, we compared the number of simulated deaths and antibiotic use by prognostic tool under fixed clinical and biological conditions using a first-order Monte Carlo Markov state-transition model. Through analyses varying the examined biological and clinical conditions, we explored the range of clinical and biological conditions for which our conclusions held.

Materials and Methods

Model Development

We simulated the first 100 days post-HCT for adult allogeneic HCT recipients using a first-order Monte Carlo Markov state-transition model with fixed biologically and clinically plausible parameters. Figure 1 is a detailed schematic of the model. Briefly, the simulation begins with a cohort of HCT recipients on day of transplantation (all considered at risk for high-risk bacteremia) and follows the cohort to day 100 post-transplantation. Throughout the simulation, patients move between at-risk, infected with high-risk bacteremia, and not infected but treated (with antibiotic) states based on transition probabilities derived from retrospectively collected data and previously published literature.[6](Paper 1, 2) Each state is exclusive and while infected or on antibiotic treatment, patients were not considered at risk for an infection. However, once recovered from the infection or no longer on treatment, patients returned to the at-risk state. These transitions rely on the model simplifying assumptions that (1) patients could not develop new infections while currently infected, (2) antibiotics were appropriate (no resistance) and prevent infection, and (3) there was no lingering infection related risk associated with previous infections (*Supplement: Modeling Assumptions*).

For each patient, the simulation terminated at day 100 post transplantation or death, whichever occurred first. Mortality was dictated by the current infection and treatment state of the patient and, once on antibiotics or recovered from an infection, no lingering increased mortality risk was incorporated (*Supplement: Modeling Assumptions*). At the time of culture collection, antibiotic use was specified by the sensitivity and specificity of the prognostic tool and antibiotic use was considered to continue through a) the day negative culture results returned, b) infection termination, or c) death (*Supplement: Modeling Assumptions*).

Cohort (Simulated Population)

To account for differences in the target populations of the population-specific and existing tools, we ran two variations of the model. In our primary exploration, we compared the mortality and antibiotic use resulting from the employment of different sepsis prognostic tools among any simulated patient with a blood culture collection. Because the target setting for the population-specific bedside tools was locations with varied support structures (ie. outside the ICU), we, secondarily, compared the frequency of these outcomes among patients with PBIs identified outside the ICU. Specifically, we ran the presented model with culture, culture result and tool accuracy parameters estimated among allogeneic HCT recipients located and PBIs identified outside of ICUs but assumed the remaining parameters and model structure matched that of the primary analysis.

Parameters

We estimated the sample size as well as the daily culture collection and high-risk bacteremia positive proportions using retrospectively collected data previously described by Lind et al.(Paper 1, 2) Briefly, data from the first 100 days following transplantation for 1943 adult (≥ 18 years) allogeneic HCT recipients transplanted at Seattle Cancer Care Alliance / Fred Hutchinson Cancer Research Center between June 2010 and 2019 were retrospectively collected from a

prospectively maintained transplant database, electronic medical records, and laboratory records. Of the 1943 observed HCT recipients, 1594 experience at least one PBI and 1563 experienced at least one PBI outside the ICU. In total, 8131 PBIs were observed, 7253 of which occurred outside the ICU. Two hundred and thirty-eight of the 8131 PBIs (3.2%) were defined as high-risk bacteremia and 174 of the 7253 non-ICU PBIs (2.4%) were defined as high-risk bacteremia.

Based on the size of the observed sample ($N = 1943$), we set the sample size to be 2000 HCT recipients. Using the observed data, we estimated the daily culture collection proportion using Poisson regressions with natural splines for time in days since transplantation and an offset term of patient time. Similarly, we estimated the daily high-risk bacteremia proportion among all and non-ICU specific PBIs using Poisson regressions with natural splines for time since transplant and an offset term of collected cultures (Figure 2 A/C). The daily culture and culture positive estimates were then defined as the day-specific point estimate predicted from the appropriate Poisson regression. These parameters were estimated and incorporated into the modeling process under the assumption that previous infection or treatment does not impact risk of future infection or treatment (*Supplement: Model Assumptions*).

The remaining parameter estimates were derived from previously published literature and center practices over the time of observed data collection and can be found in Table 1. The infection duration and time to culture result were defined as 14 and 3 days, respectively, based on antibiotic treatment practices within our center over the span of data collection and published estimates of bacterial growth times.[7–10]

Tool accuracy metrics (sensitivity and specificity) were taken from Lind et al.'s previously published validation of the examined tools.(Paper 1, 2) For the primary analysis, we included the sensitivity and specificity for SIRS/qSOFA 2+ Ref. [11], NEWS 4+ Ref. [12],

HBS2 2+, HBST 3.3% (Paper 1, 2) and for the automated tools we selected the probability cut-points with specificity closest to that of SIRS 2+ among all allogeneic HCT recipients with PBIs (76%; SHBSL: 5.5%, C-SHBSL: 11.1%; Table 1).(Paper 1, 2) For the secondary analysis (non-ICU PBIs), we included the sensitivity and specificity for SIRS/qSOFA 2+ Ref. [11], NEWS 4+ Ref. [12] and the probability cut-point with the specificity closest to that of SIRS 2+ among allogeneic HCT recipients with PBIs identified outside the ICU (76%; HBS2: 2+, HBST: 3.3%). Because the automated tools have not been validated among allogeneic HCT recipients with PBIs identified outside the ICU, we chose the previously published estimates for recipients with PBIs identified in outpatient settings (approximated from Lind et al. (CITE AUTO) Figure 2: receiver operating characteristic curve).(Paper 2) While this cohort is exclusive of non-ICU inpatient identified PBIs and the accuracy metrics resulting from the non-ICU specific PBI cohort may differ slightly from the outpatient PBIs, we concluded that the accuracy metrics from this cohort were likely the best surrogate for the non-ICU PBI metrics. To account for uncertainty in the tools' prognostic accuracy metric estimates (sensitivity and specificity), we ran 1000 iterations selecting sensitivity and specificity at random from the assumed normal distribution of each metric.

Daily mortality proportions were estimated based on survival results from published literature of HCT recipients in the following three scenarios: bloodstream infection naïve, infected with appropriate antibiotics, and infected without appropriate antibiotics.[13,14] For each of these scenarios, we identified periods of approximately linear survival proportion decline and calculated mortality constants for those periods as:

$$\text{Equation 1: } \sum_{d=0}^D (x / (1 - d * x) * (100 - d * 100)) - z * 100$$

Where x is the mortality constant, d is the number of days since event (transplant or culture

collection and D is final day), and z is the mortality probability for a given time period.

If we observed multiple distinct linear periods for a given group (as we did with true positive and false negatives), we calculated constants for both periods and adjusted the equation for the second period to account for the reduced population:

$$\text{Equation 2: } q + \sum_{d=0}^D (x / (1 - d * x)) * (100 - d * 100) - z * 100$$

Where q is the total death count from the prior linear period.

The daily mortality proportions were then calculated using the appropriate mortality constant and the following equation.

$$\text{Equation 3: } x / (1 - d * x)$$

Where d (day) equaled zero for day since transplant or infection onset (culture collection)

The published point value for uninfected individuals was a year-long mortality probability of 35%.[13] The 7-day mortality probability for infected and treated individuals was 16% and for infected and untreated individuals, it was 45%.[14,15] The 30-day mortality probability from which the constant for the remaining infection time was calculated was 21% for infected and treated individuals and 64% for infected and untreated individuals (Table 1).[14,15] We developed our model in R 4.0.2 using base functions and data management packages. This work was approved by the FHCRC Institutional Review Board.

Sensitivity Analyses – Varying non-Tool accuracy parameters

To test the robustness of our results to varying biological and clinical parameter settings, we performed numerous sensitivity analyses setting the non-tool accuracy parameters to a wide array of parameter estimates ranging from highly plausible to implausible. Specifically, we examined time to culture results of 2, 5, and 10 days; infection durations of 5, 7, 30, 50, 100 days; and infected and untreated mortality proportions based on published data suggesting that

the odds of mortality increase by 1.03 (CI: 1.00, 10.5) times each hour antibiotics are delayed.[16] For the varying mortality risk analyses, we set the time zero (day of culture collection) mortality proportion for the infected and untreated group equal to that of the infected and treated group, estimated the hourly mortality proportion using odds ratios (ORs) of 1, 1.03, and 1.05, and because our model runs on day intervals, set each days mortality proportion as the median hourly proportion for that day since culture collection.

Along with sensitivity analyses looking at different mortality proportions, infection durations and culture result times, we recognize that clinicians' treatment related actions may not align with tool recommendations. We explored such disagreements between tool recommendation and clinical action in two ways – 1) assuming that the clinicians and tools lack of alignment was random and 2) assuming that the clinician knows more than the tool and correctly reclassifies patients. For the first scenario (random disagreement between clinical action and tool recommendation), we reclassified 10, 30, 50, 70, and 100% of each classification group (ie. 10% true positives reclassified false negatives). For the scenario in which disagreement stems from clinical recognition of inappropriate tool-based classification, we ran simulations where correct reclassification occurred among 5, 10, 15, 20% for the infected and uninfected populations (ie. 5% of false negatives reclassified to true positives). Each of these analyses were run with an infection duration of 14 days, a culture result time of 3 days, and in the primary analysis cohort (patients with any PBIs).

Model Verification

We performed model checks to ensure that our constructed equations (1-3) and code were performing as desired. To test the mortality equations, we ran a single iteration under the following scenarios: all uninfected with perfect prediction (all true negatives), all uninfected with perfectly imperfect prediction (all false positives), all infected with perfect prediction (all true

positives), and all infected with perfectly imperfect prediction (all false negatives). Additionally, we verified that our antibiotic days were accumulating as desired. As with the mortality verification, we looked at antibiotic accumulation by tool prediction group. All checks were run on a sample of 1000 patients, a 3-day culture return, and a 14-day infection duration.

As desired, the uninfected individuals (true negative and false positives) had matching linear cumulative mortality growth and infected individuals had step like cumulative mortality growth. For the model version with all true positives, we observed a steep linear growth in cumulative mortality for the first seven days of an infection cycle followed by 7 days (day 8-14 of infection) of a shallower linear growth. The model version with all false negatives had three linear growth sections per infection cycle: the steepest being the untreated time between culture collection (day 0) and day of culture results (day 3), followed by 4 days of slightly shallower cumulative mortality growth (within 7 days of infection onset but now treated), and then the shallow linear growth for the remaining 7 days of a cycle. Over the 100 days of follow-up the steps became shorter as the baseline population declined (e.Figure 1).

Regarding antibiotic use, we found that no use occurred when all patients were true negatives and saw a linear decline in the number of antibiotic days matching the slope of the linear growth in cumulative mortality for the false positive group. Similarly, the daily antibiotic decline matched the slope and curvature of the cumulative mortality curve for the model with all true positive individuals. From the model version with all false negatives, we found that the first 3 days are antibiotic free (period between culture collection and result) followed by 11 days of antibiotic use matching the inverse of the cumulative mortality curve (e.Figure 2).

Evaluation

For each prognostic tool, we summarized the cumulative mortality and antibiotic use estimates from the 1000 iterations using median and 95% prediction intervals (2.5 and 97.5% percentiles).

Daily events were first averaged over all days for a given tool/iteration combination and then summarized over the 1000 iterations using the median and 95% prediction intervals. We visually compared the median daily deaths and antibiotic use estimates by infection status using jittered line graphs with 95% percentile error bars.

In order to compare the mortality / antibiotic use trade-off between each examined tool, we compared the estimated total and daily deaths for each tool against those estimated in the absence of a prognostic tool (ie. 100% specificity and 0% sensitivity; referred to as base scenario). From this, we compared the estimated number of prevented deaths per 2000 antibiotic use days beyond the base scenario (Equation 4) by prognostic tool. Moving forward, we refer to the 2000 antibiotic use days beyond the base scenario as “2000 tool specified antibiotic days” as they represent the antibiotic days that occurred due to the use of a given tool. We selected 2000 antibiotic days as it represents one antibiotic day per simulated HCT recipient.

Equation 4: $(\text{Base scenario deaths} - \text{tool model deaths}) / (\text{Tool model antibiotic days} - \text{Base model antibiotic days}) * 2000$

For each tool, we plotted the difference in median mortality and antibiotic outcome estimates between the tool and the base scenario simulation using a two-group lollipop plot with median base scenario estimates set as a reference (zero). All evaluations were conducted, and figures generated in R 4.0.2 using the ggplot2, GGally, and hrbrthemes packages.

Observed Mortality Proportion

Additionally, as a supplement to this work, we estimated the mortality proportion by each tool’s prediction category (tested positive or negative) and correctness (false or true prediction) in relation to high sepsis risk bacteremia (referred to as high-risk bacteremia) among an observed cohort of allogeneic HCT recipients with potential infections (*Supplement: Observed Mortality Proportions*). However, due to the frequency of antibiotic use around the time of culture

collection, these estimates do not reflect the impact tool related treatment would have on mortality and we provided them solely as observational data.

Results

Primary Analysis

Mirroring the observed trends in high-risk bacteremia incidence (Figure 2.A), daily deaths and daily deaths among infected individuals peaked in the first 20 days following transplantation (Figure 3.A). However, during this period, the daily deaths among uninfected individuals dipped (Figure 3.C). All tools led to fewer deaths and the population-specific tools led to fewer deaths than existing tools (Figure 3.A). Under all sensitivity and specificity combinations, SHBSL led to fewer deaths than C-SHBSL, HBST and the examined existing tools. Specifically, SHBSL led to 3 [95% Prediction Interval (PI): 1, 5] fewer cumulative deaths than C-SHBSL, 3 (PI: 1, 5) fewer than HBST, and 3 (PI: 1, 5) fewer deaths than SIRS – the existing tool resulting in the fewest deaths (Table 2). SHBSL did not result in fewer deaths than HBS2 under all examined sensitivity and specificity combinations. Compared with HBS2, SHBSL led to between 1 more to 4 fewer deaths.

qSOFA led to the fewest number of total antibiotic days (1297, PI: 1219, 1388) and HBS2 led to the most (8497, PI: 8040, 8946). SHBSL led to the greatest number of antibiotic days among individuals with high-risk bacteremia (1561, PI: 1506, 1627) and HBS2 led to the greatest number among patients without high-risk bacteremia (6992 PI: 6533, 7429; Table 3). Relative to the base scenario (sensitivity 0%, specificity 100%), qSOFA prevented the most deaths per 2000 tool specified antibiotic days (14, PI: 5, 24) followed by SHBSL (4, PI: 3, 5; Table 4, Figure 5).

Secondary Analysis

We observed similar trends in daily deaths among our simulated cohort of patients at risk for

PBIs outside ICUs as we observed in the primary analysis (Figure 4.A). However, the peak in deaths around day 75 was visibly lower among patients at risk for non-ICU PBIs than among patients at risk for any PBIs (Figures 3/4, A) As with the primary analysis, each of the population-specific tools led to fewer cumulative deaths than existing tools and all tools led to fewer deaths than the base scenario (Figure 4.A). SHBSL led to the fewest deaths among all examined tools (213, PI: 212, 214), followed by HBS2 (214, PI: 213, 215) and HBST (215, PI: 214, 216; Table 2).

HBS2 led to the greatest number of total antibiotic days 7304 (PI: 6891, 7702) and qSOFA led to the fewest 880 (PI: 828, 938). SHBSL led to the greatest number of antibiotic days among infected individuals 1133 (PI: 1099, 1175) followed by HSB2 1093 (PI: 1049, 1126). As with the primary analysis, qSOFA prevented the greatest number of deaths per 2000 tool specified antibiotic days (10; PI: 1, 22) and SHBSL, HBST, and HBS2 all prevented 3. HBS2 and HBST prevented between 0 and 2 more deaths per 2000 tool specified antibiotic days than NEWS or SIRS (Table 4, Figure 6).

Sensitivity Analyses

In the sensitivity analyses, the timing of mortality and antibiotic use peaks mirrored the primary analysis, but the magnitude of these peaks varied. Model versions with longer infection durations, days until culture results, and greater differences in the mortality risk between treated and untreated infected individuals (larger ORs) had higher mortality and antibiotic use peaks (e.Figures 3-14). Compared with smaller percentages, larger random disagreement percentages between treatment and tool recommendations had larger peaks and larger correct reclassification percentages led to smaller peaks among all and infected individuals.

SHBSL led to the fewest deaths under the majority of examined scenarios. As in the primary analysis, SHBSL did not lead to fewer deaths than HBS2 for all sensitivity and

specificity combinations but did lead to fewer deaths on average under all examined scenarios except when treatment was not considered associated with mortality (OR 1) and when random disagreement between tool classification and treatment exceeded 50%. Under these scenarios, HBS2 led to fewer deaths for all examined sensitivity and specificity combinations (e.Tables 9, 13) and when random disagreement between treatment and tool recommendation exceeded 50%, SHBSL led to more deaths than all of the other examined tools (e.Table 9).

Discussion

Using a Monte Carlo Markov state-transition model we examined if the simulated implementation of recently developed population-specific sepsis prognostic tools (HBST, HBS2, SHBSL, and C-SHBSL) would lead to meaningfully fewer deaths and better antibiotic efficiency than the continued use of SIRS, NEWS, or qSOFA among allogeneic HCT recipients under biologically and clinically plausible scenarios. We then varied the fixed clinical and biological parameters to explore the range of scenarios under which are results held and through the estimation of a base scenario in which no prognostic tool was used (i.e. sensitivity of 0%, specificity of 100%), we compared the mortality related impact of tool specific antibiotic use.

Under the fixed parameters and assumptions of the primary and secondary scenarios, the population-specific tools, namely SHBSL, prevented more deaths with little to no additional antibiotic use per simulated patient than the examined existing tools. While we found qSOFA prevented the most deaths per 2000 days of tool specified antibiotic use (Table 4), the lollipop charts displaying the difference in the total number of deaths and antibiotic days per simulated patient between the tool specific scenarios and the base scenario [Figures 5 (Primary), 6 (Secondary)], show that, qSOFA prevented very few deaths (<2.5). Conversely, 2000 days of SHBSL specified antibiotic use prevented the second greatest number of deaths (Table 4) and the use of SHBSL resulted in the fewest number of deaths overall (Table 2). Highlighting these results, the lollipop charts show that, SHBSL prevented visibly more

deaths than SIRS, NEWS, and qSOFA and did so with few additional days of antibiotic use per simulated patient (<3 for qSOFA, <2 NEWS, ~ 0 SIRS). This suggests that the use of SHBSL would result in fewer deaths among patients with high-risk bacteremia than existing tools without resulting in excessive additional antibiotic use among patients without high-risk bacteremia. The fact that these results held under all plausible clinical and biological parameter scenarios (only under scenarios in which antibiotics did not reduce the risk of mortality among patients with high-risk bacteremia or the random disagreement between tool recommendation and treatment exceeded 50% did SHBSL result in more deaths than other examined tools) supports the replacement of existing tools with SHBSL.

Compared with the other population-specific tools (C-SHBSL, HBST, HBS2), SHBSL led to fewer deaths and prevented more deaths per 2000 tool specified antibiotic days. Specifically, among the primary cohort, HBS2 and HBST led to 1 more death per 2000 days of tool specified antibiotic use. Though the differences in number of deaths between SHBSL and the population-specific bedside tools were smaller among the secondary cohort (the cohort the bedside tools were built to target) than in the primary cohort, SHBSL still resulted in fewer deaths and prevented more deaths per tool specified antibiotic day than either bedside tool. While these results highlight the superior nature of SHBSL and support the use of this automated tool in all scenarios, SHBSL requires extensive computational resources and relies on data that may not be available in ambulatory or emergency settings.

In such data- or computationally-limited settings, the existing tools and the population-specific bedside tools provide clinically practical alternatives to SHBSL. When comparing these tools in our data-limited cohort (the secondary, non-ICU PBI cohort), we find that HBS2 and HBST resulted in better mortality to antibiotic use tradeoffs than existing tools. For example, compared with SIRS, the existing tool that led to the fewest number of deaths, HBS2 led to 2 fewer deaths and prevented 0 more deaths per 2000 tool specified antibiotic days on average.

Similarly, HBST, on average, led to 1 less death and prevented 1 more death per 2000 tool specified antibiotic days than SIRS. This indicates that, HBS2 and HBST resulted in fewer deaths than SIRS and did so with similarly precise antibiotic use (HBS2) and more precise antibiotic use (HBST).

While this does not negate the superior performance of SHBSL, it suggests that, under scenarios in which SHBSL based predictions are impossible, HBS2 or HBST provide the best alternatives to SHBSL.

Strengths and Limitations

Our study had multiple limitations. First, our model relied on fixed biological and clinical parameters. To combat this limitation, we examined alternative parameters to test the robustness of our findings and found that under all biologically plausible scenarios, SHBSL led to fewer deaths with limited additional antibiotic initiation among uninfected individuals than other examined tools. In fact, only under scenarios when mortality among treated and untreated infected patients is equal (OR 1) or when random disagreement between treatment and tool recommendation exceeds 50% did SHBSL lead to more deaths than other examined tools (e.Tables 9; 13). However, we were unable to run all scenarios and our results are biased to the examined fixed parameters. Second, our model relied upon multiple model assumptions that likely do not reflect real clinical scenarios (for a full list of modeling assumptions see *Supplement: Modeling Assumptions*). These assumptions increased the interpretability of our results and model simplicity but limit the value of the individual mortality and antibiotic use estimates – as they may not reflect clinical reality. Instead, since the simulation of each tool's employment were subject to the same assumptions and fixed parameters, the value of our results lies in their comparison (as presented in the discussion). Third, our analysis relied upon prognostic accuracy metrics calculated within similar cohorts as were used in the population-specific tool development (the held-out validation component of the same population). Though

we incorporated the uncertainty of the prognostic accuracy metrics in this analysis and Lind et al. showed, through numerous sensitivity analyses, that their prognostic findings were robust to varying data collection and cleaning assumptions, the true prognostic accuracy metrics for each examined tool are unknown and estimates from external data would strengthen our understanding of each tool's utility.

Our study also had many strengths. First, our culture collection and infection proportion estimates were calibrated to ten years of follow-up for a large cohort of allogeneic HCT recipients from a leading transplant center. Second, through the development of our Monte Carlo Markov state-transition model, we were able to compare the mortality and targeted antibiotic use resulting from different sepsis prognostic tools. Due to the frequency of antibiotic use among our population at the time of or shortly following culture collection, such comparison is impossible among retrospectively collected data. Third, by comparing each tool to a base scenario in which no tool was used, we were able to estimate the number of prevented deaths relative to tool specified antibiotic use. Four, through the incorporation of prognostic accuracy metric uncertainty through 1000 iterations with random samples from each accuracy metric distribution, we produced prediction intervals reflective of the varying combinations of the accuracy metrics.

Conclusion

Under all examined plausible biological and clinical parameter combinations, we found that the use of SHBSL would likely result in fewer deaths relative to targeted antibiotic use – especially among uninfected individuals - than SIRS, NEWS, or qSOFA. Further, we found that, under scenarios with limited data or computational ability (ie. outside of ICUs), HBST and HBS2 provided the best alternatives to SHBSL. These results support the integration of SHBSL into EMR systems where possible and HBS2 or HBST when automated prediction is impossible.

References

1. Dadwal S, Eftekhari Z, Tushondra T, Johnson D. A Machine-Learning Based Prediction Model for Sepsis in Patients Undergoing Hematopoietic Stem Cell Transplantation. ASH Annual Meeting. 2018;the Systemic Inflammatory Response Syndrome (SIRS) and quick-Sequential Organ Failure Assessment (qSOFA) criteria. Available: <https://ash.confex.com/ash/2018/webprogram/Paper117002.html>
2. Kumar G, Ahmad S, Taneja A, Patel J, Guddati AK, Nanchal R, et al. Severe sepsis in hematopoietic stem cell transplant recipients*. *Crit Care Med.* 2015;43: 411–421. doi:10.1097/CCM.0000000000000714
3. Kochanek M, Schalk E, von Bergwelt-Baildon M, Beutel G, Buchheidt D, Hentrich M, et al. Management of sepsis in neutropenic cancer patients: 2018 guidelines from the Infectious Diseases Working Party (AGIHO) and Intensive Care Working Party (iCHOP) of the German Society of Hematology and Medical Oncology (DGHO). *Ann Hematol.* 2019;98: 1051–1069. doi:10.1007/s00277-019-03622-0
4. Lind ML, Phipps AI, Mooney S, Liu C, Fohner A, Patel K, et al. Predictive value of three clinical criteria for sepsis (qSOFA, SIRS, and NEWS) with respect to short-term mortality in allogeneic hematopoietic cell transplant recipients with suspected infections. *Clin Infect Dis.* 2020. doi:10.1093/cid/ciaa214
5. Pastores SM, Papadopoulos E, van den Brink M, Alicea M, Halpern NA. Septic shock and multiple organ failure after hematopoietic stem cell transplantation: treatment with recombinant human activated protein C. *Bone Marrow Transplant.* 2002;30: 131–134. doi:10.1038/sj.bmt.1703618
6. Thomas D, Hilgsmann M, John D, Al Ahdab OG, Li H. Chapter 18 - Pharmacoeconomic Analyses and Modeling. In: Thomas D, editor. *Clinical Pharmacy Education, Practice and Research.* Elsevier; 2019. pp. 261–275. doi:10.1016/B978-0-12-814276-9.00018-0
7. Tabak YP, Vankeepuram L, Ye G, Jeffers K, Gupta V, Murray PR. Blood Culture Turnaround Time in U.S. Acute Care Hospitals and Implications for Laboratory Process Optimization. *Journal of Clinical Microbiology.* 2018;56. doi:10.1128/JCM.00500-18
8. Kim TJ, Weinstein MP. Update on blood cultures: how to obtain, process, report, and interpret. *Clinical Microbiology and Infection.* 2013;19: 513–520. doi:10.1111/1469-0691.12180
9. Lind ML, Roncaioli S, Phipps A, Liu C, Bryant A, Sweet A, et al. Trends in the Proportion of Time Spent Outpatient While Receiving Antibiotics for Treatment of Gram-Negative Rod Bacteremia Among Allogeneic Hematopoietic Cell Transplant Recipients. *Biology of Blood and Marrow Transplantation.* 2020;26: S320–S321. doi:10.1016/j.bbmt.2019.12.343
10. Olearo F, Kronig I, Masouridi-Levrat S, Chalandon Y, Khanna N, Passweg J, et al. Optimal Treatment Duration of *Pseudomonas aeruginosa* Infections in Allogeneic Hematopoietic Cell Transplant Recipients. *Open Forum Infectious Diseases.* 2020;7. doi:10.1093/ofid/ofaa246
11. Seymour CW, Liu VX, Iwashyna TJ, Brunkhorst FM, Rea TD, Scherag A, et al. Assessment of Clinical Criteria for Sepsis: For the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA.* 2016;315: 762–774. doi:10.1001/jama.2016.0288
12. Royal College of Physicians. National Early Warning Score (NEWS) Standardising the assessment of acute-illness severity in the NHS. London; 2012.
13. Ustun C, Young J-AH, Papanicolaou GA, Kim S, Ahn KW, Chen M, et al. Bacterial blood stream infections (BSIs), particularly post-engraftment BSIs, are associated with increased

- mortality after allogeneic hematopoietic cell transplantation. *Bone Marrow Transplant*. 2019;54: 1254–1265. doi:10.1038/s41409-018-0401-4
14. Mikulska M, Del Bono V, Bruzzi P, Raiola AM, Gualandi F, Van Lint MT, et al. Mortality after bloodstream infections in allogeneic haematopoietic stem cell transplant (HSCT) recipients. *Infection*. 2012;40: 271–278. doi:10.1007/s15010-011-0229-y
 15. Kang C-I, Kim S-H, Park WB, Lee K-D, Kim H-B, Kim E-C, et al. Bloodstream Infections Caused by Antibiotic-Resistant Gram-Negative Bacilli: Risk Factors for Mortality and Impact of Inappropriate Initial Antimicrobial Therapy on Outcome. *Antimicrob Agents Chemother*. 2005;49: 760–766. doi:10.1128/AAC.49.2.760-766.2005
 16. Seymour CW, Gesten F, Prescott HC, Friedrich ME, Iwashyna TJ, Phillips GS, et al. Time to Treatment and Mortality during Mandated Emergency Care for Sepsis. *New England Journal of Medicine*. 2017;376: 2235–2244. doi:10.1056/NEJMoa1703058

Tables

Table 1: Semi-Stochastic equation-based Model Parameters

	Varied by Iteration	All Allogeneic HCT Recipients (Primary Analysis)		Non-ICU Allogeneic HCT Recipients (Secondary Analysis)	
		Point Value	Standard Deviation	Point Value	Standard Deviation
Sample Size	No	2000	-	2000	-
Duration of infection	No	5 ^a	-	5 ^a	-
		7 ^a	-	7 ^a	-
		14	-	14	-
Time to Culture Results	No	2 ^a	-	2 ^a	-
		3	-	3	-
Total Percent Dead	Yes				
Uninfected (Year)		35.0%	0.7%	35.0%	0.7%
Infected Treated					
0-7 Days		16.0%	5.9%	16.0%	5.9%
8-30 Days		21.0%	6.6%	21.0%	6.6%
Infected Not Treated					
0-7 Days		45.0%	15.0%	45.0%	15.0%
8-30 Days		64.0%	14.5%	64.0%	14.5%
All PBI Accuracy Metrics					
Sensitivities	Yes				
HBS2		65.7%	5.5%	72.4%	5.2%
HBST		54.3%	5.8%	62.1%	5.6%
SHBSL		72.9%	5.2%	80.0%	4.6%
C-SHBSL		57.1%	5.8%	40.0%	5.7%
SIRS		54.3%	5.8%	51.7%	5.8%
qSOFA		7.1%	3.0%	3.4%	2.1%
NEWS		31.4%	5.4%	29.3%	5.3%
Specificities	Yes				
HBS2		70.0%	1.0%	70.6%	0.9%
HBST		77.2%	0.9%	78.3%	0.9%
SHBSL		75.8%	0.9%	76.0%	0.9%
C-SHBSL		74.7%	0.9%	74.0%	0.9%
SIRS		75.9%	0.9%	76.2%	0.9%
qSOFA		99.5%	0.2%	99.7%	0.1%
NEWS		81.9%	0.8%	80.6%	0.8%
Daily Culture Collection Rate					
	Yes	Variable by day		Variable by day	
Daily High-Risk Bacteremia Proportion					
	Yes	Variable by day		Variable by day	

Systemic Inflammatory Response Syndrome (SIRS), quick Sequential Organ Failure Assessment (qSOFA), National Early Warning Score (NEWS) Super HCT Bacterial Sepsis Learner (SHBSL), Clinical Factor-specific Super HCT Bacterial Sepsis Learner (C-SHBSL), HCT Bacteremia Sepsis Score (HBS2), HCT Bacteremia Sepsis Tree (HBST)

^a Secondary analysis parameter

^b Non-ICU estimates: Estimated from ROC curve of outpatient identified PBIs (Figure 2)

^c Taken from Paper One Supplement (e.Table 11)

Table 2: Simulated Mortality and Mortality Difference For 2000 Allogeneic HCT Recipients with Potential Bloodstream Infections

			Mortality Difference with Prediction Interval			
	Count	Prediction Intervals ^a	SHBSL Reference	C-SHBSL Reference	HBS2 Reference	HBST Reference
All Potential Bloodstream Infections (Primary Analysis)						
<i>Cumulative Deaths</i>						
SHBSL 5.5%+	222	(221, 224)	-	-3 (-5, -1)	-1 (-4, 1)	-3 (-5, -1)
C-SHBSL 11.1%+	225	(223, 227)	3 (1, 5)	-	1 (-1, 4)	0 (-3, 2)
HBS2 2+	223	(222, 225)	1 (-1, 4)	-1 (-4, 1)	-	-2 (-4, 0)
HBST 3.3%+	225	(224, 227)	3 (1, 5)	0 (-2, 3)	2 (0, 4)	-
SIRS 2+	225	(224, 227)	3 (1, 5)	0 (-2, 3)	2 (-1, 4)	0 (-2, 3)
qSOFA 2+	233	(232, 234)	11 (9, 13)	8 (6, 10)	10 (8, 11)	8 (6, 10)
NEWS 4+	229	(228, 231)	7 (5, 9)	4 (2, 7)	6 (3, 7)	4 (2, 6)
Base Scenario ^b	234	(234, 234)	12 (11, 14)	9 (8, 11)	11 (9, 12)	9 (8, 11)
Non-ICU Potential Bloodstream Infections (Secondary Analysis)						
<i>Cumulative Deaths</i>						
SHBSL ^c	213	(212, 214)	0 (0, 0)	-4 (-6, -3)	-1 (-3, 0)	-2 (-3, -1)
C-SHBSL ^c	217	(216, 219)	4 (3, 6)	0 (0, 0)	3 (2, 5)	2 (1, 4)
HBS2 2+	214	(213, 215)	1 (0, 3)	-3 (-5, -2)	0 (0, 0)	-1 (-3, 0)
HBST 3.3%+	215	(214, 216)	2 (1, 3)	-2 (-4, -1)	1 (0, 3)	0 (0, 0)
SIRS 2+	216	(215, 217)	3 (2, 4)	-1 (-3, 1)	2 (1, 4)	1 (0, 3)
qSOFA 2+	222	(221, 222)	9 (8, 10)	4 (3, 5)	8 (7, 9)	7 (5, 8)
NEWS 4+	219	(218, 220)	6 (4, 7)	1 (0, 3)	5 (3, 6)	4 (2, 5)
Base Scenario ^b	222	(222, 222)	9 (8, 10)	5 (4, 6)	8 (7, 9)	7 (6, 8)

Systemic Inflammatory Response Syndrome (SIRS), quick Sequential Organ Failure Assessment (qSOFA), National Early Warning Score (NEWS) Super HCT Bacterial Sepsis Learner (SHBSL), Clinical Factor-specific Super HCT Bacterial Sepsis Learner (C-SHBSL), HCT Bacteremia Sepsis Score (HBS2), HCT Bacteremia Sepsis Tree (HBST)

^a 2.5-97.5% Percentiles for Mortality and Antibiotic Days

^b Base scenario: simulation run without a prognostic tool (sensitivity: 0%, specificity 100%)

^c Accuracy metrics estimated from Lind et al. 2021 JAMA Open Figure 2

Table 3: Simulated Antibiotic Use For 2000 Allogeneic HCT Recipients

	Count	Cumulative Prediction Intervals ^a	Count	Average Daily Prediction Intervals ^a
All Potential Bloodstream Infections (Primary Analysis)				
<i>Antibiotic Days</i>				
SHBSL 5.5%+	7160	(6794, 7597)	71	(44, 89)
C-SHBSL 11.1%+	7341	(6936, 7703)	73	(45, 92)
HBS2 2+	8497	(8040, 8946)	84	(52, 107)
HBST 3.3%+	6734	(6336, 7074)	67	(41, 84)
SIRS 2+	7017	(6605, 7449)	69	(43, 88)
qSOFA 2+	1297	(1219, 1388)	13	(1, 33)
NEWS 4+	5503	(5163, 5917)	54	(33, 68)
Base Scenario ^b	1138	(1138, 1138)	11	(0, 31)
<i>Antibiotic Days - Infected Individuals</i>				
SHBSL 5.5%+	1561	(1506, 1627)	15	(2, 38)
C-SHBSL 11.1%+	1457	(1380, 1533)	14	(2, 36)
HBS2 2+	1504	(1438, 1554)	15	(2, 37)
HBST 3.3%+	1452	(1393, 1514)	14	(2, 36)
SIRS 2+	1444	(1366, 1507)	14	(2, 36)
qSOFA 2+	1182	(1150, 1218)	12	(0, 32)
NEWS 4+	1306	(1242, 1368)	13	(1, 34)
Base Scenario ^b	1138	(1138, 1138)	11	(0, 31)
<i>Antibiotic Days - Uninfected Individuals</i>				
SHBSL 5.5%+	5599	(5205, 6033)	55	(23, 76)
C-SHBSL 11.1%+	5883	(5495, 6298)	58	(24, 79)
HBS2 2+	6992	(6533, 7429)	69	(29, 94)
HBST 3.3%+	5282	(4889, 5662)	52	(22, 72)
SIRS 2+	5573	(5165, 5999)	55	(23, 76)
qSOFA 2+	115	(52, 187)	1	(0, 2)
NEWS 4+	4196	(3895, 4534)	42	(17, 57)
Base Scenario ^b	0	(0, 0)	0	(0, 0)
Non-ICU Potential Bloodstream Infections (Secondary Analysis)				
<i>Antibiotic Days</i>				
SHBSL ^c	6161	(5826, 6556)	61	(35, 76)
C-SHBSL ^c	6425	(6059, 6769)	64	(36, 80)
HBS2 2+	7304	(6891, 7702)	72	(41, 91)
HBST 3.3%+	5616	(5264, 5921)	56	(32, 69)
SIRS 2+	5993	(5622, 6383)	59	(34, 75)
qSOFA 2+	880	(828, 938)	9	(1, 31)
NEWS 4+	4981	(4678, 5342)	49	(28, 62)
Base Scenario ^b	804	(804, 804)	8	(0, 30)
<i>Antibiotic Days - Infected Individuals</i>				
SHBSL ^c	1133	(1099, 1175)	11	(2, 37)
C-SHBSL ^c	957	(903, 1010)	9	(1, 33)
HBS2 2+	1093	(1049, 1126)	11	(1, 36)
HBST 3.3%+	1060	(1020, 1103)	10	(1, 35)
SIRS 2+	1010	(955, 1054)	10	(1, 34)
qSOFA 2+	819	(804, 836)	8	(0, 31)
NEWS 4+	914	(870, 957)	9	(1, 32)

Table 3: Simulated Antibiotic Use For 2000 Allogeneic HCT Recipients (Continued)

	Count	Cumulative Prediction Intervals ^a	Count	Average Daily Prediction Intervals ^a
Base Scenario ^b	804	(804, 804)	8	(0, 30)
<i>Antibiotic Days - Uninfected Individuals</i>				
SHBSL ^c	5028	(4671, 5421)	50	(21, 68)
C-SHBSL ^c	5468	(5113, 5848)	54	(23, 73)
HBS2 2+	6211	(5796, 6606)	61	(27, 83)
HBST 3.3%+	4555	(4204, 4895)	45	(19, 62)
SIRS 2+	4983	(4615, 5368)	49	(21, 68)
qSOFA 2+	62	(18, 112)	1	(0, 1)
NEWS 4+	4067	(3787, 4380)	40	(17, 55)
Base Scenario ^b	0	(0, 0)	0	(0, 0)

Systemic Inflammatory Response Syndrome (SIRS), quick Sequential Organ Failure Assessment (qSOFA), National Early Warning Score (NEWS) Super HCT Bacterial Sepsis Learner (SHBSL), Clinical Factor-specific Super HCT Bacterial Sepsis Learner (C-SHBSL), HCT Bacteremia Sepsis Score (HBS2), HCT Bacteremia Sepsis Tree (HBST)

^a 2.5-97.5% Percentiles for Mortality and Antibiotic Days

^b Base scenario: simulation run without a prognostic tool (sensitivity: 0%, specificity 100%)

^c Accuracy metrics estimated from Lind et al. 2021 JAMA Open Figure 2

Table 4: Prevented Deaths per 2000 Antibiotic Days Beyond the Base Scenario for Simulated Allogeneic HCT Recipient Population

Difference in Prevented Deaths per 2000 Additional Antibiotic Day						
	Prevented Deaths	Prediction Intervals ^a	SHBSL Reference	C-SHBSL Reference	HBS2 Reference	HBST Reference
All Potential Bloodstream Infections (Primary Analysis)						
SHBSL 5.5%+	4	(3, 5)	-	-1 (-2, 0)	-1 (-2, 0)	-1 (-1, 0)
C-SHBSL 11.1%+	3	(2, 4)	1 (0, 2)	-	0 (-1, 1)	0 (-1, 1)
HBS2 2+	3	(2, 3)	1 (0, 2)	0 (-1, 1)	-	0 (0, 1)
HBST 3.3%	3	(3, 4)	1 (0, 1)	0 (-1, 1)	0 (-1, 0)	-
SIRS 2+	3	(2, 4)	1 (0, 2)	0 (-1, 1)	0 (-1, 0)	0 (-1, 1)
qSOFA 2+	14	(5, 24)	-10 (-19, -2)	-11 (-21, -3)	-11 (-21, -3)	-11 (-20, -2)
NEWS 4+	2	(2, 3)	2 (1, 2)	1 (0, 2)	1 (0, 1)	1 (0, 2)
Non-ICU Potential Bloodstream Infections (Secondary Analysis)						
SHBSL ^b	3	(3, 4)	-	-2 (-2, -1)	-1 (-1, 0)	0 (-1, 0)
C-SHBSL ^b	2	(1, 2)	2 (1, 2)	-	1 (0, 1)	1 (1, 2)
HBS2 2+	3	(2, 3)	1 (0, 1)	-1 (-1, 0)	-	0 (0, 1)
HBST 3.3%+	3	(2, 3)	0 (0, 1)	-1 (-2, -1)	0 (-1, 0)	-
SIRS 2+	2	(2, 3)	1 (1, 2)	-1 (-1, 0)	0 (0, 1)	1 (0, 1)
qSOFA 2+	10	(1, 22)	-7 (-18, 3)	-9 (-20, 1)	-8 (-19, 2)	-7 (-19, 3)
NEWS 4+	2	(1, 2)	2 (1, 2)	0 (0, 1)	1 (0, 1)	1 (1, 2)

Systemic Inflammatory Response Syndrome (SIRS), quick Sequential Organ Failure Assessment (qSOFA), National Early Warning Score (NEWS) Super HCT Bacterial Sepsis Learner (SHBSL), Clinical Factor-specific Super HCT Bacterial Sepsis Learner (C-SHBSL), HCT Bacteremia Sepsis Score (HBS2), HCT Bacteremia Sepsis Tree (HBST)

^a Prevented deaths per 2000 additional days of antibiotics beyond the base scenario; 2.5-97.5% Percentiles for Mortality and Antibiotic Days

^b Base scenario: simulation run without a prognostic tool (sensitivity: 0%, specificity 100%)

^c Accuracy metrics estimated from Lind et al. 2021 JAMA Open Figure 2

Figure Legends

Figure 1: Schematic outlining the stochastic equation-based model used to estimate mortality and antibiotic use resulting from the use of a prognostic tool for bacterial sepsis among a simulated cohort of allogeneic hematopoietic stem cell transplant recipients followed for their first 100-days post transplantation.

Figure 2: Bar graphs of the observed blood culture and high-risk bacteremia daily events by day since transplant over the estimated number of events resulting from the Poisson regression with a natural spline for time. Figures A and B display the observed and estimated number of blood cultures (A) and bacteremia events (B) among all observed patient time (Primary Cohort). Figures C and D display the blood culture (C) and bacteremia events (D) among patients outside of intensive care units (ICUs; Secondary Cohort).

Figure 3: Daily mortality (A), daily mortality among infected patients (B), daily mortality among uninfected patients (C), daily antibiotic use (D), daily antibiotic use among infected patients (E), daily antibiotic use among uninfected patients (F) results by sepsis prognostic tool (SIRS: Systemic Inflammatory Response Syndrome, qSOFA: quick Sequential Organ Failure Assessment, NEWS: National Early Warning Score, SHBSL: Super HCT Bacterial Sepsis Learner, C-SHBSL: Clinical Factor-specific Super HCT Bacterial Sepsis Learner, HBS2: HCT Bacteremia Sepsis Score, HBST: HCT Bacteremia Sepsis Tree) from the stochastic equation-based model for patients at risk for any potential bloodstream infections (primary cohort), infection durations of 14 days, and 3-day culture result wait times.

Figure 4: Daily mortality (A), daily mortality among infected patients (B), daily mortality among uninfected patients (C), daily antibiotic use (D), daily antibiotic use among infected patients (E), daily antibiotic use among uninfected patients (F) results by sepsis prognostic tool (SIRS: Systemic Inflammatory Response Syndrome, qSOFA: quick Sequential Organ Failure Assessment, NEWS: National Early Warning Score, SHBSL: Super HCT Bacterial Sepsis Learner, C-SHBSL: Clinical Factor-specific Super HCT Bacterial Sepsis Learner, HBS2: HCT Bacteremia Sepsis Score, HBST: HCT Bacteremia Sepsis Tree) from the stochastic equation-based model for patients at risk for potential bloodstream infections identified outside of intensive care units (secondary cohort), infection durations of 14 days, and 3-day culture result wait times.

Figure 5: Faceted (A) and Overlapping (B) lollipop charts depicting the difference in mortality and antibiotic use between a prognostic tool (SIRS: Systemic Inflammatory Response Syndrome, qSOFA: quick Sequential Organ Failure Assessment, NEWS: National Early Warning Score, SHBSL: Super HCT Bacterial Sepsis Learner, C-SHBSL: Clinical Factor-specific Super HCT Bacterial Sepsis Learner, HBS2: HCT Bacteremia Sepsis Score, HBST: HCT Bacteremia Sepsis Tree) and the base scenario (no tool – sensitivity 0%, specificity 100%) for simulated allogeneic HCT recipients at risk for any potential bloodstream infections (primary cohort), infection durations of 14 days, and 3-day culture result wait times. The charts are standardized to base scenario event count of zero and antibiotic use is presented in days per simulated patients (antibiotic days / 2000 simulated patients).

Figure 6: Faceted (A) and Overlapping (B) lollipop charts depicting the difference in mortality

and antibiotic use between a prognostic tool (SIRS: Systemic Inflammatory Response Syndrome, qSOFA: quick Sequential Organ Failure Assessment, NEWS: National Early Warning Score, SHBSL: Super HCT Bacterial Sepsis Learner, C-SHBSL: Clinical Factor-specific Super HCT Bacterial Sepsis Learner, HBS2: HCT Bacteremia Sepsis Score, HBST: HCT Bacteremia Sepsis Tree) and the base scenario (no tool – sensitivity 0%, specificity 100%) for simulated allogeneic HCT recipients at risk for potential bloodstream infections identified outside of intensive care units (secondary cohort), infection durations of 14 days, and 3-day culture result wait times. The charts are standardized to base scenario event count of zero and antibiotic use is presented in days per simulated patients (antibiotic days / 2000 simulated patients).

Figure 1: Schematic of Bacterial Sepsis Stochastic Equation-Based Model

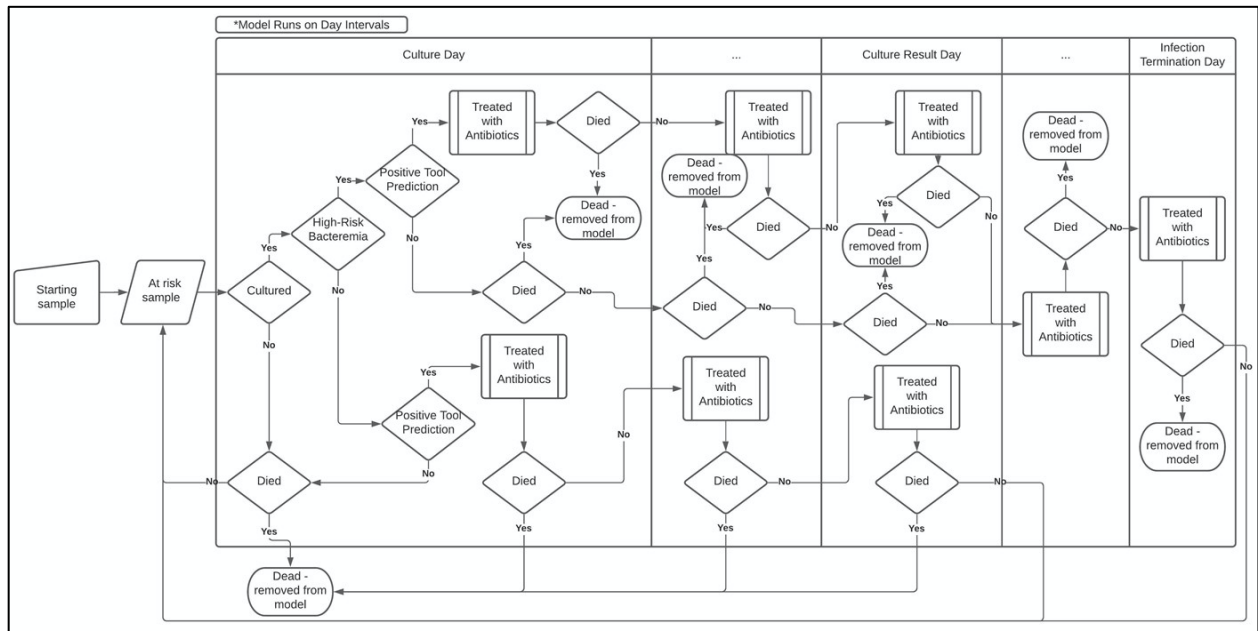


Figure 2: Estimated Blood Culture and High-risk Bacteremia Daily Events by Day Since Transplant - Shown as events per 100 units (patients or cultures)

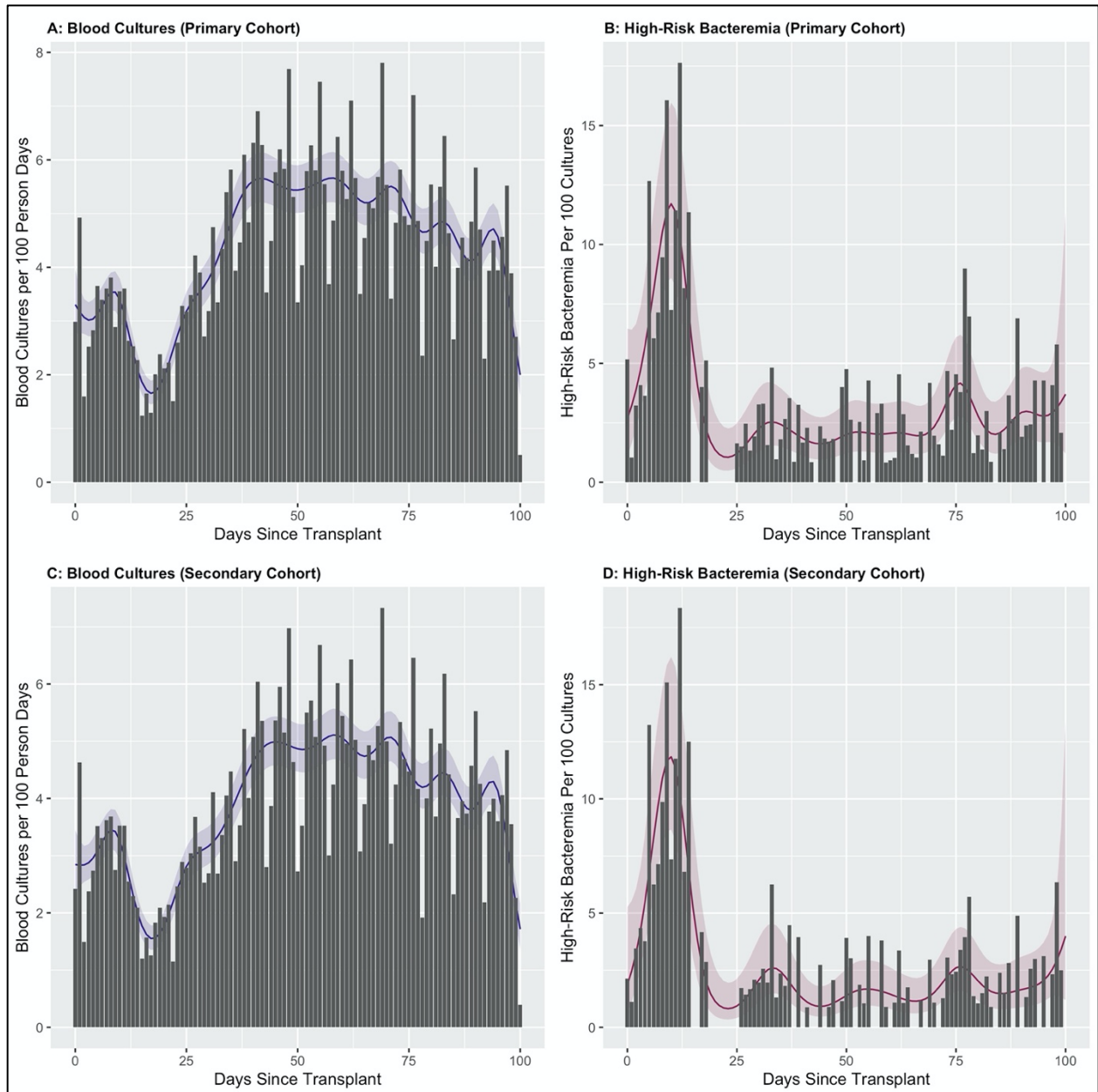


Figure 3: Stochastic equation-based Model Results for Patients at risk for Potential Bloodstream Infections (Primary Analysis Cohort)
 - Time to culture results: 3 days; Infection duration: 14 days

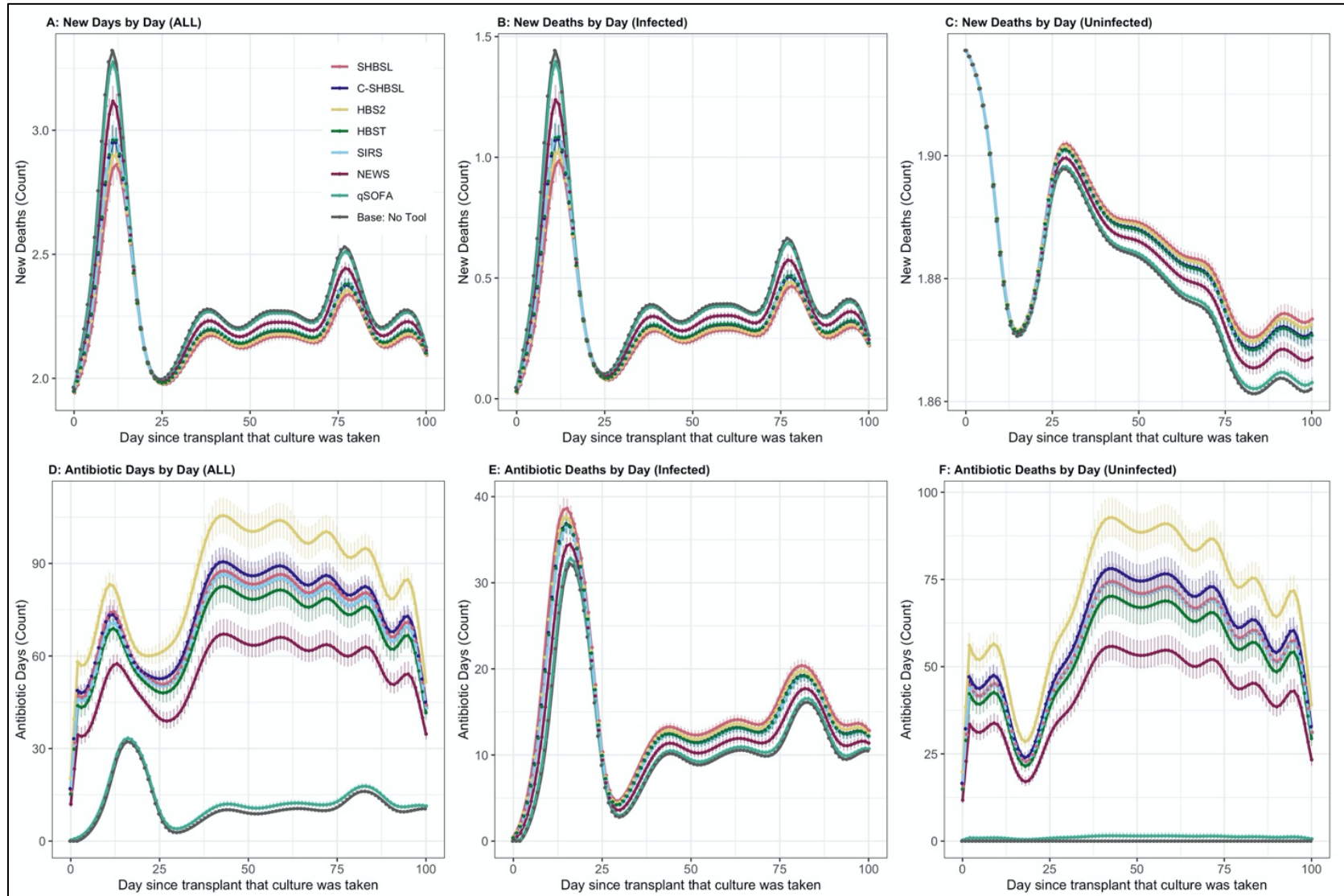


Figure 4: Stochastic equation-based Model Results for Patients at risk for Potential Bloodstream Infections Identified Outside of the ICU (Secondary Analysis Cohort) - Time to culture results: 3 days; Infection duration: 14 days

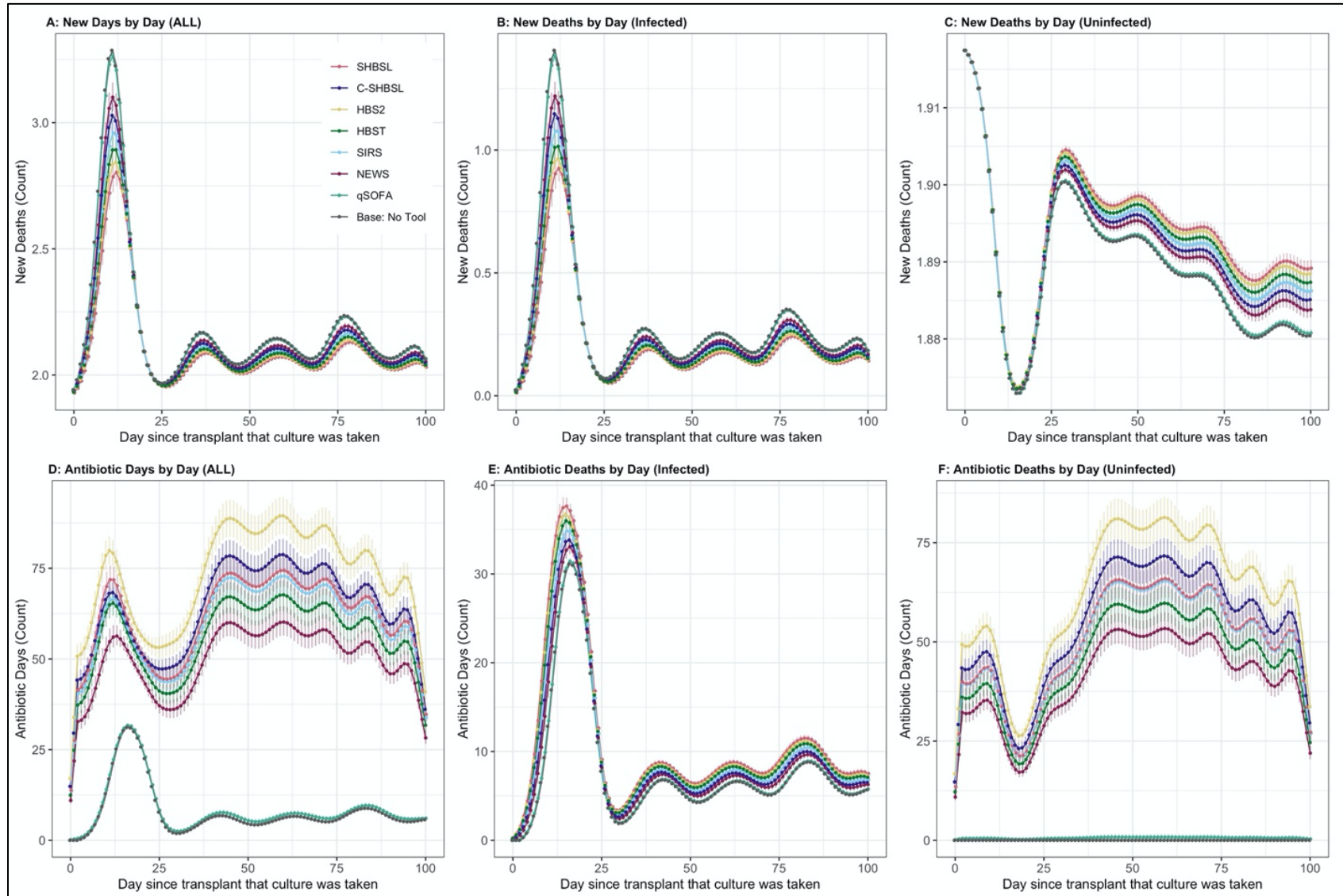


Figure 5: Cumulative mortality and antibiotic use difference between tool employed and base scenario simulation for patients with any potential bloodstream infections (Primary Analysis Cohort) - Standardized to base scenario event count of zero | Antibiotic use in days per 2000 patients | Time to culture results: 3 days; Infection duration: 14 days

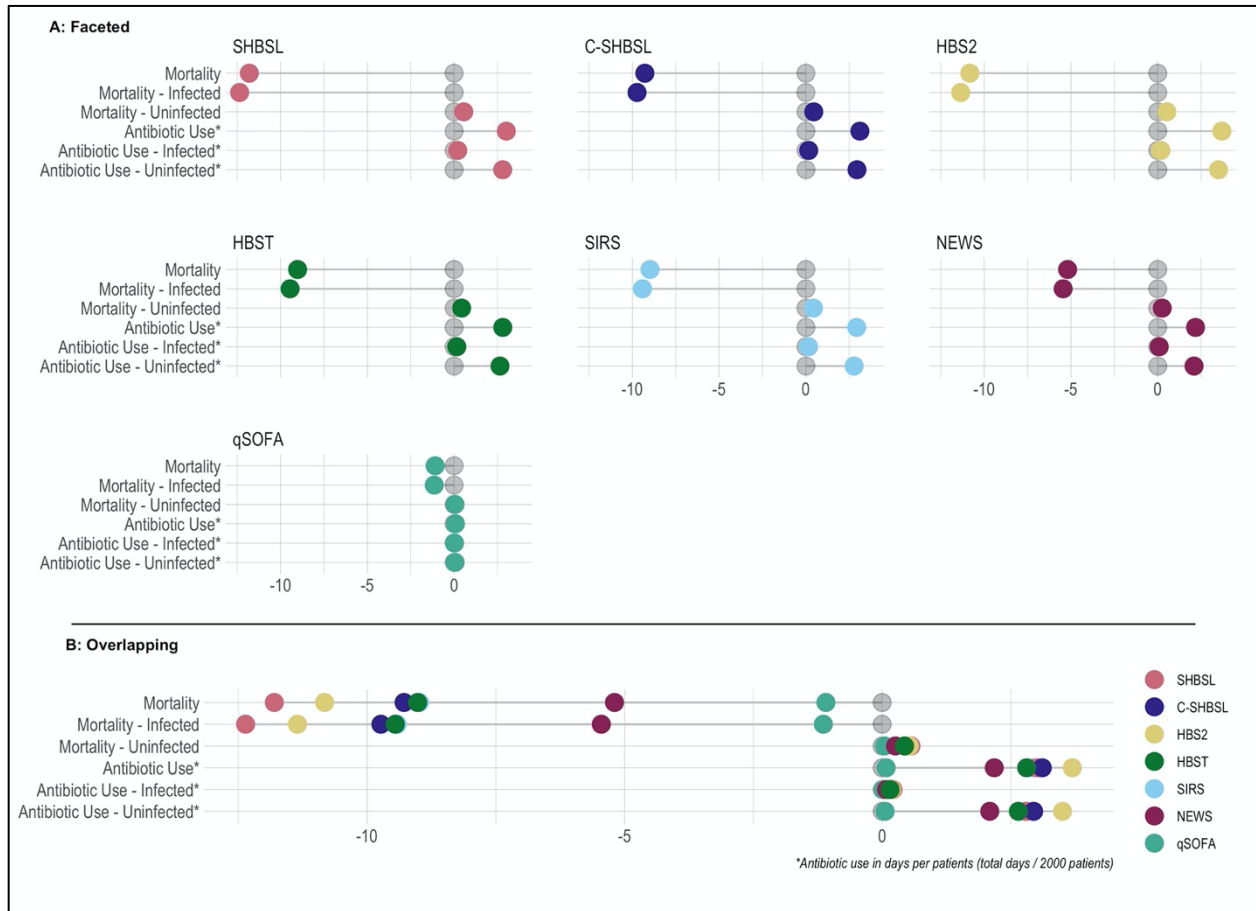
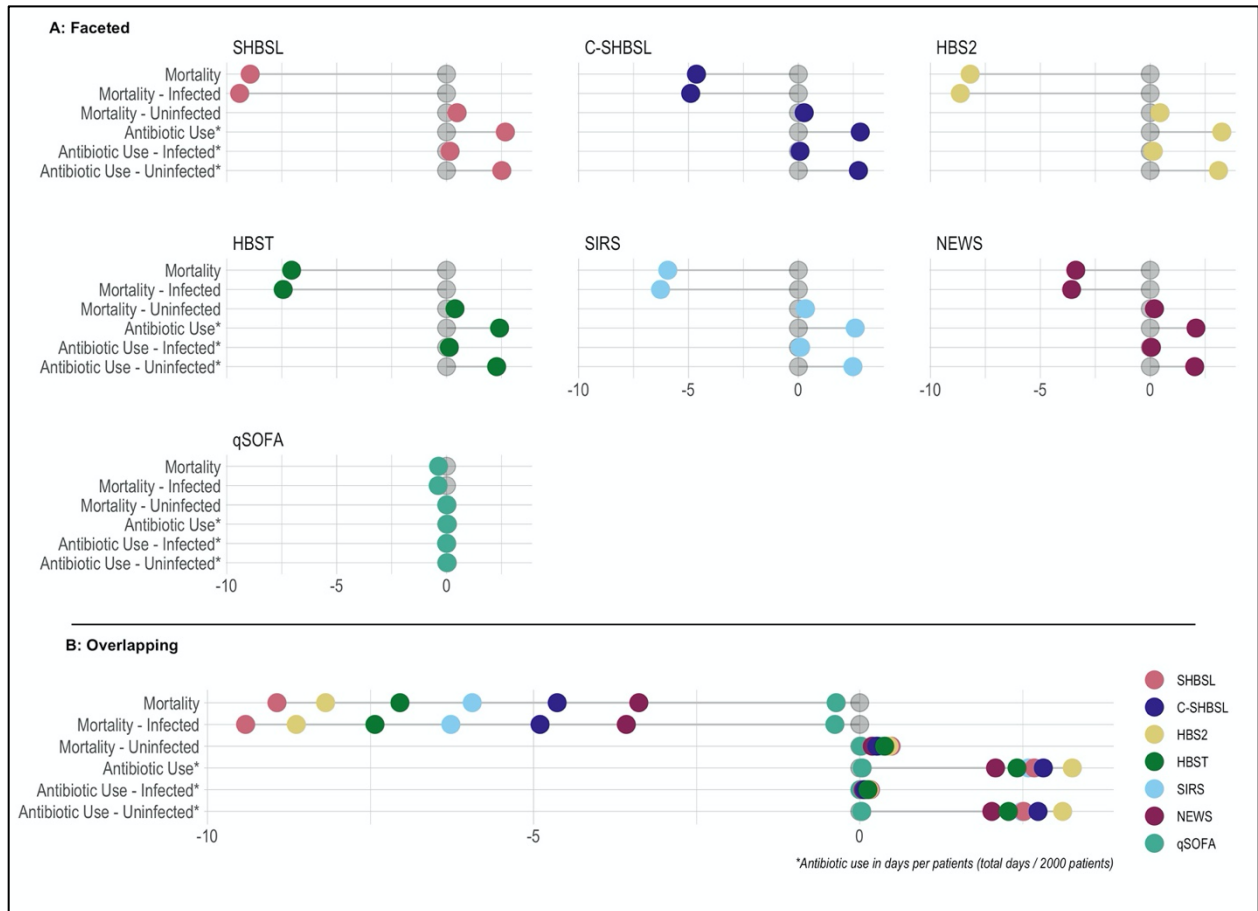


Figure 6: Cumulative mortality and antibiotic use difference between tool employed and base scenario simulation for patients with potential bloodstream infections identified outside of the ICU (Secondary Analysis Cohort) - Standardized to base scenario event count of zero | Antibiotic use in days per 2000 patients; Time to culture results: 3 days; Infection duration: 14 days



Supplement Material

Supplemental Paper 1: Immunocompromised cancer patient specific rapid clinical tools for bacteremia with high risk for sepsis

Missing Data

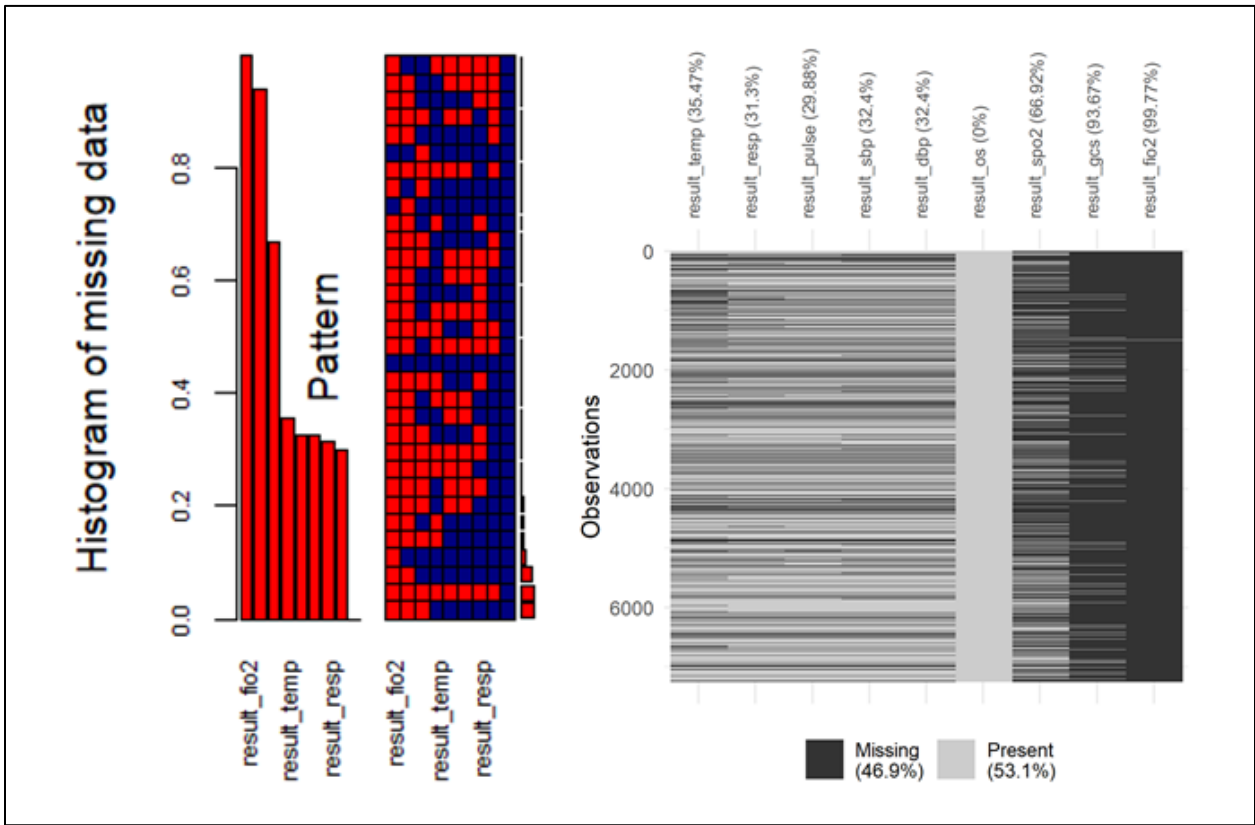
A limitation of using electronic medical record (EMR) data in development and evaluation of prediction tools, as we did in this analysis, is the large amount of structurally ambiguous data missingness. eTable 1 shows the amount of missingness for each examined predictor and summarizes their observed values.

e.Table 1: Frequency of Demographic and Clinical Categories Screened for Model Inclusion
(For all potential bloodstream infections N = 7253)

	Missingness	Range	Median / Count	IQR / Percent
Age	0, 0%	18, 81	53	40, 62
Diastolic Blood Pressure	2355, 32%	40, 134	79	72, 86
Gender (Count, % male)	0, 0%	Female, Male	4202	57.8%
Mental Function Status (Glasgow Coma Score)	6809, 94%	4, 15	15	15, 15
Oxygen Saturation (SpO2)	4860, 67%	84, 100	97	95, 98
Heart Rate	2170, 30%	11, 167	90	79, 103
Respiratory Rate	2275, 31%	6, 56	16	16, 18
Systolic Blood Pressure	2355, 32%	76, 206	128	116, 140
Temperature	2578, 35%	34.6, 40.1	36.9	36.6, 37.3
Ethnicity/Race (Count, %)	0, 0%			
American Indian or Alaskan Native ^b	-	-	138	1.90%
Asian ^b	-	-	467	6.40%
Black or African America ^b	-	-	126	1.70%
Multiple ^b	-	-	132	1.80%
Native Hawaiian or other Pacific Islander ^b	-	-	100	1.40%
Unknown ^b	-	-	194	2.70%
White ^b	-	-	6115	84.10%

^b Non-Hispanic

e.Figure 1: Missingness Structure of Examined Prediction



We handled this missingness in two ways. First, we restricted which predictors we examined to those with relatively frequent measurement ($> 65\%$ present) during the 24 hours prior to and 2 hours following culture collection. Second, we performed single, normal value imputation using publicized normal value ranges (eTable 2). Normal value ranges are provided for examination factors we included in our model development and any additional examination or laboratory factors included in the existing criteria we examined. Each missing value was imputed with a randomly selected value from the normal range using a uniform distribution.

e.Table 2: Normal Value Ranges for Single, Normal Value Imputation

	Normal Value Range	Source
Temperature	36.1 - 37.2°C	MedlinePlus
Respiratory Rate	12-20 breaths/minute	Cleveland Clinic
Pulse (Heart Rate)	60-100 beats/minute	Mayo Clinic
Systolic Blood Pressure	90-120 mmHg	Havard, Tolulope et al.
Diastolic Blood Pressure	60-80 mmHg	Stensrud et al.
Oxygen Saturation	95-100%	Mayo Clinic
		Blood Groups and Red Cell Antigens, Bethesda
White Blood Cell Count	4500 - 1100 cells/mm ³	et al.
Glasgow Coma Scale Score	15	MdCalc

Validation versus Modeling Dataset

We randomly divided the study sample into a modeling (70%) and validation (30%) groups at the individual level. The modeling and validation datasets were comprised of the PBIs for the individuals selected into the corresponding groups.

eTable 3: Population Demographics and Transplant Factors by Randomly Assigned Modeling and Validation Dataset

	Modeling Dataset	Validation Dataset
Sample Count	N = 1094	N = 469
Calendar Year at Transplant (Median, IQR)	2014 (2012, 2016)	2014 (2012, 2016)
Age at Transplant (Median, IQR)	54 (42, 62)	55 (43, 63)
Gender (Count, % Male)	613 (56.0)	276 (58.8)
Ethnicity/Race (Count, %)		
American Indian or Alaskan Native	18 (1.6)	8 (1.7)
Asian	82 (7.5)	34 (7.2)
Black	17 (1.6)	10 (2.1)
Multiple	13 (1.2)	10 (2.1)
Native Hawaiian or other Pacific Islander	19 (1.7)	6 (1.3)
Unknown	45 (4.1)	13 (2.8)
White	900 (82.3)	388 (82.7)
HCT Cell Donor Type		
Bone marrow	95 (8.7)	29 (6.2)
Bone marrow, peripheral blood stem cells	2 (0.2)	1 (0.2)
Umbilical cord blood	122 (11.2)	52 (11.1)
Peripheral blood stem cells	875 (80.0)	387 (82.5)
Donor Relation (Count, % unrelated)		
Child	26 (2.4)	10 (2.1)
Half Sibling	3 (0.3)	2 (0.4)
Not Related	766 (70.0)	319 (68.0)
Parent	14 (1.3)	3 (0.6)
Relative	1 (0.1)	0 (0.0)
Sibling	284 (26.0)	135 (28.8)
Acute GVHD (Grade > 1) (Count, %) ^a	820 (75.0)	346 (73.8)
Died (Count, %) ^a	86 (7.9)	44 (9.4)

Detailed Methods

Practicality dictated that our bedside tools contain a restricted number of easily captured demographic and measurable clinical examination factor categories (score) or splits (decision tree). However, we did not know the impact of additional flexibility (complexity) on model performance and thus were unable to define the optimal flexibility *a priori*. Instead, we chose to evaluate the predictive impact of added flexibility by performing best subset selection for the decision tree and forward stepwise selection starting with two predictors for the score.¹⁰

Regardless of the predictor inclusion technique (best subset or forward stepwise), the best model of a specific size (amount of flexibility) was identified as the model with the highest 10-fold cross-validated area under the receiver operator curve (cv-AUC).¹⁵ The cv-AUCs from the best model of each size were then plotted against the model size and the performances compared. The optimal model was selected based on performance relative to its practicality (size) and, using the optimal model parameters, the final model was developed in the full modeling dataset.

Score (HBS2) Development

The score was developed using a weighted integer logistic regression method.¹¹ To ensure the bedside functionality of our score and ensure computational practicality, we evaluated predictors in binary form. For each continuous predictor, we developed binary predictor variants at even steps along the upslope of the smoothed curve of sepsis probability by that predictor. The smoothed sepsis probabilities were estimated using single predictor logistic regressions with cubic splines. The size of the steps varied by predictor and were determined based on the width of the sepsis probability upslope and the common digit of measurement for the predictor.

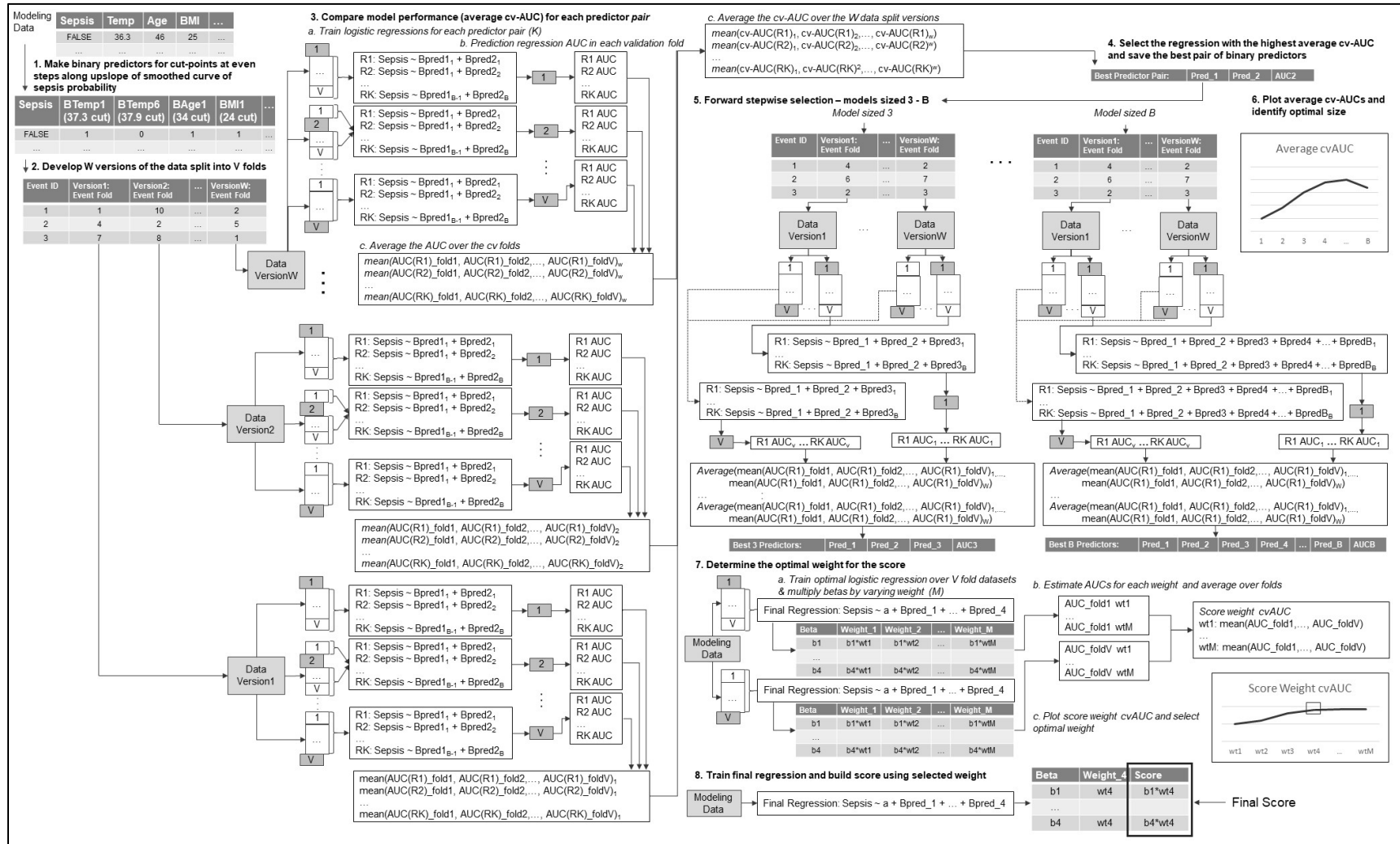
The performance of all two binary-form predictor logistic regressions were compared using an AUC optimized Discrete Super Learner.²⁶ The Discrete Super Learner uses cross-

validation to estimate the performance (cv-AUC) of user-selected algorithms and provides an easy way to compare numerous models or model variants.²⁶ Because the randomly-generated cross-validation folds may impact the performance of the resulting algorithm, the Discrete Super Learner was rerun using 10 different seeds and the average cv-AUC for each binary predictor pair was saved. The binary predictor pair with the highest average cv-AUC was then identified and the binary predictors were saved.

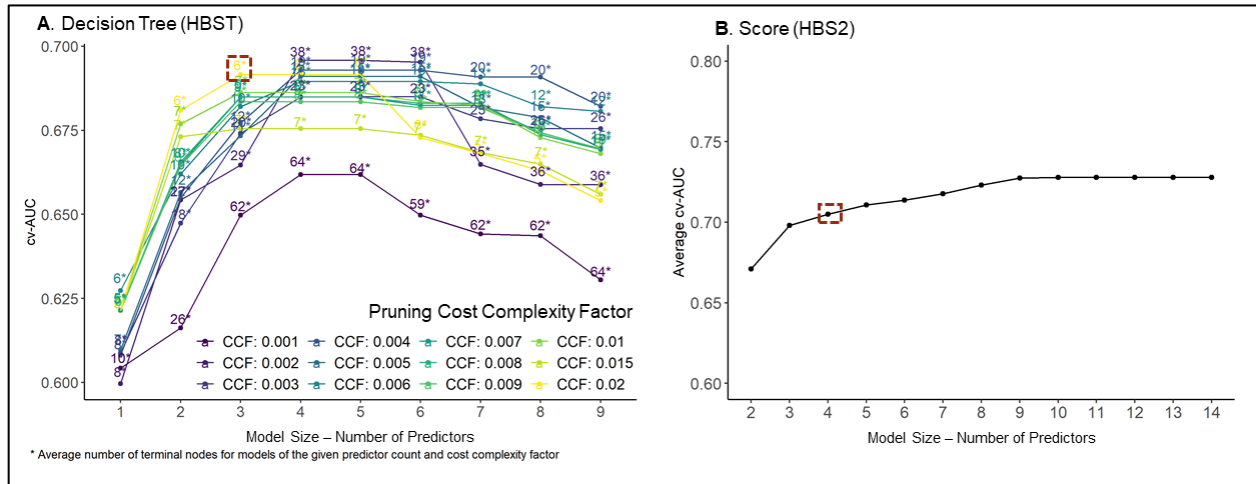
Using the Discrete Super Learner and the binary-form predictors from the best two predictor model, the best three-predictor logistic regressions were identified. As with the two-predictor model, we performed the Discrete Super Learner using 10 different seeds and the best three predictor model was the model with the highest average cv-AUC. This forward stepwise process was then repeated for the selection of the best subsequent models (e.Figure 2). All binary predictors not previously selected were considered in the forward step and we terminated the forward step process when the last 5 best models of a given size were equal to or below the single best model of any given size (14 models examined, eFigure 3).

From the selected “best” logistic regression, we developed the scores using a beta weighted approach.¹¹ This entailed multiplying the regression betas by a specific weight and rounding the products to the nearest whole number. To select the optimal weight, we estimated the cv-AUC for the scores resulting from the following weights: 1, 5, 10, 20, 50, 100, 1000, 10000, 10000000 (full betas) (e.Figure 3). We then selected the lowest weight that led to a similar cv-AUC as full beta result (selected weight: 1).

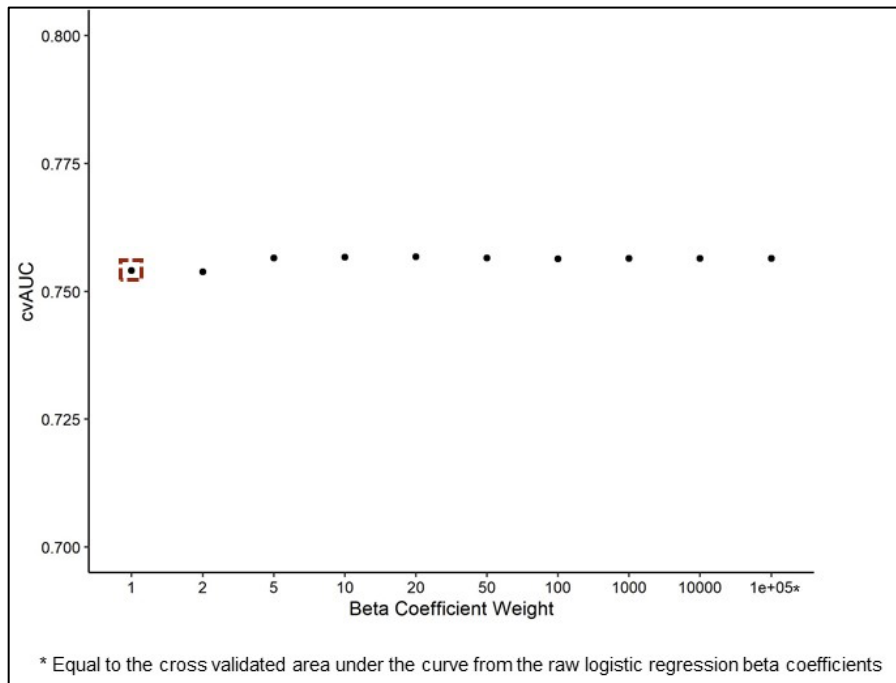
e.Figure 2: High-Risk Bacteremia Score Development Schematic



eFigure 3: Maximum High-risk Bacteremia Cross-Validated Area Under the Curve (cv-AUC) Estimates for Decision Trees and average cv-AUC Scores of Different Sizes Among Allogeneic Hematopoietic Stem Cell Transplant Recipients with Potential Bloodstream Infections



eFigure 4: Cross Validated Area Under the Curve for Varying Final Logistic Regression Beta Coefficient Integer Weights



Decision Tree (HBST) Development

To balance our highly imbalanced data (events ~2.4% of PBIs) and have a sufficient number of events to train our decision tree, we re-sampled our data to an approximate 50/50 split. We performed this re-sampling using the Synthetic Minority Over-sampling Technique (SMOTE) method (combination of over-sampling the minority class through example synthesis and under-sampling the majority class).^{27,28} The resampling was performed on factorized data ensuring that the simulated minority group observations were generated as the most common surrounding factors and not an average of the numeric values. To ensure our cv-AUC were reflective of testing performance, we SMOTED (re-sampled) the training set for each k-fold iteration but not the testing set. This allowed us to build the model in SMOTED data but test the model in raw data.

Trees were grown using recursive binary splitting and tree flexibility resulted from a combination of the number of included predictors and the cost complexity factor (range: 0.001-0.005) used in tree pruning. Because one predictor with multiple splits may have been more predictive than a tree with multiple predictors with few splits, we compared the cv-AUC for each model size combination (number of terminal nodes comprised of the number of predictors / cost complexity for all predictors).

Calibration Plots for Developed and Existing Prediction Tools

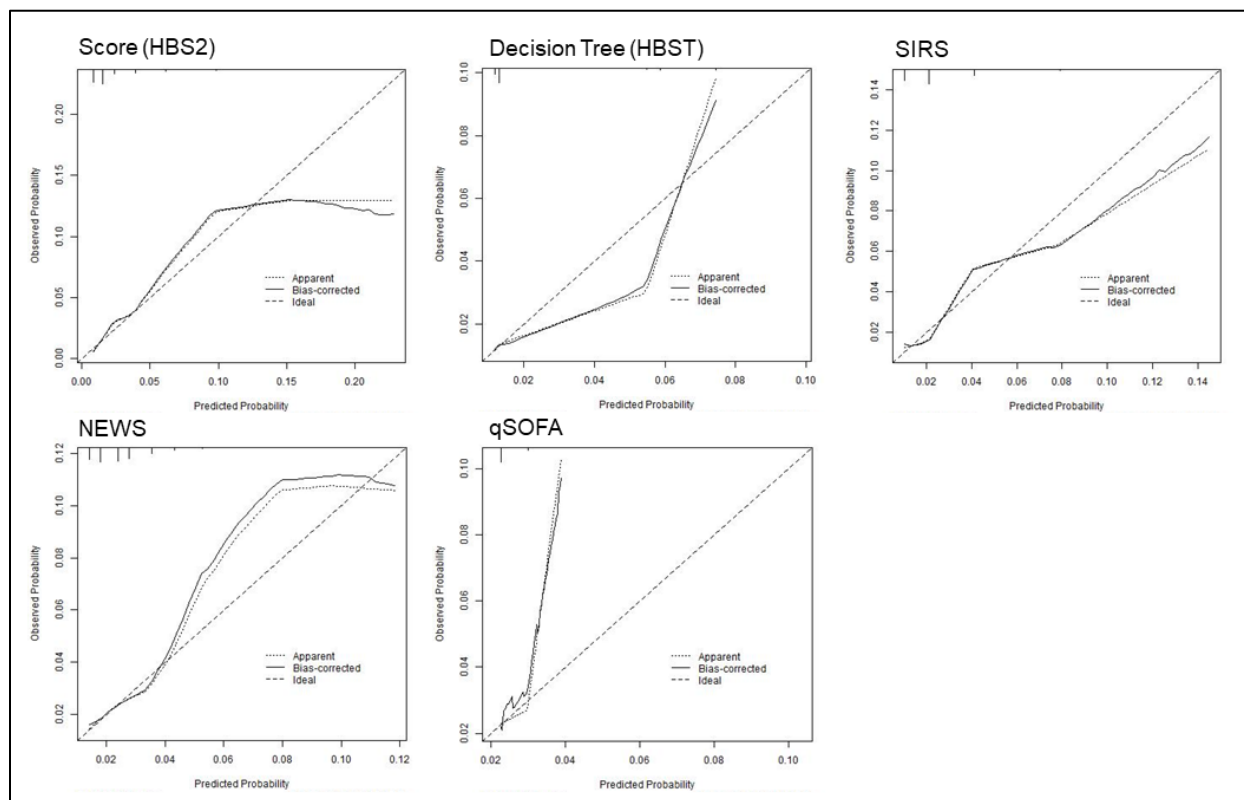
e.Table 4: Calibration - Observed Versus Predicted High-risk Bacteremia^a

Observed high-risk bacteremia (%) ^b	2.6%
<i>Sepsis prediction (mean (IQR))</i>	
Bedside Criteria (HBS2)	2.6% (0.9, 2.5%)
Bedside Tree (HBST)	2.6% (1.3, 1.3%)
SIRS	2.6% (1.2, 2.3%)
qSOFA	2.6% (2.5, 2.5%)
NEWS	2.6% (2.0, 3.0%)

^a High-risk bacteremia defined as any gram-negative bacteremia, *Staphylococcus aureus* or *Streptococcal* bacteremia

^b Performed in the validation dataset

eFigure 5: Calibration Plots by Predictive Tool



e.Table 5: Culture Result Specific Tool Results in the Validation Data

		HBS2 (Score) ^a			HBST (Tree)			SIRS			qSOFA			NEWS		
	Total (N = 2192)	Too l Test +	Too l Test -	Inappro priate Test Proporti on	Too l Test +	Too l Test -	Inappro priate Test Proporti on	Too l Test +	Too l Tes t -	Inappro priate Test Proport ion	Too l Tes t +	Too l Tes t -	Inappr opriate Test Propor tion	Tool Test +	Too l Tes t -	Inappro priate Test Proporti on
High-risk bacteremia																
Gram-negative bacteremia	30	23	7	23%	15	15	50%	15	15	50%	1	29	97%	2	28	93%
<i>Staphylococcus aureus</i>	1	1	0	0%	1	0	0%	1	0	0%	0	1	100%	2	1	100%
<i>Streptococcus species</i>	27	18	9	33%	20	7	26%	14	13	48%	1	26	96%	0	25	93%
Not High-risk bacteremia																
<i>Coagulase-neg staph^b</i>	37	19	18	51%	14	23	38%	21	16	57%	0	37	0%	1	36	3%
Other Bacteria	28	21	7	75%	15	13	54%	13	15	46%	1	27	4%	1	27	4%
			148			163			159			206			204	
Culture Negative	2069	587	2	28%	435	4	21%	474	5	23%	6	3	0%	29	0	1%

^aTest-positive/Negative according to tool (HBS2 2+, HBST 3.3%+, qSOFA 2+, SIRS, 2+, NEWS 7+)

^b Coagulase-negative staphylococci

Sensitivity Analyses Overview

Due to the retrospective nature of our analysis, we relied upon numerous assumptions. In order to test the robustness of our predictive value findings to these assumptions, we performed multiple sensitivity analyses. Each sensitivity analysis was conducted among patients in the validation cohort and the accuracy metrics were estimated for a cut-point of 2+ for SIRS and qSOFA and 7+ for NEWS. Accuracy metric estimates for our developed tools were estimated at the primary analysis selected optimal cut-point according to the Upper-Left method (HBS2: 2, HBST: 3.3%). For the non-surveillance culture and high/low-risk bacteremia discrimination sensitivity analyses, we examined the accuracy metrics for each cut-point for our developed tools and for SIRS 2+, qSOFA 2+, and NEWS 4+ and 7+.

Missing Data:

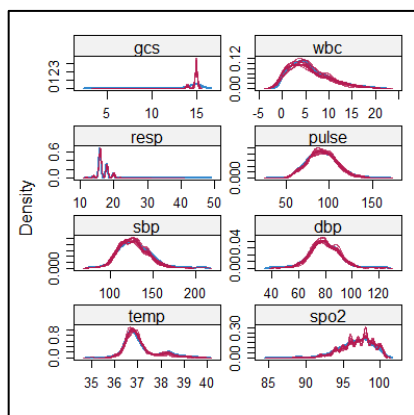
Missing data patterns and the assumptions made in the primary analysis are discussed in the Supplement subsection: *Missing Data*.

- Complete case – Potential bloodstream infections were restricted to those with recorded measurements. This missing data approach assumes a missing completely at random missingness structure.
- Pattern mixture modeling – Using Pattern Mixture Modeling we estimated the predictive value of each tool under a missing at random (MAR) and multiple missing not at random (MNAR) assumptions. We selected 9 different MNAR scenarios, each accounting for a different missing data assumption (i.e., patients with missing data were likely to be sicker or healthier than patients with observed data) – eTable6.

We imputed 10 datasets using multiple imputations by chained equations (MAR assumption) and adjusted the resulting datasets to reflect our various MNAR scenarios (through weight

specification). This resulted in 10 weighted versions of each imputed dataset, or 100 datasets in total. We modeled our missing data using CART (classification and regression trees) and included the predictors of all examined tools, our primary outcome, unique patient identification number, donor relation and demographic predictors in the imputation procedure. The density plot (e.Figure 6) shows the imputed (red) and observed (blue) data for each criteria predictor. From the 100 datasets, we estimated AUCs, sensitivities and specificities and, used Rubin's rule to pool our estimates.²⁹

e.Figure 6: Observed and Imputed Predictive Factor Distributions



Window of Time for Data Inclusion:

For the primary evaluation and for model development, we included data from the 24 hours prior to and the 2 hours following PBI onset (culture collection). Here we examine the impact of varying the time window of data inclusion by estimating the predictive value of our tools and the existing criteria (SIRS, qSOFA, NEWS) in the following time windows: 48 hours prior, 24 hours prior, and 12 hours prior and 2 hours following. For each analysis, the measurement collected closest in time to the time of culture collection was used for evaluation and missing data was imputed using single, normal value imputation.

Alternative Examination Measurement:

In the primary model evaluation and in the model development, we used the predictor

measurement taken closest in time to the culture collection. While we selected this measure due to its proximity to the decision our tools are built to inform (treatment at the time of culture collection), many treating teams may be more interested in 24-hour measurement minimums or maximums. As the use of minimums and maximums may reflect the processing of some tool users, we needed to test the impact of using these measures in the tool rather than the most proximal. In order to do so, we estimated the predictive value of the existing criteria and our tools using the maximum, minimum, and modal values collected during the 24 hours prior to or 2 hours following culture collection. As with the primary analysis, all missing values were imputed using single, normal value imputation.

Alternative Culture Collection Inclusion:

For the primary evaluation and model development, we defined PBIs as a blood culture collected in the absence of recent (last 3 days) antibiotic administration or prescription. We additionally dropped all follow-up cultures, or continuously collected cultures for the 14 days following a positive culture. We implemented these limitations to increase the probability of examining distinct PBIs and to reduce the frequency of antibiotic use related false negatives (cultures were negative because of antibiotic use not the absence of an infection). However, we chose to include cultures taken within three days of a levofloxacin administration or prescription because of the commonality of this antibiotic in neutropenic fever prophylaxis treatment (Center standard at time of study). To test the impact of these restrictions, we estimated the predictive value of our tools and the existing criteria in the following restricted data: included all blood cultures, restricted to blood cultures collected in the absence of any antibiotic use in the previous three days (including levofloxacin), and restricted to blood cultures collected in the absence of antibiotic use in the prior 7 days. Additionally, while our tools were built to be used outside of the ICU, if implemented, it would not be surprising if they were used within the ICU at times.

Thus, we felt the need to estimate the predictive accuracy among all PBIs regardless of ICU status at the time of culture collection. Each of these analyses were conducted using single, normal value imputed data and data from the 24 hours prior to and 2 hours following culture collection. As with the primary analysis, the factor measurements collected closest in time to culture collection were used for evaluation.

Non-Surveillance Cultures:

Surveillance cultures are being collected less frequently and a sample in which surveillance cultures are not included may better reflect culture practices at other institutions or centers. To test the predictive validity of our tools and that of the examined existing tools without surveillance cultures, we restricted our validation sample to exclude cultures with site “Blood, Surveillance.” While site of culture is selected by clinician and may not fully capture all surveillance cultures, this exclusion provided us with the best sample of the examined tools under a non-surveillance included scenario. This analysis was conducted using single, normal value imputed data and data from the 24 hours prior to and 2 hours following culture collection. As with the primary analysis, the factor measurements collected closest in time to culture collection were used for evaluation.

Discrimination of High-risk versus Low-risk Bacteremia:

A primary goal of our tools is to provide guidance on antibiotic initiation needs at the time of culture collection. In order to do so, we defined our outcome as high-sepsis risk bacteremia and estimated the predictive value of our tools and existing tools in relation to this outcome among all PBIs collected outside of the ICU. While this gives the best overall estimation of the tools’ utility, it does not shine a light on the ability of the tools to discriminate between high and low risk bacteremia events. To perform such an evaluation, we restrict our “negative” PBIs to cultures that tested positive for *Coagulase-negative staphylococci* (a low infection commonly

considered low risk and without need of antibiotic treatment). As with the other sensitivity analysis, this analysis was conducted using single, normal value imputed data and data from the 24 hours prior to and 2 hours following culture collection. As with the primary analysis, the factor measurements collected closest in time to culture collection were used for evaluation.

Sensitivity Analyses Results

Missing Data:

Complete Case

e.Table 6: Predictive Value of Developed Tools in Relation to High-Risk Bacteremia^a Among Allogeneic HCT Recipients with Potential Bloodstream Infection Using the Complete Case Missing Data Method

	AUC ^b	Sensitivity ^c	Specificity ^c	Positive Predictive Value ^c	Negative Predictive Value ^c
HBS2 - Bedside Score (2+) ^a	0.78 (0.72, 0.84)	81.2 (67.9, 89.9%)	59.3 (55.6, 62.8%)	6.7 (4.9, 9.1%)	98.9 (97.9, 99.4%)
HBST - Bedside Decision Tree (3.3%+) ^a	0.73 (0.66, 0.80)	68.8 (55.4, 79.6%)	71.2 (67.4, 74.7%)	7.9 (5.7, 10.9%)	98.4 (97.4, 99.1%)
SIRS (2+)	0.66 (0.59, 0.74)	60.4 (46.5, 72.8%)	68.5 (65.4, 71.4%)	6.4 (4.5, 9.1%)	98.0 (96.9, 98.7%)
qSOFA (2+)	0.55 (0.49, 0.61)	4.2 (1.0, 15.2%)	99.5 (98.9, 99.8%)	22.2 (5.6, 57.9%)	96.7 (95.5, 97.5%)
NEWS (7+)	0.66 (0.59, 0.74)	6.3 (2.0, 17.7%)	97.8 (95.8, 98.9%)	9.4 (2.8, 27.4%)	96.7 (95.5, 97.5%)

^a Based on Upper Left Method Optimal Cut-points from primary analysis

^b 95% Confidence Interval estimated using 2000 stratified bootstrapped replicates

^c 95% Confidence Intervals estimated using generalized estimating equations with robust standard errors

Pattern Mixture Models

e.Table 7: Pattern-mixture Model MNAR Scenarios

			Glasgow Coma Scale Score			White Blood Cell Count		
Scenario Number	Type of weight	MNAR rescaling parameters	Percent Missingness	MAR MI Means (SE) ^a	Adjusted MI Means (SE) ^b	Percent Missingness	MAR MI Means (SE) ^a	Adjusted MI Means (SE) ^b
1	-	1	94%	14.75 (0.11)	14.75 (0.11)	11%	5.84 (0.09)	5.84 (0.09)
2	Multiplicative	1.05	94%	14.75 (0.11)	14.91 (0.05)	11%	5.84 (0.09)	5.96 (0.10)
3	Multiplicative	1.1	94%	14.75 (0.11)	14.94 (0.03)	11%	5.84 (0.09)	6.15 (0.11)
4	Multiplicative	0.95	94%	14.75 (0.11)	13.36 (0.10)	11%	5.84 (0.09)	5.78 (0.09)
5	Multiplicative	0.9	94%	14.75 (0.11)	11.96 (0.09)	11%	5.84 (0.09)	5.71 (0.09)
6	Additive	0.5	94%	14.75 (0.11)	14.76 (0.11)	11%	5.84 (0.09)	5.84 (0.09)
7	Additive	1	94%	14.75 (0.11)	14.84 (0.09)	11%	5.84 (0.09)	5.87 (0.09)
8	Additive	2	94%	14.75 (0.11)	14.88 (0.07)	11%	5.84 (0.09)	5.90 (0.10)
9	Additive	-0.5	94%	14.75 (0.11)	14.06 (0.11)	11%	5.84 (0.09)	5.81 (0.09)
10	Additive	-1	94%	14.75 (0.11)	13.36 (0.10)	11%	5.84 (0.09)	5.78 (0.09)
			Respiratory Rate (Breaths per Minute)			Heart Rate (Beats per Minute)		
			Percent Missingness	MAR MI Means (SE) ^a	Adjusted MI Means (SE) ^b	Percent Missingness	MAR MI Means (SE) ^a	Adjusted MI Means (SE) ^b
1	-	1	31%	17.1 (0.07)	17.10 (0.07)	30%	91.95 (0.41)	91.95 (0.41)
2	Multiplicative	1.05	31%	17.1 (0.07)	18.16 (0.08)	30%	91.95 (0.41)	97.57 (0.48)
3	Multiplicative	1.1	31%	17.1 (0.07)	19.76 (0.12)	30%	91.95 (0.41)	105.99 (0.68)
4	Multiplicative	0.95	31%	17.1 (0.07)	16.56 (0.07)	30%	91.95 (0.41)	89.14 (0.40)
5	Multiplicative	0.9	31%	17.1 (0.07)	16.03 (0.07)	30%	91.95 (0.41)	86.34 (0.42)
6	Additive	0.5	31%	17.1 (0.07)	17.12 (0.07)	30%	91.95 (0.41)	92.09 (0.41)
7	Additive	1	31%	17.1 (0.07)	17.36 (0.07)	30%	91.95 (0.41)	93.36 (0.42)
8	Additive	2	31%	17.1 (0.07)	17.63 (0.08)	30%	91.95 (0.41)	94.76 (0.44)
9	Additive	-0.5	31%	17.1 (0.07)	16.83 (0.07)	30%	91.95 (0.41)	90.55 (0.40)
10	Additive	-1	31%	17.1 (0.07)	16.56 (0.07)	30%	91.95 (0.41)	89.14 (0.40)
			Systolic Blood Pressure			Diastolic Blood Pressure		
			Percent Missingness	MAR MI Means (SE) ^a	Adjusted MI Means (SE) ^b	Percent Missingness	MAR MI Means (SE) ^a	Adjusted MI Means (SE) ^b
1	-	1	32%	127.74 (0.46)	127.74 (0.46)	32%	79.16 (0.34)	79.16 (0.34)
2	Multiplicative	1.05	32%	127.74 (0.46)	136.00 (0.57)	32%	79.16 (0.34)	84.34 (0.43)
3	Multiplicative	1.1	32%	127.74 (0.46)	148.39 (0.86)	32%	79.16 (0.34)	92.11 (0.62)
4	Multiplicative	0.95	32%	127.74 (0.46)	123.61 (0.46)	32%	79.16 (0.34)	76.57 (0.33)
5	Multiplicative	0.9	32%	127.74 (0.46)	119.48 (0.49)	32%	79.16 (0.34)	73.98 (0.33)
6	Additive	0.5	32%	127.74 (0.46)	127.94 (0.46)	32%	79.16 (0.34)	79.29 (0.35)
7	Additive	1	32%	127.74 (0.46)	129.80 (0.48)	32%	79.16 (0.34)	80.45 (0.36)
8	Additive	2	32%	127.74 (0.46)	131.87 (0.50)	32%	79.16 (0.34)	81.75 (0.38)
9	Additive	-0.5	32%	127.74 (0.46)	125.67 (0.45)	32%	79.16 (0.34)	77.86 (0.33)
10	Additive	-1	32%	127.74 (0.46)	123.61 (0.46)	32%	79.16 (0.34)	76.57 (0.33)
			Oxygen Saturation			Temperature (Celsius)		
			Percent Missingness	MAR MI Means (SE) ^a	Adjusted MI Means (SE) ^b	Percent Missingness	MAR MI Means (SE) ^a	Adjusted MI Means (SE) ^b
1	-	1	67%	96.9 (0.09)	96.90 (0.09)	35%	37.08 (0.03)	37.08 (0.03)
2	Multiplicative	1.05	67%	96.9 (0.09)	98.94 (0.04)	35%	37.08 (0.03)	39.68 (0.08)
3	Multiplicative	1.1	67%	96.9 (0.09)	98.94 (0.04)	35%	37.08 (0.03)	43.57 (0.19)
4	Multiplicative	0.95	67%	96.9 (0.09)	90.27 (0.12)	35%	37.08 (0.03)	35.78 (0.05)
5	Multiplicative	0.9	67%	96.9 (0.09)	83.65 (0.20)	35%	37.08 (0.03)	34.48 (0.08)
6	Additive	0.5	67%	96.9 (0.09)	97.19 (0.08)	35%	37.08 (0.03)	37.14 (0.03)
7	Additive	1	67%	96.9 (0.09)	98.73 (0.05)	35%	37.08 (0.03)	37.73 (0.04)
8	Additive	2	67%	96.9 (0.09)	98.94 (0.04)	35%	37.08 (0.03)	38.38 (0.05)
9	Additive	-0.5	67%	96.9 (0.09)	93.59 (0.10)	35%	37.08 (0.03)	36.43 (0.03)
10	Additive	-1	67%	96.9 (0.09)	90.27 (0.12)	35%	37.08 (0.03)	35.78 (0.05)

^a The Rubin's rule pooled summary estimate for each imputed factor, continuous variables were imputed using linear, binary using logistic regression, and non-normal continuous variables (GCS) with predictive mean matching

^b The Rubin's rule pooled summary estimate for each imputed factor following MNAR scenario adjustment, categorical variables were rounded to the nearest integer within the bounds of the factor

e.Table 8: Predictive Value of Developed Tools in Relation to High-Risk Bacteremia^a Under Varying Missing Data Scenarios (Pattern Mixture Modeling)

	Scenario 1			Scenario 2			Scenario 3		
	AUC ^b	Sensitivity ^c	Specificity ^c	AUC ^b	Sensitivity ^c	Specificity ^c	AUC ^b	Sensitivity ^c	Specificity ^c
Bedside Score (HBS2) ^d	0.78 (0.76, 0.79)	77.7% (61.9, 88.2%)	57.0% (53.8, 60.1%)	0.74 (0.73, 0.75)	82.0% (69.9, 89.9%)	38.1% (35.2, 41.2%)	0.71 (0.71, 0.72)	82.1% (69.9, 90.0%)	37.9% (35.0, 41.0%)
Bedside Tree (HBST) ^d	0.73 (0.71, 0.74)	66.2% (52.9, 77.4%)	70.4% (67.1, 73.6%)	0.68 (0.67, 0.68)	69.3% (57.5, 79.1%)	46.5% (43.4, 49.6%)	0.68 (0.67, 0.68)	69.5% (57.7, 79.3%)	46.1% (43.1, 49.2%)
SIRS	0.66 (0.64, 0.69)	64.9% (49.7, 77.6%)	67.7% (64.8, 70.4%)	0.62 (0.61, 0.64)	68.6% (55.1, 79.6%)	45.4% (42.4, 48.5%)	0.60 (0.59, 0.61)	69.3% (55.8, 80.1%)	42.9% (40.0, 45.9%)
qSOFA	0.56 (0.50, 0.61)	6.1% (1.6, 21.0%)	98.6% (97.6, 99.2%)	0.55 (0.53, 0.57)	4.2% (1.0, 16.1%)	99.4% (98.8, 99.7%)	0.45 (0.44, 0.47)	5.2% (1.7, 15.1%)	99.6% (98.8, 99.9%)
NEWS	0.67 (0.58, 0.75)	7.4% (1.7, 26.8%)	97.3% (95.9, 98.2%)	0.58 (0.56, 0.60)	6.3% (1.5, 23.1%)	95.8% (93.8, 97.1%)	0.53 (0.52, 0.54)	9.5% (3.3, 24.5%)	78.2% (75.3, 80.8%)
	Scenario 4			Scenario 5			Scenario 6		
	AUC ^b	Sensitivity ^c	Specificity ^c	AUC ^b	Sensitivity ^c	Specificity ^c	AUC ^b	Sensitivity ^c	Specificity ^c
Bedside Score (HBS2) ^d	0.78 (0.77, 0.78)	73.9% (59.6, 84.4%)	65.8% (62.8, 68.6%)	0.78 (0.77, 0.78)	73.4% (58.9, 84.2%)	68.5% (65.7, 71.3%)	0.78 (0.76, 0.79)	78.1% (61.6, 88.8%)	56.9% (53.8, 60.0%)
Bedside Tree (HBST) ^d	0.70 (0.69, 0.70)	69.2% (57.3, 79.0%)	46.7% (43.6, 49.8%)	0.70 (0.69, 0.70)	69.2% (57.3, 79.0%)	46.7% (43.6, 49.8%)	0.73 (0.71, 0.75)	66.3% (53.0, 77.4%)	70.0% (66.4, 73.5%)
SIRS	0.64 (0.63, 0.66)	65.9% (53.1, 76.8%)	56.3% (53.6, 59.1%)	0.66 (0.64, 0.67)	68.1% (54.5, 79.2%)	59.4% (56.5, 62.2%)	0.67 (0.64, 0.70)	66.1% (50.9, 78.6%)	65.7% (62.6, 68.6%)
qSOFA	0.50 (0.47, 0.52)	13.7% (6.4, 26.9%)	89.7% (87.2, 91.7%)	0.47 (0.44, 0.50)	14.6% (6.9, 28.2%)	79.8% (77.0, 82.4%)	0.56 (0.50, 0.61)	5.9% (1.4, 21.8%)	98.6% (97.6, 99.2%)
NEWS	0.59 (0.57, 0.61)	41.3% (29.5, 54.2%)	45.6% (42.5, 48.7%)	0.60 (0.59, 0.61)	41.3% (29.5, 54.2%)	45.4% (42.3, 48.5%)	0.66 (0.58, 0.75)	7.6% (1.9, 25.8%)	97.3% (96.0, 98.3%)
	Scenario 7			Scenario 8			Scenario 9		
	AUC ^b	Sensitivity ^c	Specificity ^c	AUC ^b	Sensitivity ^c	Specificity ^c	AUC ^b	Sensitivity ^c	Specificity ^c
Bedside Score (HBS2) ^d	0.76 (0.75, 0.76)	82.0% (69.8, 90.0%)	38.7% (35.8, 41.8%)	0.75 (0.75, 0.76)	82.0% (69.8, 89.9%)	38.2% (35.3, 41.2%)	0.78 (0.77, 0.79)	73.9% (59.7, 84.5%)	63.2% (60.1, 66.3%)
Bedside Tree (HBST) ^d	0.68 (0.68, 0.69)	69.5% (57.4, 79.5%)	48.3% (45.1, 51.5%)	0.68 (0.67, 0.68)	69.3% (57.5, 79.1%)	46.5% (43.4, 49.6%)	0.70 (0.68, 0.71)	67.4% (53.3, 78.9%)	50.3% (46.7, 54.0%)
SIRS	0.64 (0.62, 0.65)	68.0% (53.7, 79.6%)	49.7% (46.6, 52.7%)	0.63 (0.62, 0.65)	68.0% (53.4, 79.8%)	48.1% (45.1, 51.2%)	0.65 (0.64, 0.67)	64.4% (49.9, 76.6%)	63.0% (60.0, 66.0%)
qSOFA	0.56 (0.50, 0.61)	5.9% (1.4, 21.8%)	98.7% (97.7, 99.3%)	0.55 (0.53, 0.57)	4.4% (1.1, 16.3%)	99.3% (98.5, 99.7%)	0.50 (0.49, 0.52)	12.9% (6.0, 25.6%)	92.5% (90.6, 94.0%)
NEWS	0.64 (0.57, 0.71)	8.3% (2.2, 26.9%)	96.9% (95.4, 97.9%)	0.61 (0.59, 0.63)	5.3% (1.0, 23.5%)	97.6% (96.0, 98.5%)	0.51 (0.45, 0.56)	32.4% (18.4, 50.4%)	65.0% (61.2, 68.7%)
	Scenario 10								
	AUC ^b	Sensitivity ^c	Specificity ^c						
Bedside Score (HBS2) ^d	0.78 (0.77, 0.78)	73.9% (59.6, 84.4%)	65.8% (62.8, 68.6%)						
Bedside Tree (HBST) ^d	0.70 (0.69, 0.70)	69.2% (57.3, 79.0%)	46.7% (43.6, 49.8%)						
SIRS	0.65 (0.64, 0.67)	65.9% (53.1, 76.8%)	56.3% (53.6, 59.1%)						
qSOFA	0.50 (0.47, 0.52)	13.7% (6.4, 26.9%)	89.7% (87.2, 91.7%)						
NEWS	0.59 (0.57, 0.61)	41.3% (29.5, 54.2%)	45.6% (42.5, 48.7%)						

^a High-risk bacteremia defined as gram-negative, *Streptococcus* species, or *Staphylococcus aureus* culture confirmed bacteremia

^b 95% Confidence Interval estimated using 2000 stratified bootstrapped replicates

^c 95% Confidence Intervals estimated using generalized estimating equations with robust standard errors

^d Based on Upper Left Method Optimal Cut-points from primary analysis

Window of Time for Data Inclusion:

e.Table 9: Predictive Value of Developed Tools in Relation to High-Risk Bacteremia^b Using Bedside Examination Data from Varying Windows of Time

	AUC ^b	Sensitivity ^c	Specificity ^c	Positive Predictive Value ^c	Negative Predictive Value ^c
<i>Prior 12 & After 2</i>					
HBS2 - Bedside Score (2+) ^d	0.75 (0.68, 0.82)	70.7 (54.6, 82.8%)	72.2 (69.5, 74.7%)	6.5 (4.7, 8.8%)	98.9 (98.0, 99.4%)
HBST - Bedside Decision Tree (3.3%+) ^d	0.74 (0.68, 0.81)	62.1 (46.6, 75.4%)	78.4 (75.2, 81.2%)	7.2 (5.2, 9.9%)	98.7 (97.8, 99.3%)
SIRS (2+)	0.65 (0.57, 0.72)	48.3 (35.9, 60.9%)	78.5 (76.4, 80.5%)	5.8 (4.0, 8.2%)	98.2 (97.4, 98.8%)
qSOFA (2+)	0.49 (0.44, 0.55)	3.4 (0.9, 12.9%)	99.7 (99.4, 99.9%)	25 (6.3, 62.3%)	97.4 (96.5, 98.1%)
NEWS (7+)	0.58 (0.51, 0.66)	6.9 (2.6, 17.2%)	98.6 (97.3, 99.3%)	11.8 (4, 29.9%)	97.5 (96.6, 98.2%)
<i>Prior 24</i>					
HBS2 - Bedside Score (2+) ^d	0.67 (0.59, 0.75)	53.4 (40.7, 65.8%)	77.4 (74.9, 79.6%)	6.0 (4.1, 8.7%)	98.4 (97.6, 98.9%)
HBST - Bedside Decision Tree (3.3%+) ^d	0.64 (0.56, 0.71)	41.4 (28.6, 55.4%)	81.4 (78.2, 84.2%)	5.7 (3.8, 8.4%)	98.1 (97.1, 98.7%)
SIRS (2+)	0.62 (0.54, 0.69)	34.5 (22.9, 48.2%)	82.7 (80.6, 84.6%)	5.1 (3.3, 8.0%)	97.9 (97.0, 98.5%)
qSOFA (2+)	0.48 (0.43, 0.53)	0.0 (0.0, 6.2) ^e	99.7 (99.2, 99.9%)	0.0 (0.0, 41) ^e	97.3 (96.4, 98.0%)
NEWS (7+)	0.51 (0.44, 0.59)	0.0 (0.0, 6.2) ^e	99.1 (98.0, 99.6%)	0 (0.0, 17.6%) ^e	97.3 (96.4, 98.0%)
<i>Prior 48</i>					
HBS2 - Bedside Score (2+) ^d	0.67 (0.59, 0.75)	53.4 (38.9, 67.4%)	75.0 (72.4, 77.4%)	5.5 (3.8, 7.8%)	98.3 (97.4, 99.0%)
HBST - Bedside Decision Tree (3.3%+) ^d	0.65 (0.58, 0.72)	43.1 (30.0, 57.2%)	80.9 (77.7, 83.7%)	5.8 (3.9, 8.4%)	98.1 (97.2, 98.8%)
SIRS (2+)	0.63 (0.56, 0.70)	39.7 (27.3, 53.5%)	79.5 (77.3, 81.5%)	5.0 (3.3, 7.5%)	98.0 (97.0, 98.6%)
qSOFA (2+)	0.48 (0.43, 0.53)	0.0 (0.0, 6.2) ^e	99.6 (99.1, 99.8%)	0.0 (0.0, 36.9) ^e	97.3 (96.4, 98.0%)
NEWS (7+)	0.55 (0.48, 0.62)	0.0 (0.0, 6.2) ^e	99.0 (97.9, 99.5%)	0.0 (0.0, 15.4) ^e	97.3 (96.4, 98.0%)
<i>Prior 48 & After 2</i>					
HBS2 - Bedside Score (2+) ^d	0.70 (0.62, 0.78)	70.7 (54.6, 82.8%)	69.4 (66.6, 72.1%)	5.9 (4.3, 8.0%)	98.9 (97.9, 99.4%)
HBST - Bedside Decision Tree (3.3%+) ^d	0.74 (0.68, 0.81)	62.1 (46.6, 75.4%)	77.8 (74.7, 80.6%)	7.1 (5.1, 9.7%)	98.7 (97.7, 99.2%)
SIRS (2+)	0.67 (0.59, 0.74)	55.2 (41.2, 68.4%)	74.8 (72.5, 77.0%)	5.6 (4.0, 7.9%)	98.4 (97.5, 99.0%)
qSOFA (2+)	0.51 (0.45, 0.56)	3.4 (0.9, 12.9%)	99.6 (99.2, 99.8%)	20 (4.9, 55.0%)	97.4 (96.5, 98.1%)
NEWS (7+)	0.60 (0.53, 0.68)	6.9 (2.6, 17.2%)	98.6 (97.3, 99.3%)	11.8 (4.0, 29.9%)	97.5 (96.6, 98.2%)

^aHigh-risk bacteremia defined as gram-negative, Streptococcus species, or Staphylococcus aureus culture confirmed bacteremia

^b 95% Confidence Interval estimated using 2000 stratified bootstrapped replicates

^c 95% Confidence Intervals estimated using generalized estimating equations with robust standard errors

^d Based on Upper Left Method Optimal Cut-points from primary analysis

^e 95% Confidence Intervals estimated using Clopper Pearson methods (zeros in either the denominator or numerator prevented reliable generalized estimating equations uncertainty estimation)

Alternative Examination Measurement:

e.Table 10: Predictive Value of Developed Tools in Relation to High-Risk Bacteremia^a Using Maximum, Minimum, and Mode Measures Collected During the 24 Hours Prior to and Two Hours Following Culture Collection for All Potential Bloodstream Infections

	AUC ^b	Sensitivity ^c	Specificity ^c	Positive Predictive Value ^c	Negative Predictive Value ^c
<i>Maximum Value</i>					
HBS2 - Bedside Score (2+) ^d	0.78 (0.71, 0.84)	93.1 (83.1, 97.4%)	32.0 (29.6, 34.4%)	3.6 (2.7, 4.8%)	99.4 (98.5, 99.8%)
HBST - Bedside Decision Tree (3.3%+) ^d	0.76 (0.70, 0.82)	67.2 (50.7, 80.4%)	73.3 (70.0, 76.4%)	6.4 (4.7, 8.7%)	98.8 (97.8, 99.4%)
SIRS (2+)	0.67 (0.60, 0.74)	58.6 (43.9, 71.9%)	70.3 (67.8, 72.7%)	5.1 (3.6, 7.1%)	98.4 (97.4, 99.0%)
qSOFA (2+)	0.53 (0.47, 0.58)	0.0 (0.0, 6.2) ^e	99.4 (98.6, 99.8%)	0.0 (0.0, 26.5) ^e	97.3 (96.4, 98.0%)
NEWS (7+)	0.59 (0.51, 0.66)	5.2 (1.7, 15.0%)	98.2 (96.6, 99.1%)	7.3 (2.1, 22.6%)	97.4 (96.5, 98.1%)
<i>Minimum Value</i>					
HBS2 - Bedside Score (2+) ^d	0.66 (0.59, 0.73)	37.9 (25.3, 52.4%)	83.7 (81.4, 85.7%)	5.9 (3.9, 9.0%)	98.0 (97.1, 98.7%)
HBST - Bedside Decision Tree (3.3%+) ^d	0.58 (0.50, 0.65)	29.3 (18.8, 42.7%)	84.8 (81.6, 87.5%)	5.0 (3.1, 7.9%)	97.8 (96.8, 98.5%)
SIRS (2+)	0.54 (0.47, 0.62)	24.1 (14.2, 38.0%)	84.9 (82.9, 86.7%)	4.2 (2.4, 7.1%)	97.6 (96.7, 98.3%)
qSOFA (2+)	0.53 (0.47, 0.59)	1.7 (0.2, 11.3%)	99.5 (99.1, 99.8%)	9.1 (1.2, 44.3%)	97.4 (96.5, 98.1%)
NEWS (7+)	0.48 (0.40, 0.55)	1.7 (0.2, 11.3%)	98.8 (98.0, 99.3%)	3.8 (0.5, 23.4%)	97.4 (96.5, 98.0%)
<i>Mode Value</i>					
HBS2 - Bedside Score (2+) ^d	0.69 (0.62, 0.77)	60.3 (46.3, 72.9%)	70.1 (67.1, 72.9%)	5.2 (3.7, 7.2%)	98.5 (97.5, 99.1%)
HBST - Bedside Decision Tree (3.3%+) ^d	0.67 (0.60, 0.75)	46.6 (32.9, 60.7%)	80.6 (77.5, 83.5%)	6.1 (4.2, 8.9%)	98.2 (97.3, 98.8%)
SIRS (2+)	0.59 (0.51, 0.66)	37.9 (25.6, 52.1%)	76.9 (74.6, 79.0%)	4.3 (2.8, 6.5%)	97.9 (96.9, 98.5%)
qSOFA (2+)	0.50 (0.45, 0.55)	0.0 (0.0, 6.2) ^e	99.9 (99.6, 100%)	0.0 (0.0, 70.8) ^e	97.4 (96.5, 98.0%)
NEWS (7+)	0.53 (0.45, 0.62)	8.6 (3.6, 19.4%)	98.2 (97.2, 98.9%)	11.6 (4.6, 26.2%)	97.5 (96.6, 98.2%)

^aHigh-risk bacteremia defined as gram-negative, *Streptococcus* species, or *Staphylococcus aureus* culture confirmed bacteremia

^b95% Confidence Interval estimated using 2000 stratified bootstrapped replicates

^c95% Confidence Intervals estimated using generalized estimating equations with robust standard errors

^dBased on Upper Left Method Optimal Cut-points from primary analysis

^e95% Confidence Intervals estimated using Clopper Pearson methods (zeros in either the denominator or numerator prevented reliable generalized estimating equations uncertainty estimation)

Alternative Culture Collection Inclusion:

e.Table 11: Predictive Value of Developed Tools in Relation to High-Risk Bacteremia^a Using Cultures Under Varying Antibiotic Use Inclusion Restriction

	AUC ^b	Sensitivity ^c	Specificity ^c	Positive Predictive Value ^c	Negative Predictive Value ^c
<i>All Cultures</i>					
HBS2 - Bedside Score (2+) ^d	0.67 (0.62, 0.72)	72.9 (61.8, 81.7%)	53.9 (50.6, 57.3%)	4.3 (3.4, 5.4%)	98.6 (97.8, 99.1%)
HBST - Bedside Decision Tree (3.3%+) ^d	0.64 (0.59, 0.69)	60.7 (48.8, 71.5%)	61.2 (57.7, 64.5%)	4.2 (3.3, 5.4%)	98.2 (97.3, 98.8%)
SIRS (2+)	0.59 (0.54, 0.64)	57.0 (46.9, 66.5%)	62.3 (59.3, 65.2%)	4.1 (3.2, 5.2%)	98.1 (97.3, 98.7%)
qSOFA (2+)	0.49 (0.45, 0.52)	2.8 (0.9, 8.4%)	99.1 (98.5, 99.5%)	8.1 (2.7, 21.8%)	97.3 (96.6, 97.9%)
NEWS (7+)	0.58 (0.52, 0.63)	9.3 (4.8, 17.4%)	93.7 (91.3, 95.4%)	4 (2.1, 7.5%)	97.4 (96.6, 97.9%)
<i>No Use^b in Last 3 Days (Including Levofloxacin)</i>					
HBS2 - Bedside Score (2+) ^d	0.73 (0.65, 0.81)	60.5 (44.9, 74.2%)	73.4 (70.3, 76.3%)	4.7 (3.1, 6.9%)	98.9 (98.0, 99.4%)
HBST - Bedside Decision Tree (3.3%+) ^d	0.71 (0.63, 0.78)	48.8 (32.5, 65.4%)	80.7 (77.3, 83.7%)	5.1 (3.4, 7.8%)	98.7 (97.6, 99.2%)
SIRS (2+)	0.62 (0.53, 0.71)	44.2 (29.2, 60.3%)	79.8 (77.4, 81.9%)	4.5 (2.8, 7.1%)	98.5 (97.6, 99.1%)
qSOFA (2+)	0.52 (0.45, 0.58)	7.0 (2.2, 19.9%)	99.8 (99.5, 99.9%)	42.9 (17.5, 72.7%)	98.0 (97.1, 98.7%)
NEWS (7+)	0.57 (0.49, 0.66)	9.3 (3.4, 22.8%)	98.7 (97.2, 99.4%)	13.3 (4.3, 34.3%)	98.1 (97.1, 98.7%)
<i>No Use^b in Last 7 Days (Excluding Levofloxacin)</i>					
HBS2 - Bedside Score (2+) ^d	0.76 (0.69, 0.83)	69.1 (55.1, 80.3%)	71.2 (68.4, 73.9%)	7.0 (5.0, 9.7%)	98.7 (97.8, 99.2%)
HBST - Bedside Decision Tree (3.3%+) ^d	0.72 (0.65, 0.80)	61.8 (49.1, 73.1%)	78.6 (75.4, 81.6%)	8.3 (6.0, 11.5%)	98.5 (97.6, 99.1%)
SIRS (2+)	0.67 (0.60, 0.74)	50.9 (37.9, 63.8%)	76.3 (73.8, 78.6%)	6.3 (4.3, 9.1%)	98.0 (97.1, 98.7%)
qSOFA (2+)	0.53 (0.47, 0.59)	1.8 (0.3, 11.9%)	99.8 (99.4, 99.9%)	20.0 (2.7, 69.1%)	97.0 (96.0, 97.8%)
NEWS (7+)	0.59 (0.52, 0.66)	3.6 (0.9, 13.5%)	98.6 (97.4, 99.3%)	7.7 (1.8, 28.0%)	97.0 (96.0, 97.8%)
<i>All Potential Bloodstream Infections Regardless of ICU Status at Time of Culture Collection</i>					
HBS2 - Bedside Score (2+) ^d	0.74 (0.68, 0.80)	65.7 (51.9, 77.3%)	70.0 (67.2, 72.5%)	6.5 (4.8, 8.7%)	98.5 (97.5, 99.1%)
HBST - Bedside Decision Tree (3.3%+) ^d	0.69 (0.62, 0.75)	54.3 (41.5, 66.5%)	77.2 (74.0, 80.1%)	7.0 (5.1, 9.5%)	98.2 (97.2, 98.8%)
SIRS (2+)	0.63 (0.56, 0.69)	45.7 (34.2, 57.7%)	75.5 (73.1, 77.7%)	5.6 (4.0, 7.8%)	97.8 (96.8, 98.5%)
qSOFA (2+)	0.52 (0.47, 0.57)	2.9 (0.7, 10.8%)	99.7 (99.3, 99.8%)	22.2 (5.6, 57.9%)	97.0 (96.0, 97.7%)
NEWS (7+)	0.60 (0.54, 0.67)	5.7 (2.1, 14.4%)	98.3 (96.9, 99.1%)	9.8 (3.4, 25.2%)	97.0 (96.1, 97.8%)

^aHigh-risk bacteremia defined as gram-negative, *Streptococcus* species, or *Staphylococcus aureus* culture confirmed bacteremia

^b 95% Confidence Interval estimated using 2000 stratified bootstrapped replicates

^c 95% Confidence Intervals estimated using generalized estimating equations with robust standard errors

^d Based on Upper Left Method Optimal Cut-points from primary analysis

Surveillance Culture Exclusion:

e.Table 12: Predictive Value of Developed Tools in Relation to High-Risk Bacteremia^a Among Non-Surveillance Cultures

	AUC ^b	Sensitivity ^c	Specificity ^c	Positive Predictive Value ^c	Negative Predictive Value ^c
HBS2 - Bedside Score	0.64 (0.59, 0.69)				
Cut-point 0	-	100 (100, 100%)	0.0 (0.0, 0.1)	-	-
Cut-point 1	-	90.0 (79.8, 95.3%)	23.6 (21.1, 26.3%)	3.9 (3.1, 4.9%)	98.6 (96.8, 99.4%)
Cut-point 2	-	78.0 (67.6, 85.8%)	45.2 (41.6, 48.8%)	4.7 (3.7, 5.9%)	98.3 (97.2, 99.0%)
Cut-point 3	-	68.0 (56.2, 77.9%)	58.8 (55.2, 62.3%)	5.4 (4.2, 6.9%)	98.2 (97.1, 98.8%)
Cut-point 4	-	53.0 (42.2, 63.5%)	68.8 (65.4, 72.0%)	5.5 (4.2, 7.3%)	97.7 (96.7, 98.4%)
Cut-point 5	-	34.0 (25.0, 44.3%)	79.0 (76.1, 81.7%)	5.3 (3.8, 7.4%)	97.2 (96.3, 97.9%)
Cut-point 6	-	14.0 (8.4, 22.3%)	89.5 (87.3, 91.4%)	4.4 (2.6, 7.5%)	96.8 (95.9, 97.5%)
Cut-point 7	-	9.0 (4.9, 15.9%)	94.4 (93.0, 95.5%)	5.2 (2.7, 9.8%)	96.8 (95.9, 97.4%)
HBST - Bedside Decision Tree	0.6 (0.55, 0.66)				
Cut-point 0.83%	-	100 (100, 100%)	0.0 (0.0, 0.1)	-	-
Cut-point 0.95%	-	92.0 (84.0, 96.2%)	15.6 (12.7, 19.0%)	3.6 (2.9, 4.6%)	98.3 (96.4, 99.2%)
Cut-point 3.3%	-	62.0 (50.0, 72.7%)	53.0 (49.2, 56.7%)	4.4 (3.4, 5.6%)	97.6 (96.4, 98.4%)
Cut-point 3.4%	-	60.0 (48.2, 70.7%)	57.3 (53.8, 60.8%)	4.6 (3.6, 6.0%)	97.6 (96.5, 98.4%)
Cut-point 3.7%	-	46.0 (35.2, 57.1%)	70.5 (67.7, 73.1%)	5.1 (3.8, 6.9%)	97.4 (96.4, 98.1%)
Cut-point 3.8%	-	45.0 (34.3, 56.1%)	72.4 (69.6, 75.1%)	5.3 (3.9, 7.3%)	97.4 (96.5, 98.2%)
SIRS (2+)	0.56 (0.51, 0.61)	61.0 (51.0, 70.2%)	53.9 (50.7, 57.2%)	4.4 (3.4, 5.6%)	97.6 (96.5, 98.3%)
qSOFA (2+)	0.47 (0.43, 0.51)	4.0 (1.5, 10.3%)	96.2 (93.9, 97.6%)	3.5 (1.5, 8.0%)	96.7 (95.8, 97.4%)
NEWS	0.56 (0.50, 0.61)				
Cut-point of 4	-	33.0 (24.7, 42.5%)	71.1 (67.5, 74.4%)	3.8 (2.7, 5.3%)	96.8 (95.9, 97.6%)
Cut-point of 7	-	9.0 (4.5, 17.3%)	91.7 (88.2, 94.2%)	3.6 (1.8, 7.2%)	96.7 (95.8, 97.4%)

^aHigh-risk bacteremia defined as gram-negative, *Streptococcus* species, or *Staphylococcus aureus* culture confirmed bacteremia

^b 95% Confidence Interval estimated using 2000 stratified bootstrapped replicates

^c 95% Confidence Intervals estimated using generalized estimating equations with robust standard errors

^d Based on Upper Left Method Optimal Cut-points from primary analysis

High and Low Risk Bacteremia Evaluation:

e.Table 13: Discrimination Between High-risk^a and Low-risk^a Bacteremia

	AUC ^b	Sensitivity ^c	Specificity ^c	Positive Predictive Value ^c	Negative Predictive Value ^c
HBS2 - Bedside Score	0.63 (0.55, 0.72)				
Cut-point 0	-	100 (100, 100%)	0.0 (0.0, 6.1)	-	-
Cut-point 1	-	88.7 (79.1, 94.2%)	22.0 (13.2, 34.5%)	67.1 (58.5, 74.7%)	52.0 (30.4, 72.9%)
Cut-point 2	-	73.6 (63.3, 81.8%)	45.8 (33.3, 58.7%)	70.9 (61.8, 78.6%)	49.1 (34.1, 64.3%)
Cut-point 3	-	63.2 (51.2, 73.8%)	62.7 (50.4, 73.6%)	75.3 (65.3, 83.1%)	48.7 (35.3, 62.2%)
Cut-point 4	-	50.0 (39.3, 60.7%)	66.1 (53.5, 76.7%)	72.6 (61.7, 81.3%)	42.4 (31.0, 54.7%)
Cut-point 5	-	33.0 (24.2, 43.2%)	83.1 (72.0, 90.3%)	77.8 (63.6, 87.5%)	40.8 (31.3, 51.1%)
Cut-point 6	-	13.2 (7.9, 21.3%)	93.2 (83.6, 97.4%)	77.8 (54.7, 91.0%)	37.4 (29.1, 46.5%)
Cut-point 7	-	8.5 (4.6, 15.1%)	96.6 (87.4, 99.2%)	81.8 (49.3, 95.4%)	37.0 (28.9, 45.9%)
HBST - Bedside Decision Tree	0.60 (0.51, 0.68)				
Cut-point 0.83%	-	100 (100, 100%)	0.0 (0.0, 6.1)	-	-
Cut-point 0.95%	-	93.4 (85.9, 97.1%)	16.9 (9.0, 29.7%)	66.9 (58.3, 74.5%)	58.8 (32.7, 80.8%)
Cut-point 3.3%	-	60.4 (48.4, 71.2%)	52.5 (39.7, 65.0%)	69.6 (59.8, 77.8%)	42.5 (29.4, 56.7%)
Cut-point 3.4%	-	56.6 (45.0, 67.5%)	59.3 (46.4, 71.1%)	71.4 (61.4, 79.7%)	43.2 (30.9, 56.4%)
Cut-point 3.7%	-	42.5 (32.2, 53.5%)	71.2 (58.9, 81.0%)	72.6 (60.7, 81.9%)	40.8 (30.4, 52.1%)
Cut-point 3.8%	-	41.5 (31.3, 52.5%)	71.2 (58.9, 81.0%)	72.1 (60.1, 81.6%)	40.4 (30.1, 51.6%)
SIRS (2+)	0.50 (0.41, 0.59)	55.7 (45.4, 65.4%)	44.1 (32.3, 56.5%)	64.1 (53.7, 73.4%)	35.6 (24.5, 48.5%)
qSOFA (2+)	0.45 (0.39, 0.52)	3.8 (1.4, 9.8%)	98.3 (88.9, 99.8%)	80.0 (30.9, 97.3%)	36.2 (28.3, 45.0%)
NEWS	0.53 (0.44, 0.62)				
Cut-point of 4	-	29.2 (21.1, 38.9%)	83.1 (72.0, 90.3%)	75.6 (61.3, 85.8%)	39.5 (30.5, 49.4%)
Cut-point of 7	-	9.4 (4.9, 17.5%)	93.2 (83.4, 97.4%)	71.4 (45.2, 88.3%)	36.4 (28.3, 45.4%)

^aHigh-risk bacteremia defined as gram-negative, *Streptococcus* species, or *Staphylococcus aureus* culture confirmed bacteremia, low-risk bacteremia defined as *Coagulase-Negative Staphylococcus*

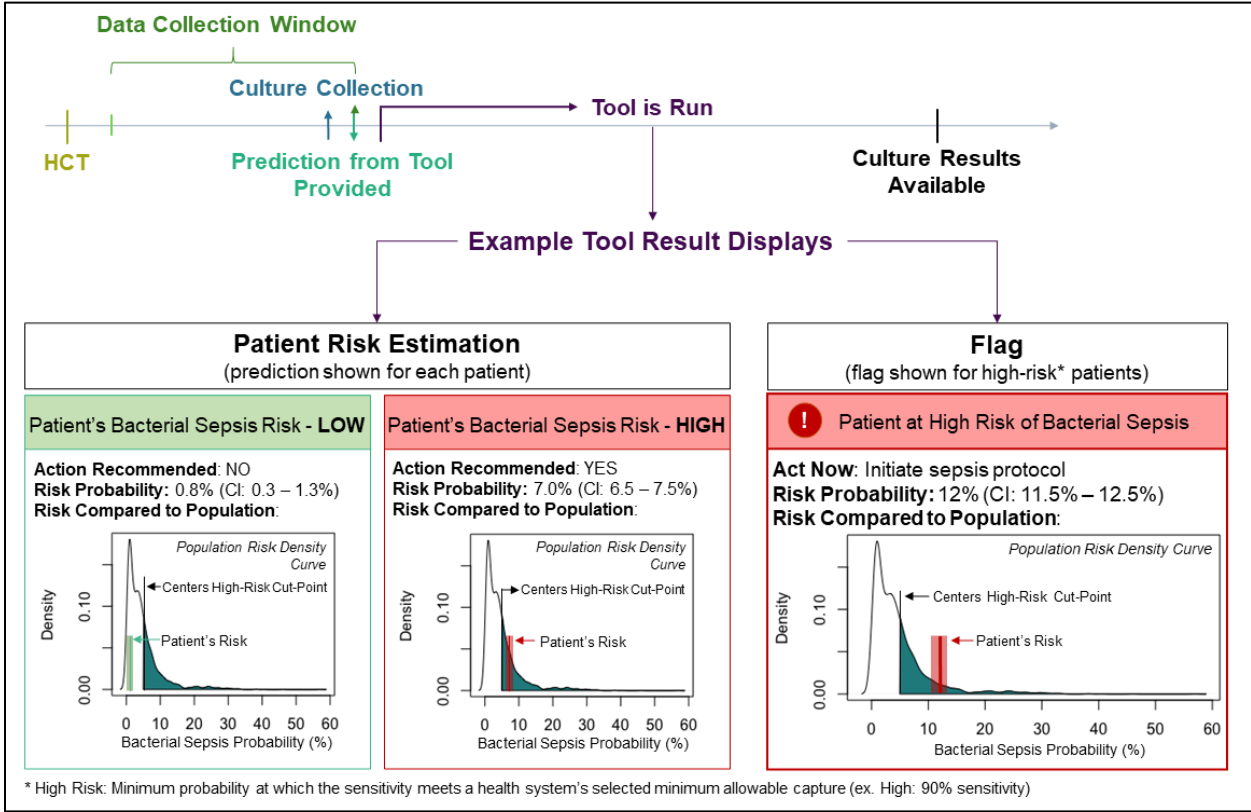
^b 95% Confidence Interval estimated using 2000 stratified bootstrapped replicates

^c 95% Confidence Intervals estimated using generalized estimating equations with robust standard errors

^d Based on Upper Left Method Optimal Cut-points from primary analysis

Supplemental Paper 2: Development and Validation of a Machine Learning Model to Predict Bacterial Sepsis among Immunocompromised Stem Cell Transplant Recipients

eFigure 1. Bacterial Sepsis Prediction Tool Schematic



Outcome Justification

We selected bacteremia with high risk for sepsis - blood culture with gram-negative species, *Staphylococcus aureus*, or *Streptococcus* species bacteremia – based on evidence that sepsis is significantly more likely among HCT recipients with PBIs and the selected bloodstream infections than without. Neutropenic HCT recipients with positive blood cultures have been found to be 9.6 times and recipients with gram-negative bacteremia 12.9 times more likely to have severe sepsis than individuals with negative blood cultures.¹ Additionally, previous studies show gram-negative bacteremia results in significantly higher C-reactive protein levels (a recognized indicator of sepsis²) and is associated with a significantly higher risk of severe sepsis and non-relapse mortality than gram-positive bacteremia.^{1,3}

Potential Predictors and Missingness

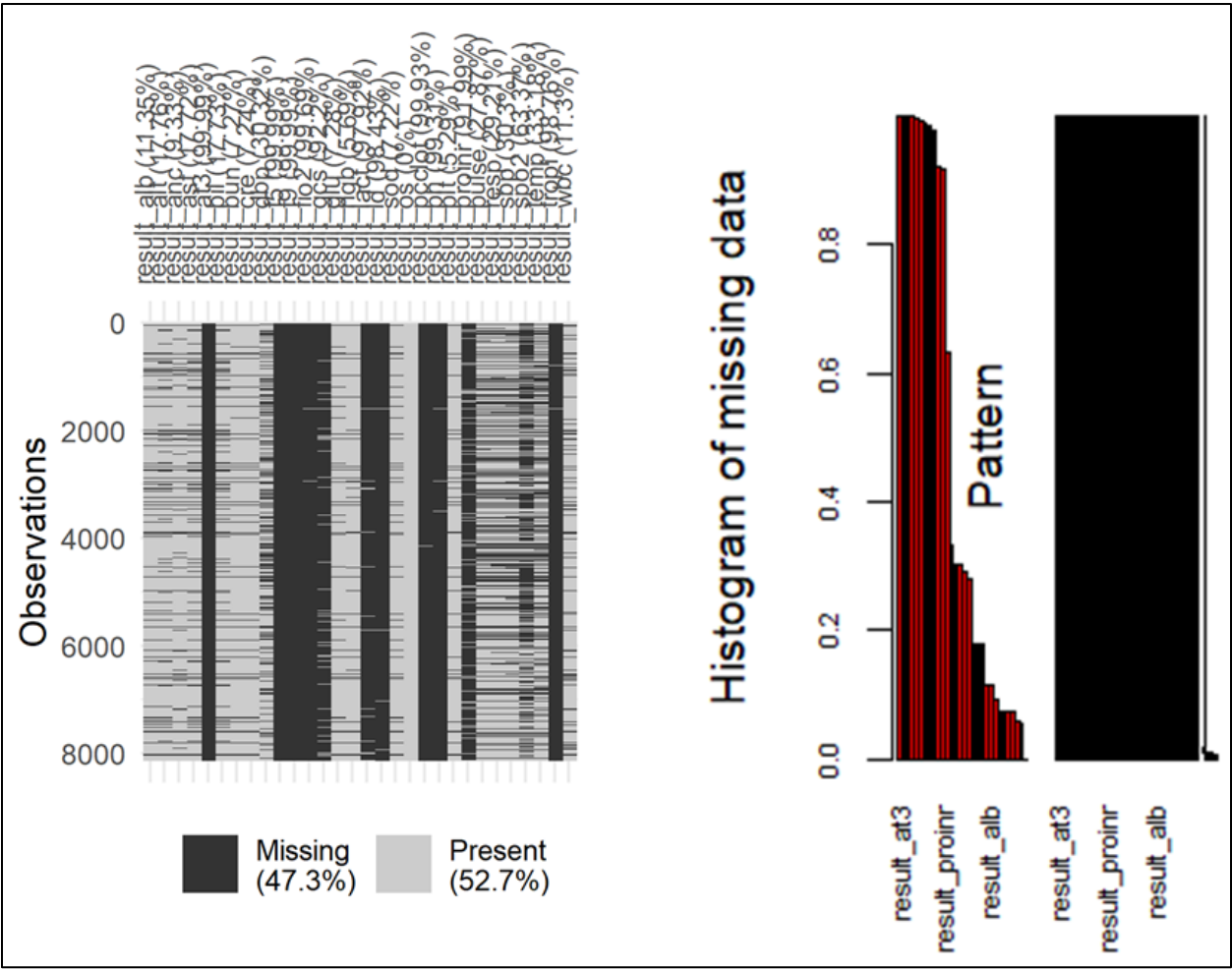
eTable 1. Frequency of Potential Predictors Screened for Model Inclusion

Factor	Considered in:	Included in:	Missing	Range	Median/Most Frequent Category	IQR/Percent of Most Frequent
Demographic						
Age	C-SHBSL/SHBSL	C-SHBSL/SHBSL	0, 0%	17.9, 80.4	53.1	40.6, 61.9
Ethnicity/Race	C-SHBSL/SHBSL	C-SHBSL/SHBSL	0, 0%	-	White	78.8%
Gender	C-SHBSL/SHBSL	C-SHBSL/SHBSL	0, 0%	-	Male	57.5%
Body Structure						
BMI at time of transplant	C-SHBSL/SHBSL	C-SHBSL/SHBSL	0, 0%	12.8, 44	24.7	21.7, 28.1
Vitals and Biomarkers						
Albumin	SHBSL	SHBSL	923, 11.4%	1, 4.9	3.5	3.1, 3.8
Alanine Aminotransferase	SHBSL	SHBSL	1444, 17.8%	3, 9380	26	16, 46
Antithrombin III	SHBSL		8130, 100%	136, 136	136	136, 136
Asparate Transaminase	SHBSL	SHBSL	1441, 17.7%	4, 11290	21	16, 29
Bicarbonate	SHBSL		8131, 100%	-	-	-
Bilirubin	SHBSL	SHBSL	1442, 17.7%	0.1, 26.5	0.7	0.5, 1
Blood Urea Nitrogen	SHBSL	SHBSL	591, 7.3%	1, 159	22	16, 29
C-Reactive Protein	SHBSL		8131, 100%	-	-	-
Creatinine	SHBSL	SHBSL	589, 7.2%	0.2, 8.9	1	0.8, 1.2
Diastolic Blood Pressure	C-SHBSL/SHBSL	C-SHBSL/SHBSL	2465, 30.3%	13, 134	79	72, 86
Factor IX	SHBSL		8130, 100%	90, 90	90	90, 90
Factor V	SHBSL		8130, 100%	38, 38	38	38, 38
Glasgow Coma Score	C-SHBSL/SHBSL		7497, 92.2%	4, 15	15	15, 15
Glucose	SHBSL	SHBSL	592, 7.3%	54, 556	132	111, 164
Heart Rate (Pulse)	C-SHBSL/SHBSL	C-SHBSL/SHBSL	2266, 27.9%	11, 170	90	79, 103
Hemoglobin	SHBSL	SHBSL	463, 5.7%	5.7, 16	10.2	9.4, 11.3
International Normalized Ratio	SHBSL		8131, 100%	-	-	-
Interleukin-6	SHBSL		8131, 100%	-	-	-
Lactate (volumn) within last 24 hours	SHBSL		7962, 97.9%	0.3, 11.8	1.3	0.9, 2.1
Lactate Dehydrogenase	SHBSL		8004, 98.4%	65, 1495	51	25.5, 78.5
Neutrophil	SHBSL	SHBSL	759, 9.3%	0, 43	3.8	1.9, 6.5

Factor	Considered in:	Included in:	Missing	Range	Median/Most Frequent Category	IQR/Percent of Most Frequent
Oxygen Saturation	C-SHBSL/SHBSL		5153, 63.4%	84, 100	97	95, 98
Plasminogen activator inhibitor 1	SHBSL		8131, 100%	-	-	-
Plasminogen activity	SHBSL		8131, 100%	-	-	-
Platelet	SHBSL	SHBSL	430, 5.3%	1, 745	88	46, 145
Protein C Activity	SHBSL		8125, 99.9%	167, 242	201	195.8, 212.2
Protein S Activity	SHBSL		8131, 100%	-	-	-
Prothrombin	SHBSL		7480, 92%	0.8, 2.5	1.1	1, 1.2
Respiratory Rate	C-SHBSL/SHBSL	C-SHBSL/SHBSL	2375, 29.2%	6, 56	16	16, 18
Sodium	SHBSL	SHBSL	587, 7.2%	117, 152	136	134, 138
Supplemental Oxygen (Y/N)	C-SHBSL/SHBSL	C-SHBSL/SHBSL	0, 0%	0, 100	0	0, 0
Systolic Blood Pressure	C-SHBSL/SHBSL	C-SHBSL/SHBSL	2464, 30.3%	73, 206	128	116, 139.5
Temperature	C-SHBSL/SHBSL	C-SHBSL/SHBSL	2698, 33.2%	34.5, 40.1	36.9	36.6, 37.3
Troponin	SHBSL					
White Blood Cell Count	SHBSL	SHBSL	919, 11.3%	0.1, 19.9	5	2.8, 8.1
Recent Events						
Previous Bacterial Sepsis Event	SHBSL	SHBSL	0, 0%	0, 4	0	0, 0
Antibiotics within last 14 days	SHBSL	SHBSL	0, 0%	0, 1	1	0, 1
Days Since Transplant	SHBSL	SHBSL	0, 0%	0, 100	54	34, 74
Time of Day (in hours 0-24)	SHBSL	SHBSL	0, 0%	0, 24	10	9, 12
Additional Factors						
Location (IUC, inpatient, v outpatient)	SHBSL	SHBSL	0, 0%	-	Outpatient	74.4%
Acute GVHD ^a	SHBSL	SHBSL	0, 0%	0, 4	2	2, 2
Acute Gut GVHD ^a	SHBSL	SHBSL	0, 0%	0, 4	1	1, 1
Acute Skin GVHD ^a	SHBSL	SHBSL	0, 0%	0, 4	1	0, 3
Acute Liver GVHD ^a	SHBSL	SHBSL	0, 0%	0, 4	0	0, 0
Donor Relation (related/unrelated)	SHBSL	SHBSL	0, 0%	-	Not Related	69.9%
HCT Cell Donor Type (ex. BM/PBSC)	SHBSL	SHBSL	0, 0%	-	Unrelated	69.9%
Underlying disease/cancer type	SHBSL	SHBSL	0, 0%	-	ANL	33.2%
Disease Status	SHBSL	SHBSL	0, 0%	-	Remission	54.8%

^aGVHD: Graft vs. Host Disease

eFigure 2: Missingness Frequency and Pattern for Potential Model Predictors



Imputation Approach

Given this data limitation, we employed a two-step missing data approach. First, we restricted to predictor that had measurements for at least 65% of the examined PBIs. Second, we filled the remaining missing measurement observations using a single, normal value imputation approach. Specially for each missing value, we randomly selected a value from the normal biological range of that predictor. The specific normal value rangers used can be found in are e.Table2.

e.Table 2: Normal Value Ranges for Single, Normal Value Imputation

	Normal Value Range	Source
Temperature	36.1 - 37.2°C	MedlinePlus
Respiratory Rate	12-20 breaths/minute	Cleveland Clinic
Pulse (Heart Rate)	60-100 beats/minute	Mayo Clinic
Systolic Blood Pressure	90-120 mmHg	Havard, Tolulope et al.
Diastolic Blood Pressure	60-80 mmHg	Stensrud et al.
Oxygen Saturation	95-100%	Mayo Clinic
White Blood Cell Count	4.4 - 11 billion cells/L	UpToDate
Albumin	3.4 - 5.4 g/dL	Rochester
Alanine Aminotransaminase	7 - 55 u/l	Mayo Clinic
Asparate Transaminase	8 - 48 u/l	Mayo Clinic
Bilirubin	0.1 - 1.2 mg/dL	Mayo Clinic
Blood Urea Nitrogen	7 - 20 mg/dL	Mayo Clinic
Creatinine	0.84 - 1.21 mg/dL	Mayo Clinic
Glucose	62 - 125 mg/dL	Mindscape
Sodium	135 - 145 mEq/L	Mayo Clinic
Absolute Neutrophil Count	1.5 - 8 n/nL	Healthline
Platelet	150 - 450 p/nL	Hopkins Medicine
Hemoglobin		Mayo Clinic
Female	12- 15.5 g/dL	
Male	13.5 - 17.5 g/dL	
Glasgow Coma Scale Score ^a	15	MdCalc

^a For existing criteria comparisons

Detailed Description of Super Learner Library

Screens: predictor selection step that can be applied to included algorithms thus reducing the predictors included in algorithm learning

- Correlation pvalue (corP): predictors with a pearson correlation pvalue of 0.05 or less are included (user edited)
- Regularization (glmnet): performs L1 (Lasso) penalized regression and keeps predictors with coefficients larger than zero (original)

Algorithms: varying algorithms and algorithm forms (screened and hypermeters)

- Linear Models
 - Logistic regression: SL.glm
 - Base
 - Screen: corP & glmnet
 - Stepwise regression: SL.stepAIC
 - Base
 - Generalized additive model: SL.gam
 - Base
 - Screen: corP & glmnet
 - Penalized generalized linear model using elastic net: SL.glmnet
 - Base: regularization L1 (alpha 1)
 - Hyperparameter
 - Regularization: L2 (alpha 0)
 - Screen: corP
- Network and Trees
 - Single Node Neural Network: SL.nnet
 - Hyperparameters: Decays (regularization): 0.9, Size (number of units in hidden layer): 6
 - Screen: corP
 - Recursive Binary Split Tree (Classification): SL.rpart
 - Base
 - Random Forest (classification tree): SL.randomForest
 - Base
 - Screen: corP & glmnet
 - Hyperparameters: group size of 100
 - Screen: corP & glmnet
 - Bagging classification tree: SL.ipredbagg
 - Base
 - Screen: corP

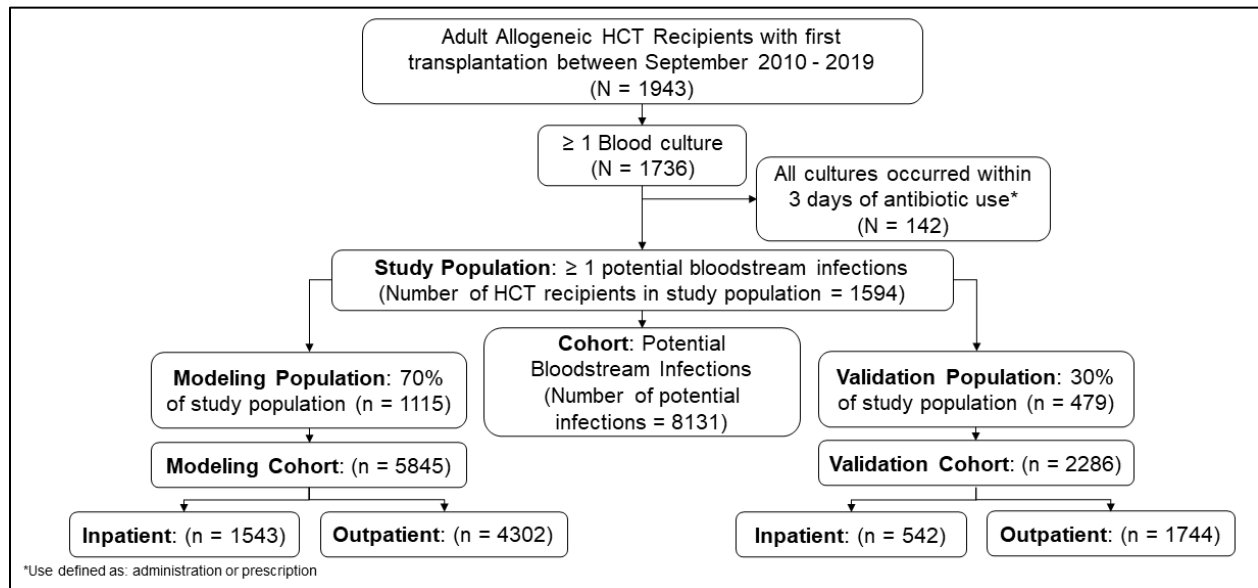
The parameters used in the base form of each algorithm can be found in Dr. Eric Polley's Github SuperLearner R code repository.

Net Reclassification

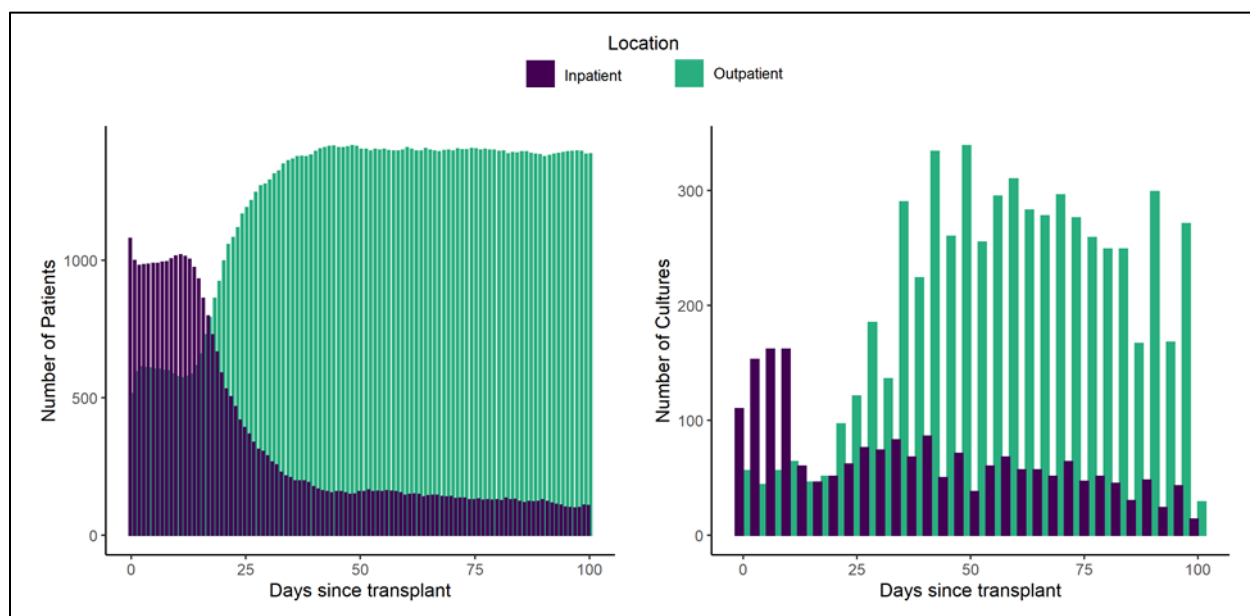
Net Reclassification Index (NRI) is a measure of tool discrimination that works by comparing how patients distributed into risk probability categories by two models (say SIRS and SHBSL) actually fare (in relation to a specific outcome – such as high-risk bacteremia). The NRI conducts this comparison by measuring the net improvement of appropriate patient placement (i.e. events in higher prediction categories) of one model (say SHBSL) over the other (say SIRS).

To calculate this metric, we considered SHBSL to be the “updated” model and the existing tool or C-SHBSL to be the “base” model. In order to get risk probability categories for existing tools, we estimated the risk probabilities associated with each score for SIRS, NEWS, and qSOFA using single variable logistic regressions. Using the resulting risk probability categories, we estimated the reclassification of SHBSL in reference to the existing tools. In order to calculate the NRI between SHBSL and C-SHBSL, we defined our categories as the quartiles of C-SHBSL’s risk probabilities.

eFigure 3. Flowchart of Hematopoietic Cell Transplant Recipient (HCT) Study Population and Potential Bloodstream Infection Cohort



eFigure 4. Histograms of Location of Patient and Collected Cultures by Day Since Transplant



Patient location was defined using full calendar days and patients were considered to be inpatient for the full day of hospital admission and discharge. Culture location was defined based on the patient location for each day.

eTable 3: Population Demographics and Transplant Factors by Randomly Assigned Modeling and Validation Dataset

	Modeling	Validation
	N = 1115	N = 479
Calendar Year at Tx (Median, IQR)	2014 (2012, 2016)	2014 (2012, 2016)
Age at Tx (Median, IQR)	55 (43, 63)	54 (41, 62)
Gender (Count, % male)	644 (57.8)	267 (55.7)
Ethnicity/Race (Count., %)		
American Indian or Alaskan Native ^b	19 (1.7)	5 (1.0)
Asian ^b	73 (6.5)	46 (9.6)
Black ^b	18 (1.6)	7 (1.5)
Multiple ^b	16 (1.4)	6 (1.3)
Native Hawaiian or other Pacific Islander ^b	20 (1.8)	5 (1.0)
Unknown ^b	35 (3.1)	19 (4.0)
White ^b	888 (79.6)	354 (73.9)
LatinX	46 (4.1)	37 (7.7)
Underlying Disease		
Acute Lymphoblastic Leukemia	156 (14.0)	56 (11.7)
Acute Nonlymphocytic Leukemia	404 (36.2)	176 (36.7)
Myelodysplastic Syndromes	297 (26.6)	113 (23.6)
Other	258 (23.1)	134 (28.0)
HCT Cell Donor Type		
Bone marrow	90 (8.1)	36 (7.5)
Bone marrow, peripheral blood stem cells	3 (0.3)	0 (0)
Umbilical cord blood	123 (11.0)	56 (11.7)
Peripheral blood stem cells	899 (80.6)	387 (80.8)
Donor Relation (Count, %)		
Child	22 (2.0)	16 (3.3)
Half Sibling	3 (0.3)	2 (0.4)
Not Related	781 (70.0)	326 (68.1)
Parent	11 (1.0)	6 (1.3)
Relative	1 (0.1)	0 (0)
Sibling	297 (26.6)	129 (26.9)
Acute GVHD (Count, %)		
Grade 0	238 (21.3)	90 (18.8)
Grade 1	57 (5.1)	21 (4.4)
Grade 2	649 (58.2)	303 (63.3)
Grade 3	129 (11.6)	50 (10.4)
Grade 4	42 (3.8)	15 (3.1)
Died (n, %) ^a	96 (8.6)	42 (8.8)

^a Experiences during follow-up

^b Non-Hispanic

e.Table 4: Discrimination - Predictive Accuracy of Sepsis Prediction Tools in Relation to High Sepsis-Risk Bacteremia^a

	Cut Points	Sensitivity^a	Specificity^a	Positive Predictive Value^a	Negative Predictive Value^a
Clinical Factor Specific - C-SHBSL					
	0.0215	100 (94.9, 100%)	1.8 (1.3, 2.4%)	3.1 (2.4, 3.9%)	100 (91.0, 100%)
	0.0310	98.6 (92.3, 100%)	9.9 (8.7, 11.2%)	3.3 (2.6, 4.2%)	99.5 (97.5, 100%)
	0.0407	97.1 (90.1, 99.7%)	21.7 (20.0, 23.5%)	3.8 (2.9, 4.8%)	99.6 (98.5, 99.9%)
	0.0436	95.7 (88.0, 99.1%)	24.4 (22.6, 26.3%)	3.8 (3.0, 4.9%)	99.4 (98.4, 99.9%)
	0.0439	94.3 (86.0, 98.4%)	24.8 (23.0, 26.7%)	3.8 (3.0, 4.8%)	99.3 (98.2, 99.8%)
	0.0439	92.9 (84.1, 97.6%)	24.9 (23.1, 26.8%)	3.8 (2.9, 4.8%)	99.1 (97.9, 99.7%)
	0.0450	91.4 (82.3, 96.8%)	26.3 (24.4, 28.1%)	3.8 (2.9, 4.8%)	99.0 (97.8, 99.6%)
	0.0497	90.0 (80.5, 95.9%)	31.5 (29.6, 33.5%)	4.0 (3.1, 5.1%)	99.0 (98.0, 99.6%)
	0.0502	88.6 (78.7, 94.9%)	31.9 (30.0, 33.9%)	3.9 (3.0, 5.0%)	98.9 (97.8, 99.5%)
	0.0506	87.1 (77.0, 93.9%)	32.3 (30.3, 34.3%)	3.9 (3.0, 5.0%)	98.8 (97.7, 99.4%)
	0.0522	85.7 (75.3, 92.9%)	34.0 (32.1, 36.0%)	3.9 (3.0, 5.0%)	98.7 (97.6, 99.4%)
	0.0531	84.3 (73.6, 91.9%)	34.8 (32.9, 36.9%)	3.9 (3.0, 5.0%)	98.6 (97.5, 99.3%)
	0.0548	82.9 (72.0, 90.8%)	36.6 (34.6, 38.6%)	4.0 (3.0, 5.1%)	98.5 (97.5, 99.2%)
	0.0580	81.4 (70.3, 89.7%)	40.0 (37.9, 42.1%)	4.1 (3.1, 5.3%)	98.6 (97.5, 99.2%)
	0.0582	80.0 (68.7, 88.6%)	40.3 (38.2, 42.3%)	4.1 (3.1, 5.2%)	98.5 (97.4, 99.2%)
	0.0589	78.6 (67.1, 87.5%)	41.0 (38.9, 43.1%)	4.0 (3.1, 5.2%)	98.4 (97.3, 99.1%)
	0.0651	77.1 (65.6, 86.3%)	47.1 (45.0, 49.2%)	4.4 (3.3, 5.7%)	98.5 (97.6, 99.1%)
	0.0741	75.7 (64.0, 85.2%)	54.3 (52.2, 56.4%)	5.0 (3.7, 6.5%)	98.6 (97.8, 99.2%)
	0.0787	74.3 (62.4, 84.0%)	57.4 (55.4, 59.5%)	5.2 (3.9, 6.8%)	98.6 (97.8, 99.2%)
	0.0823	72.9 (60.9, 82.8%)	59.9 (57.9, 62.0%)	5.4 (4.1, 7.1%)	98.6 (97.8, 99.1%)
	0.0836	71.4 (59.4, 81.6%)	61.2 (59.2, 63.3%)	5.5 (4.1, 7.2%)	98.5 (97.8, 99.1%)
	0.0843	70.0 (57.9, 80.4%)	61.6 (59.5, 63.6%)	5.4 (4.1, 7.1%)	98.5 (97.7, 99.1%)
	0.0853	68.6 (56.4, 79.1%)	62.5 (60.4, 64.5%)	5.5 (4.0, 7.2%)	98.4 (97.6, 99.0%)
	0.0896	67.1 (54.9, 77.9%)	65.0 (63.0, 67.0%)	5.7 (4.2, 7.5%)	98.4 (97.7, 99.0%)
	0.0926	65.7 (53.4, 76.7%)	66.9 (64.9, 68.9%)	5.9 (4.4, 7.8%)	98.4 (97.6, 99.0%)
	0.0950	64.3 (51.9, 75.4%)	68.3 (66.3, 70.2%)	6.0 (4.4, 8.0%)	98.4 (97.6, 98.9%)
	0.0961	62.9 (50.5, 74.1%)	68.7 (66.7, 70.6%)	6.0 (4.4, 7.9%)	98.3 (97.5, 98.9%)
	0.0985	61.4 (49.0, 72.8%)	69.4 (67.5, 71.4%)	6.0 (4.4, 8.0%)	98.3 (97.5, 98.9%)
	0.1002	60.0 (47.6, 71.5%)	70.4 (68.5, 72.3%)	6.0 (4.4, 8.1%)	98.2 (97.5, 98.8%)
	0.1044	58.6 (46.2, 70.2%)	73.0 (71.1, 74.8%)	6.4 (4.6, 8.6%)	98.2 (97.5, 98.8%)
	0.1081	57.1 (44.7, 68.9%)	74.7 (72.9, 76.5%)	6.7 (4.8, 9.0%)	98.2 (97.5, 98.8%)
	0.1110	55.7 (43.3, 67.6%)	76.0 (74.2, 77.8%)	6.8 (4.9, 9.2%)	98.2 (97.4, 98.8%)
	0.1120	54.3 (41.9, 66.3%)	76.6 (74.8, 78.3%)	6.8 (4.9, 9.2%)	98.1 (97.4, 98.7%)
	0.1146	52.9 (40.6, 64.9%)	77.8 (76.0, 79.5%)	7.0 (5.0, 9.5%)	98.1 (97.4, 98.7%)
	0.1223	51.4 (39.2, 63.6%)	80.7 (79.0, 82.4%)	7.8 (5.5, 10.6%)	98.1 (97.4, 98.7%)

Cut Points	Sensitivity^a	Specificity^a	Positive Predictive Value^a	Negative Predictive Value^a
0.1223	50.0 (37.8, 62.2%)	80.8 (79.1, 82.4%)	7.6 (5.3, 10.4%)	98.1 (97.3, 98.7%)
0.1227	48.6 (36.4, 60.8%)	81.0 (79.3, 82.6%)	7.5 (5.2, 10.3%)	98.0 (97.3, 98.6%)
0.1251	47.1 (35.1, 59.4%)	81.6 (79.9, 83.2%)	7.5 (5.2, 10.3%)	98.0 (97.2, 98.6%)
0.1284	45.7 (33.7, 58.1%)	82.7 (81.1, 84.3%)	7.7 (5.3, 10.7%)	98.0 (97.2, 98.6%)
0.1297	44.3 (32.4, 56.7%)	83.2 (81.5, 84.7%)	7.7 (5.3, 10.7%)	97.9 (97.2, 98.5%)
0.1352	42.9 (31.1, 55.3%)	84.7 (83.2, 86.2%)	8.2 (5.6, 11.4%)	97.9 (97.2, 98.5%)
0.1513	41.4 (29.8, 53.8%)	88.8 (87.4, 90.0%)	10.4 (7.1, 14.6%)	98.0 (97.2, 98.5%)
0.1590	40.0 (28.5, 52.4%)	90.5 (89.2, 91.7%)	11.7 (7.9, 16.5%)	97.9 (97.2, 98.5%)
0.1747	38.6 (27.2, 51.0%)	93.0 (91.9, 94.0%)	14.8 (10, 20.8%)	98.0 (97.3, 98.5%)
0.1759	37.1 (25.9, 49.5%)	93.2 (92.1, 94.2%)	14.7 (9.8, 20.8%)	97.9 (97.2, 98.5%)
0.1780	35.7 (24.6, 48.1%)	93.6 (92.5, 94.6%)	15.0 (9.9, 21.3%)	97.9 (97.2, 98.4%)
0.1786	34.3 (23.3, 46.6%)	93.7 (92.6, 94.7%)	14.6 (9.6, 21.0%)	97.8 (97.1, 98.4%)
0.1799	32.9 (22.1, 45.1%)	93.9 (92.8, 94.9%)	14.6 (9.5, 21.0%)	97.8 (97.1, 98.4%)
0.1857	31.4 (20.9, 43.6%)	94.7 (93.7, 95.6%)	15.7 (10.1, 22.8%)	97.8 (97.0, 98.3%)
0.1877	30.0 (19.6, 42.1%)	94.9 (93.9, 95.8%)	15.7 (10.0, 23.0%)	97.7 (97.0, 98.3%)
0.1984	28.6 (18.4, 40.6%)	96.5 (95.7, 97.2%)	20.6 (13.1, 30.0%)	97.7 (97.0, 98.3%)
0.1996	27.1 (17.2, 39.1%)	96.6 (95.8, 97.3%)	20.2 (12.6, 29.8%)	97.7 (97.0, 98.3%)
0.2009	25.7 (16.0, 37.6%)	96.7 (95.9, 97.4%)	19.8 (12.2, 29.4%)	97.6 (96.9, 98.2%)
0.2012	24.3 (14.8, 36.0%)	96.8 (96.0, 97.5%)	19.3 (11.7, 29.1%)	97.6 (96.9, 98.2%)
0.2024	22.9 (13.7, 34.4%)	96.9 (96.1, 97.6%)	18.8 (11.2, 28.8%)	97.5 (96.8, 98.2%)
0.2088	21.4 (12.5, 32.9%)	97.5 (96.7, 98.1%)	21.1 (12.3, 32.4%)	97.5 (96.8, 98.1%)
0.2093	20.0 (11.4, 31.3%)	97.5 (96.7, 98.1%)	20.0 (11.4, 31.3%)	97.5 (96.7, 98.1%)
0.2155	18.6 (10.3, 29.7%)	97.9 (97.2, 98.5%)	22.0 (12.3, 34.7%)	97.4 (96.7, 98.1%)
0.2208	17.1 (9.2, 28.0%)	98.3 (97.7, 98.8%)	24.0 (13.1, 38.2%)	97.4 (96.7, 98.0%)
0.2208	15.7 (8.1, 26.4%)	98.3 (97.7, 98.8%)	22.4 (11.8, 36.6%)	97.4 (96.6, 98.0%)
0.2220	14.3 (7.1, 24.7%)	98.4 (97.8, 98.9%)	21.7 (10.9, 36.4%)	97.3 (96.6, 97.9%)
0.2237	12.9 (6.1, 23.0%)	98.5 (97.9, 98.9%)	20.9 (10.0, 36.0%)	97.3 (96.5, 97.9%)
0.2404	11.4 (5.1, 21.3%)	99.5 (99.1, 99.7%)	40.0 (19.1, 63.9%)	97.3 (96.5, 97.9%)
0.2458	10.0 (4.1, 19.5%)	99.6 (99.3, 99.8%)	46.7 (21.3, 73.4%)	97.2 (96.5, 97.9%)
0.2471	8.6 (3.2, 17.7%)	99.6 (99.3, 99.8%)	42.9 (17.7, 71.1%)	97.2 (96.4, 97.8%)
0.2499	7.1 (2.4, 15.9%)	99.7 (99.4, 99.9%)	41.7 (15.2, 72.3%)	97.1 (96.4, 97.8%)
0.2527	5.7 (1.6, 14.0%)	99.7 (99.4, 99.9%)	36.4 (10.9, 69.2%)	97.1 (96.3, 97.7%)
0.2687	4.3 (0.9, 12.0%)	99.9 (99.7, 100%)	60.0 (14.7, 94.7%)	97.1 (96.3, 97.7%)
0.2807	2.9 (0.3, 9.9%)	100 (99.7, 100%)	66.7 (9.4, 99.2%)	97.0 (96.2, 97.7%)
0.2890	1.4 (0.0, 7.7%)	100 (99.7, 100%)	50.0 (1.3, 98.7%)	97.0 (96.2, 97.6%)
0.3014	0.0 (0.0, 5.1%)	100 (99.7, 100%)	-	96.9 (96.1, 97.6%)
Full Predictor-SHBSL				
0.0057	100 (94.9, 100%)	9.9 (8.7, 11.2%)	3.4 (2.6, 4.3%)	100 (98.3, 100%)

Cut Points	Sensitivity^a	Specificity^a	Positive Predictive Value^a	Negative Predictive Value^a
0.0130	98.6 (92.3, 100%)	29.8 (27.9, 31.8%)	4.2 (3.3, 5.3%)	99.8 (99.2, 100%)
0.0194	97.1 (90.1, 99.7%)	37.1 (35.1, 39.2%)	4.7 (3.6, 5.9%)	99.8 (99.1, 100%)
0.0194	95.7 (88.0, 99.1%)	37.1 (35.1, 39.2%)	4.6 (3.6, 5.8%)	99.6 (98.9, 99.9%)
0.0256	94.3 (86.0, 98.4%)	45.2 (43.1, 47.3%)	5.2 (4.0, 6.5%)	99.6 (99.0, 99.9%)
0.0363	92.9 (84.1, 97.6%)	60.4 (58.4, 62.5%)	6.9 (5.4, 8.7%)	99.6 (99.1, 99.9%)
0.0379	91.4 (82.3, 96.8%)	62.6 (60.5, 64.6%)	7.2 (5.6, 9.1%)	99.6 (99.1, 99.8%)
0.0392	90.0 (80.5, 95.9%)	64.0 (62.0, 66%)	7.3 (5.7, 9.3%)	99.5 (99.0, 99.8%)
0.0397	88.6 (78.7, 94.9%)	64.4 (62.4, 66.4%)	7.3 (5.6, 9.2%)	99.4 (98.9, 99.8%)
0.0409	87.1 (77.0, 93.9%)	65.3 (63.3, 67.3%)	7.3 (5.7, 9.3%)	99.4 (98.8, 99.7%)
0.0440	85.7 (75.3, 92.9%)	68.5 (66.6, 70.5%)	7.9 (6.1, 10.1%)	99.3 (98.8, 99.7%)
0.0442	84.3 (73.6, 91.9%)	68.6 (66.6, 70.5%)	7.8 (6.0, 10.0%)	99.3 (98.7, 99.6%)
0.0449	82.9 (72.0, 90.8%)	69.1 (67.1, 71.0%)	7.8 (6.0, 10.0%)	99.2 (98.6, 99.6%)
0.0468	81.4 (70.3, 89.7%)	70.5 (68.6, 72.4%)	8.0 (6.1, 10.3%)	99.2 (98.6, 99.6%)
0.0498	80.0 (68.7, 88.6%)	72.8 (70.9, 74.6%)	8.5 (6.5, 10.9%)	99.1 (98.6, 99.5%)
0.0502	78.6 (67.1, 87.5%)	73.2 (71.3, 75.0%)	8.5 (6.4, 10.9%)	99.1 (98.5, 99.5%)
0.0506	77.1 (65.6, 86.3%)	73.4 (71.5, 75.2%)	8.4 (6.4, 10.8%)	99.0 (98.4, 99.4%)
0.0509	75.7 (64.0, 85.2%)	73.6 (71.7, 75.4%)	8.3 (6.3, 10.7%)	99.0 (98.4, 99.4%)
0.0512	74.3 (62.4, 84.0%)	73.9 (72.0, 75.7%)	8.2 (6.2, 10.7%)	98.9 (98.3, 99.4%)
0.0553	72.9 (60.9, 82.8%)	77.3 (75.5, 79.0%)	9.2 (6.9, 11.9%)	98.9 (98.3, 99.3%)
0.0577	71.4 (59.4, 81.6%)	79.1 (77.3, 80.7%)	9.7 (7.3, 12.6%)	98.9 (98.3, 99.3%)
0.0654	70.0 (57.9, 80.4%)	83.4 (81.8, 84.9%)	11.8 (8.8, 15.2%)	98.9 (98.3, 99.3%)
0.0663	68.6 (56.4, 79.1%)	83.7 (82.1, 85.2%)	11.7 (8.8, 15.2%)	98.8 (98.2, 99.3%)
0.0767	67.1 (54.9, 77.9%)	87.6 (86.2, 89.0%)	14.6 (11.0, 19.0%)	98.8 (98.2, 99.3%)
0.0786	65.7 (53.4, 76.7%)	88.1 (86.7, 89.4%)	14.8 (11.1, 19.3%)	98.8 (98.2, 99.2%)
0.0797	64.3 (51.9, 75.4%)	88.5 (87.1, 89.8%)	15.0 (11.2, 19.6%)	98.7 (98.1, 99.2%)
0.0835	62.9 (50.5, 74.1%)	89.3 (87.9, 90.5%)	15.6 (11.6, 20.4%)	98.7 (98.1, 99.2%)
0.0851	61.4 (49.0, 72.8%)	89.7 (88.3, 90.9%)	15.8 (11.7, 20.7%)	98.7 (98.1, 99.1%)
0.0902	60.0 (47.6, 71.5%)	90.4 (89.1, 91.6%)	16.5 (12.1, 21.6%)	98.6 (98.0, 99.1%)
0.0910	58.6 (46.2, 70.2%)	90.6 (89.3, 91.8%)	16.4 (12.0, 21.6%)	98.6 (98.0, 99.0%)
0.0911	57.1 (44.7, 68.9%)	90.6 (89.3, 91.8%)	16.1 (11.7, 21.2%)	98.5 (97.9, 99.0%)
0.0913	55.7 (43.3, 67.6%)	90.6 (89.3, 91.8%)	15.7 (11.4, 20.9%)	98.5 (97.8, 99.0%)
0.0921	54.3 (41.9, 66.3%)	90.8 (89.5, 92.0%)	15.7 (11.4, 20.9%)	98.4 (97.8, 98.9%)
0.0969	52.9 (40.6, 64.9%)	91.4 (90.2, 92.6%)	16.3 (11.7, 21.8%)	98.4 (97.8, 98.9%)
0.0984	51.4 (39.2, 63.6%)	91.8 (90.6, 92.9%)	16.5 (11.8, 22.1%)	98.4 (97.7, 98.9%)
0.0992	50.0 (37.8, 62.2%)	92.1 (90.9, 93.2%)	16.7 (11.9, 22.4%)	98.3 (97.7, 98.8%)
0.1041	48.6 (36.4, 60.8%)	93.1 (92.0, 94.1%)	18.2 (12.9, 24.5%)	98.3 (97.6, 98.8%)
0.1045	47.1 (35.1, 59.4%)	93.1 (92.0, 94.1%)	17.7 (12.5, 24.0%)	98.2 (97.6, 98.8%)
0.1060	45.7 (33.7, 58.1%)	93.3 (92.2, 94.3%)	17.7 (12.4, 24.0%)	98.2 (97.5, 98.7%)

	Cut Points	Sensitivity ^a	Specificity ^a	Positive Predictive Value ^a	Negative Predictive Value ^a
	0.1085	44.3 (32.4, 56.7%)	93.4 (92.3, 94.4%)	17.5 (12.2, 23.9%)	98.2 (97.5, 98.7%)
	0.1146	42.9 (31.1, 55.3%)	94.2 (93.1, 95.1%)	18.9 (13.1, 25.8%)	98.1 (97.4, 98.7%)
	0.1163	41.4 (29.8, 53.8%)	94.4 (93.4, 95.4%)	19.1 (13.2, 26.2%)	98.1 (97.4, 98.6%)
	0.1278	40.0 (28.5, 52.4%)	95.3 (94.3, 96.1%)	21.1 (14.5, 29.0%)	98.0 (97.4, 98.6%)
	0.1293	38.6 (27.2, 51.0%)	95.3 (94.3, 96.1%)	20.6 (14.0, 28.6%)	98.0 (97.3, 98.6%)
	0.1348	37.1 (25.9, 49.5%)	95.8 (94.8, 96.6%)	21.7 (14.7, 30.1%)	98.0 (97.3, 98.5%)
	0.1367	35.7 (24.6, 48.1%)	95.9 (95.0, 96.7%)	21.6 (14.5, 30.1%)	97.9 (97.2, 98.5%)
	0.1418	34.3 (23.3, 46.6%)	96.2 (95.3, 96.9%)	22.0 (14.6, 31.0%)	97.9 (97.2, 98.4%)
	0.1624	32.9 (22.1, 45.1%)	97.1 (96.3, 97.7%)	26.1 (17.3, 36.6%)	97.9 (97.2, 98.4%)
	0.1676	31.4 (20.9, 43.6%)	97.2 (96.4, 97.8%)	25.9 (17.0, 36.5%)	97.8 (97.1, 98.4%)
	0.1702	30.0 (19.6, 42.1%)	97.3 (96.5, 97.9%)	25.9 (16.8, 36.9%)	97.8 (97.1, 98.4%)
	0.1900	28.6 (18.4, 40.6%)	97.7 (97.0, 98.3%)	28.2 (18.1, 40.1%)	97.7 (97.0, 98.3%)
	0.2071	27.1 (17.2, 39.1%)	98.1 (97.5, 98.7%)	31.7 (20.3, 45.0%)	97.7 (97.0, 98.3%)
	0.2159	25.7 (16.0, 37.6%)	98.2 (97.6, 98.7%)	31.0 (19.5, 44.5%)	97.7 (97.0, 98.3%)
	0.2200	24.3 (14.8, 36.0%)	98.5 (97.9, 99.0%)	34.0 (21.2, 48.8%)	97.6 (96.9, 98.2%)
	0.2243	22.9 (13.7, 34.4%)	98.6 (98.0, 99.0%)	33.3 (20.4, 48.4%)	97.6 (96.9, 98.2%)
	0.2262	21.4 (12.5, 32.9%)	98.6 (98.0, 99.0%)	32.6 (19.5, 48.0%)	97.5 (96.8, 98.1%)
	0.2410	20.0 (11.4, 31.3%)	98.8 (98.2, 99.2%)	34.1 (20.1, 50.6%)	97.5 (96.8, 98.1%)
	0.2524	18.6 (10.3, 29.7%)	99.1 (98.7, 99.5%)	40.6 (23.7, 59.4%)	97.5 (96.7, 98.1%)
	0.2547	17.1 (9.2, 28.0%)	99.2 (98.7, 99.5%)	40.0 (22.7, 59.4%)	97.4 (96.7, 98.0%)
	0.2577	15.7 (8.1, 26.4%)	99.3 (98.8, 99.6%)	40.7 (22.4, 61.2%)	97.4 (96.6, 98.0%)
	0.2644	14.3 (7.1, 24.7%)	99.3 (98.9, 99.6%)	40.0 (21.1, 61.3%)	97.3 (96.6, 98.0%)
	0.2819	12.9 (6.1, 23.0%)	99.5 (99.1, 99.7%)	42.9 (21.8, 66.0%)	97.3 (96.6, 97.9%)
	0.2845	11.4 (5.1, 21.3%)	99.5 (99.1, 99.7%)	40.0 (19.1, 63.9%)	97.3 (96.5, 97.9%)
	0.2970	10.0 (4.1, 19.5%)	99.5 (99.2, 99.8%)	41.2 (18.4, 67.1%)	97.2 (96.5, 97.9%)
	0.2987	8.6 (3.2, 17.7%)	99.5 (99.2, 99.8%)	37.5 (15.2, 64.6%)	97.2 (96.4, 97.8%)
	0.3084	7.1 (2.4, 15.9%)	99.6 (99.3, 99.8%)	38.5 (13.9, 68.4%)	97.1 (96.4, 97.8%)
	0.3237	5.7 (1.6, 14.0%)	99.8 (99.5, 99.9%)	44.4 (13.7, 78.8%)	97.1 (96.3, 97.8%)
	0.3683	4.3 (0.9, 12.0%)	99.9 (99.7, 100%)	60.0 (14.7, 94.7%)	97.1 (96.3, 97.7%)
	0.3807	2.9 (0.3, 9.9%)	99.9 (99.7, 100%)	50.0 (6.8, 93.2%)	97.0 (96.2, 97.7%)
	0.3894	1.4 (0.0, 7.7%)	99.9 (99.7, 100%)	33.3 (0.8, 90.6%)	97.0 (96.2, 97.6%)
	0.5752	0.0 (0.0, 5.1%)	100 (99.7, 100%)	0.0 (0.0, 97.5%)	96.9 (96.1, 97.6%)
SIRS	0	100 (94.9, 100%)	0.0 (0.0, 0.2%)	3.1 (2.4, 3.9%)	-
	1	75.7 (64.0, 85.2%)	34.0 (32.1, 36%)	3.5 (2.6, 4.6%)	97.8 (96.5, 98.7%)
	2	54.3 (41.9, 66.3%)	75.9 (74.0, 77.6%)	6.6 (4.7, 9.0%)	98.1 (97.4, 98.7%)
	3	10.0 (4.1, 19.5%)	95.5 (94.5, 96.3%)	6.5 (2.7, 13.0%)	97.1 (96.3, 97.8%)
	4	1.4 (0.0, 7.7%)	99.8 (99.5, 99.9%)	16.7 (0.4, 64.1%)	97.0 (96.2, 97.6%)
qSOFA					

	Cut Points	Sensitivity^a	Specificity^a	Positive Predictive Value^a	Negative Predictive Value^a
NEWS	0	100 (94.9, 100%)	0.0 (0.0, 0.2%)	3.1 (2.4, 3.9%)	-
	1	18.6 (10.3, 29.7%)	81.5 (79.8, 83.1%)	3.1 (1.6, 5.2%)	96.9 (96.1, 97.7%)
	2	7.1 (2.4, 15.9%)	99.5 (99.1, 99.7%)	29.4 (10.3, 56%)	97.1 (96.4, 97.8%)
	3	1.4 (0.0, 7.7%)	99.9 (99.7, 100%)	33.3 (0.8, 90.6%)	97.0 (96.2, 97.6%)
	0	100 (94.9, 100%)	0.0 (0.0, 0.2%)	3.1 (2.4, 3.9%)	-
	1	87.1 (77.0, 93.9%)	18.2 (16.6, 19.9%)	3.3 (2.5, 4.2%)	97.8 (95.9, 99.0%)
	2	68.6 (56.4, 79.1%)	41.6 (39.5, 43.7%)	3.6 (2.6, 4.7%)	97.7 (96.5, 98.5%)
	3	48.6 (36.4, 60.8%)	64.2 (62.1, 66.2%)	4.1 (2.9, 5.7%)	97.5 (96.6, 98.3%)
	4	31.4 (20.9, 43.6%)	81.9 (80.2, 83.5%)	5.2 (3.3, 7.8%)	97.4 (96.6, 98.1%)
	5	21.4 (12.5, 32.9%)	92.2 (91.0, 93.3%)	8.0 (4.6, 12.9%)	97.4 (96.6, 98.0%)
	6	17.1 (9.2, 28.0%)	97.2 (96.5, 97.9%)	16.4 (8.8, 27.0%)	97.4 (96.6, 98.0%)
	7	8.6 (3.2, 17.7%)	98.7 (98.1, 99.1%)	17.1 (6.6, 33.6%)	97.2 (96.4, 97.8%)
	8	5.7 (1.6, 14.0%)	99.2 (98.8, 99.6%)	19.0 (5.4, 41.9%)	97.1 (96.3, 97.7%)
	9	4.3 (0.9, 12.0%)	99.6 (99.3, 99.8%)	27.3 (6.0, 61%)	97.1 (96.3, 97.7%)
	10	2.9 (0.3, 9.9%)	99.9 (99.6, 100%)	40.0 (5.3, 85.3%)	97.0 (96.2, 97.7%)
	11	-	-	-	-
	12	-	-	-	-
	13	0.0 (0.0, 5.1%)	99.9 (99.7, 100%)	0.0 (0.0, 84.2%)	96.9 (96.1, 97.6%)
	14	0.0 (0.0, 5.1%)	99.9 (99.7, 100%)	0.0 (0.0, 97.5%)	96.9 (96.1, 97.6%)
	15 ^b	-	-	-	-
	16 ^b	-	-	-	-
	17 ^b	-	-	-	-
	18 ^b	-	-	-	-
	19 ^b	-	-	-	-
	20 ^b	-	-	-	-

^a95% Confidence Intervals estimated using Clopper Pearson methods

^bNo PBIs accompanied within these scores in the validation dataset

There were multiple NEWS scores that did not appear in our validation dataset, namely higher ones (15-20). This is likely the fact that we are observed a large amount of missingness in predictors included in this tool and we imputed missingness using a single normal value. For example, we observed >90% missingness within GCS and >60% missingness within oxygen saturation (Spo2).

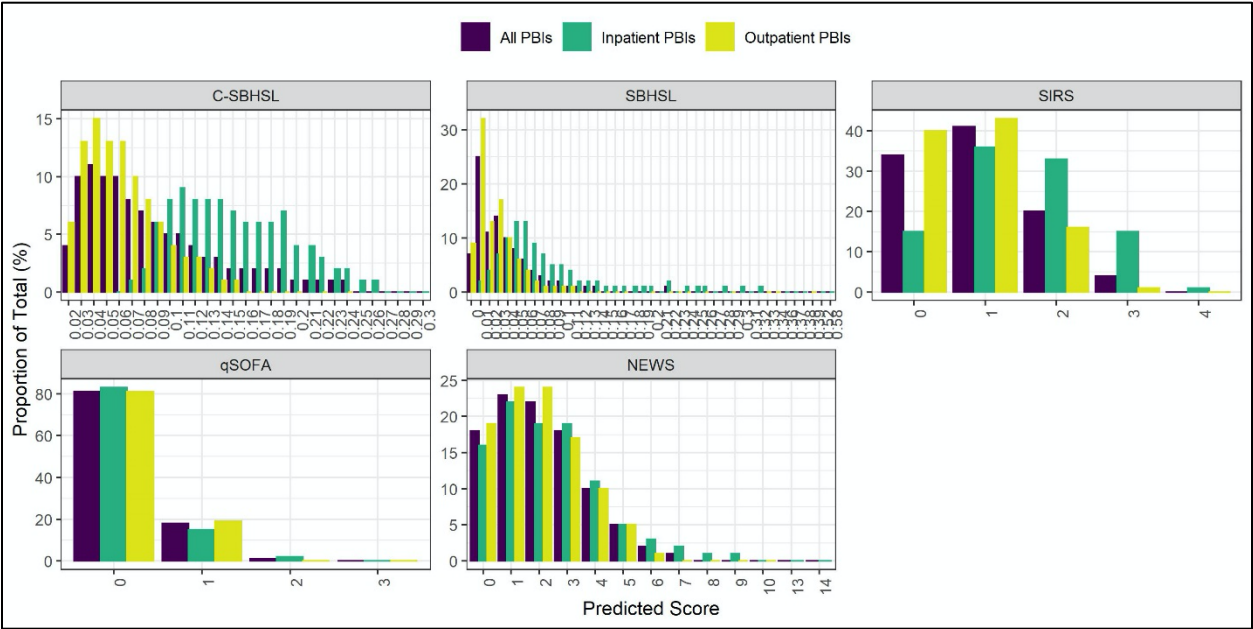
e.Table 5: Categorical Reclassification Index Comparing Risk Classification of Full Predictor Tool (SHBSL) to That of Other Examined Tools

	NRI	95% CI	Pvalue
Clinical Factor Specific - C-SHBSL ^a	0.37	0.20 - 0.54	<0.001
SIRS ^b	0.52	0.34 - 0.69	<0.001
qSOFA ^b	0.31	0.13 - 0.49	<0.001
NEWS ^b	0.35	0.20 - 0.49	<0.001

^aCategories defined as quartiles of C-SBHSI risk prediction

^bCategories defined as risk probabilities categories estimated using single variable logistic regression

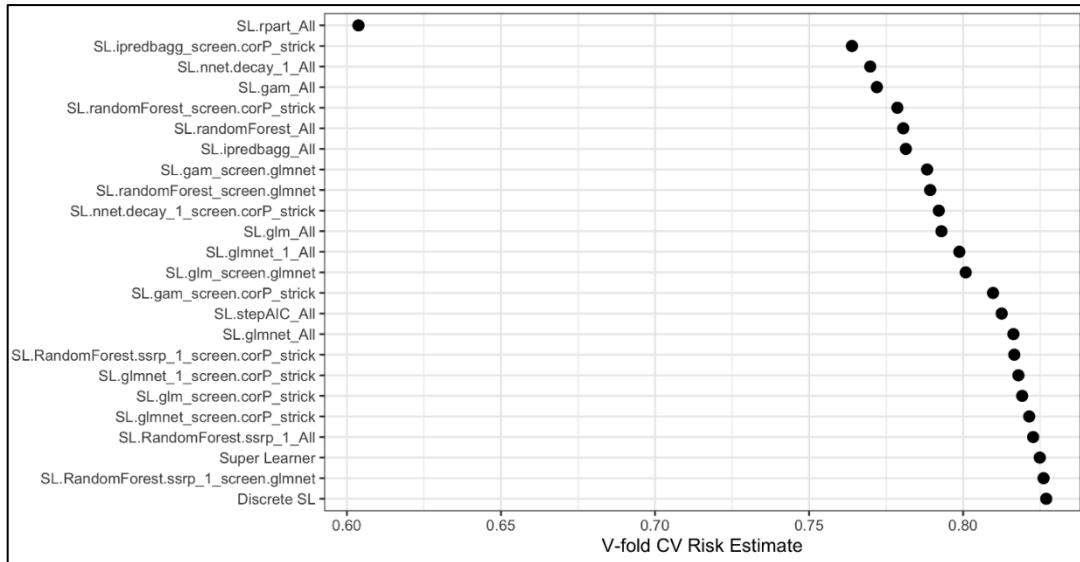
eFigure 5. Summary of Prediction Scores by Patient Location



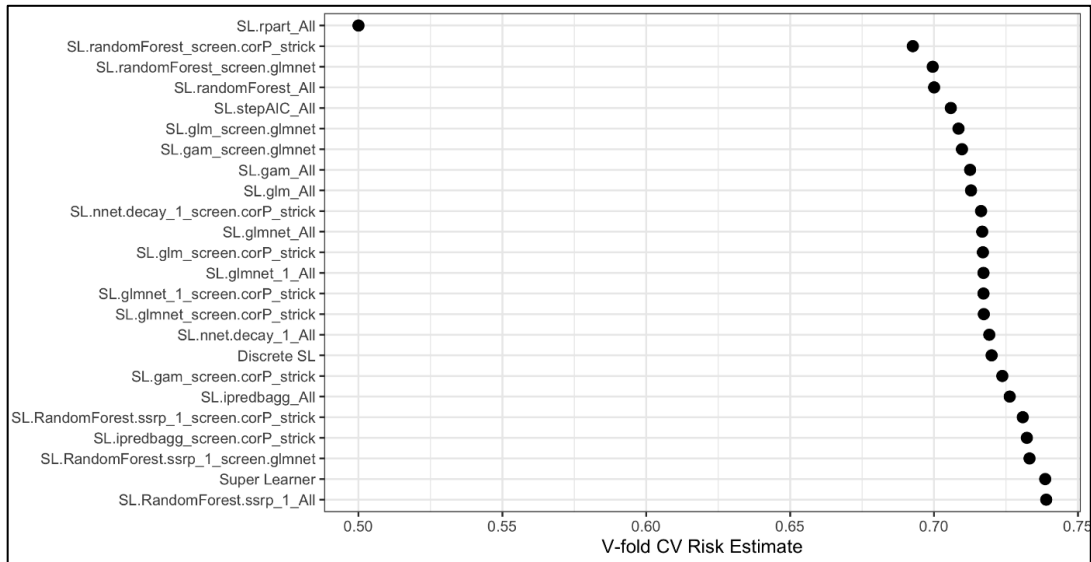
Ensemble Selection Rational

eFigure 6. Cross Validated Area Under the Curve Estimates of Super Learner and User Supplied Algorithms

A. SHBSL – Full Predictor Tool



B. C-SHBSL – Clinical Factor-Specific Tool



The cross validated area under the curve (cvAUC) estimates for each user supplied algorithm and the resulting Super Learner for the full predictor (SHBSL) and clinical factor-specific predictor

(C-SHBSL) prediction tools. The name structure for is algorithm is as follows: SL refers to the fact that the algorithms are running within the super learner framework, algorithm name (detailed list above), screen (screen.corP_strick or screen.glmnet) or absence of screen (All).

Despite the Super Learner ensembles having slightly lower cvAUCs than (screened - SHBSL and base – C-SHBSL) random forest with balanced groups of 100 samples each, we chose to present the ensembles as our final models. We made this decision because the ensembles performed essentially as well as the random forest (<1 AUC percentage point difference) and they provide the least biased option for future updates. Because excessive testing has shown that, in small data settings, the Super Learner performs essentially as well as any provided algorithm⁴ (as we saw in our data), we feel comfortable assuming that if, upon future parameter updating, the best fit provided algorithm is no longer the group-matched random forest, the Super Learner will still perform essentially as well as (if not better than) the best provided algorithm.

e.Table 6: Cross Validated and Bootstrapped Area Under the Curve Estimates in Relation to High Sepsis-Risk Bacteremia^a

	AUC	95% CIs
Clinical Factor-Specific - C-SHBSL ^a	0.74	0.59 - 0.86
Full Predictor - SHBSL ^a	0.85	0.70 - 0.91
SIRS (2+) ^b	0.64	0.52 - 0.77
qSOFA (2+) ^b	0.51	0.44 - 0.60
NEWS ^b	0.60	0.50 - 0.72

^a cvAUC estimates were generated using nested 10-fold cross validation

^b 1000 Bootstrapped Estimates

e.Table 7: Calibration - Observed Versus Predicted High Sepsis-Risk Bacteremia Probabilities^a

Observed bacterial sepsis (%)	3.1%
<i>Sepsis prediction (mean (IQR))</i>	
Clinical Factor Specific - C-SHBSL ^a	0.031 (0.013, 0.033)
Full Predictor-SHBSL ^a	0.031 (0.013, 0.024)
SIRS	0.031 (0.016, 0.048)
qSOFA	0.031 (0.028, 0.028)
NEWS	0.031 (0.022, 0.035)

^a Performed in the validation dataset

Sensitivity Analyses

Our tools were developed using data from a single center and may be biased towards data collection practices of that center. In the absence of external data (another center's) data for validation at the time of manuscript development, we performed numerous sensitivity analyses to better understand the generalizability of our developed tools. Each sensitivity analysis was conducted among patients randomized to the held-out validation dataset. We estimated each tool's AUROC and sensitivity and specificity and the following cut-points: 2+ for SIRS and qSOFA, 7+ for NEWS, and Upper-Left method selected optimal cut-point for our tools (selected in the primary analysis held-out validation data – SHBSL: 5.0%, C-SHBSL: 9.3%).

Missing Data

Because vitals and laboratory measure missingness may differ between clinical staff, our tool may be biased towards the missingness patterns of our center. We performed numerous sensitivity analyses to examine the robustness of our tool's superiority over existing tools under varying missingness scenarios. Along with testing the robustness of our results to missing data patterns, these analyses also provide insight into the predictor collection requirements of each tool (ie. does each tool require complete data to run optimally or can it produce similar estimations under varying amounts of missingness). Specifically, we examined the predictive value of each tool under the following scenarios:

1. Complete Case (restricted to only samples with full data present)
2. Partial Missingness (restricted to samples with at least one missing sample)
3. Average Missingness (restricted to samples with at least three (the average observed number of missing measurements) missing measurements)

4. Single Value from Observed Range Imputation (values within the observed predictor range were selected at random for each observation)

e.Table 8: Predictive Ability of Examined Predictions Tools in Relation to High Sepsis-risk Bacteremia Among Allogeneic HCT Recipients with PBIs Under Varying Missing Data Assumptions

	AUROC ^{b,c}	PR-AUC ^b	Sensitivity ^d	Specificity ^d
<i>Complete Case</i>				
Clinical Factor Specific - C-SHBSL ^a	0.76 (0.67, 0.85)	0.08	75.0 (50.9, 91.3%)	59.5 (56.4, 62.5%)
Full Predictor-SHBSL ^a	0.81 (0.71, 0.90)	0.11	70.0 (45.7, 88.1%)	72.3 (69.5, 75.0%)
SIRS (2+)	0.60 (0.46, 0.75)	0.07	55.0 (31.5, 76.9%)	68.8 (65.8, 71.6%)
qSOFA (2+)	0.59 (0.49, 0.69)	0.09	10.0 (1.2, 31.7%)	99.0 (98.2, 99.5%)
NEWS (4+)	0.57 (0.40, 0.74)	0.06	10.0 (1.2, 31.7%)	98.4 (97.5, 99.1%)
<i>Partial Missing for All Potential Infections</i>				
Clinical Factor Specific - C-SHBSL ^a	0.73 (0.64, 0.81)	0.23	66.0 (51.2, 78.8%)	71.0 (68.4, 73.6%)
Full Predictor-SHBSL ^a	0.86 (0.81, 0.91)	0.25	80.0 (66.3, 90.0%)	74.5 (72.0, 77.0%)
SIRS (2+)	0.67 (0.59, 0.76)	0.08	54.0 (39.3, 68.2%)	81.9 (79.6, 84.1%)
qSOFA (2+)	0.45 (0.39, 0.51)	0.11	6.0 (1.3, 16.5%)	99.8 (99.4, 100%)
NEWS (4+)	0.57 (0.48, 0.65)	0.10	8.0 (2.2, 19.2%)	99.1 (98.4, 99.5%)
<i>Average Missing for All Potential Infections</i>				
Clinical Factor Specific - C-SHBSL ^a	0.64 (0.51, 0.77)	0.17	50.0 (29.9, 70.1%)	68.6 (65.4, 71.6%)
Full Predictor-SHBSL ^a	0.85 (0.78, 0.92)	0.21	76.9 (56.4, 91.0%)	72.9 (69.8, 75.8%)
SIRS (2+)	0.64 (0.52, 0.75)	0.05	42.3 (23.4, 63.1%)	85.4 (82.9, 87.7%)
qSOFA (2+)	0.43 (0.35, 0.51)	0.12	3.8 (0.1, 19.6%)	100 (99.6, 100%)
NEWS (4+)	0.61 (0.50, 0.72)	0.04	3.8 (0.1, 19.6%)	99.3 (98.5, 99.8%)
<i>Single Value from Observed Range Imputation</i>				
Clinical Factor Specific - C-SHBSL ^a	0.67 (0.60, 0.74)	0.09	68.6 (56.4, 79.1%)	55.3 (53.2, 57.4%)
Full Predictor-SHBSL ^a	0.84 (0.79, 0.88)	0.15	80.0 (68.7, 88.6%)	73.7 (71.8, 75.5%)
SIRS (2+)	0.60 (0.54, 0.67)	0.05	65.7 (53.4, 76.7%)	52.6 (50.5, 54.7%)
qSOFA (2+)	0.48 (0.41, 0.54)	0.07	27.1 (17.2, 39.1%)	72.9 (71.0, 74.7%)
NEWS (4+)	0.51 (0.44, 0.57)	0.03	45.7 (33.7, 58.1%)	59.7 (57.6, 61.8%)

^a Based on Upper-Left Selected Cut-points from primary validation analysis (C-SHBSL: 9.3%; SHBSL: 5.0%)

^b AUROC: area under the receiver operating characteristic curve, PR-AUC: area under the precision recall curve

^c 95% Confidence Interval estimated using 2000 stratified bootstrapped replicates

^d 95% Confidence Intervals estimated using Clopper Pearson methods

Race Stratified

Despite the Fred Hutchinson Cancer Research Center / Seattle Cancer Care Alliance

(FHCRC/SCCA) having a large catchment area in the Northwest region of the United States, its

transplant patient population is largely non-Hispanic white and may not reflect

(racially/ethnically) the national transplant population. This is concerning within the context of

sepsis prediction modeling because of evidence suggesting that there are racial/ethnic differences

in both the rate of transplant complications and sepsis.^{5–8} To examine the generalizability of our

tool to a more racially/ethnically diverse population, we estimated the predictive value of our

tool stratified by race/ethnicity. Due to limited data for non-white racial/ethnic groups, we

examined the following stratification:

1. non-Hispanic White
2. Other

e.Table 9: Predictive Ability of Sepsis Prediction Tools Among Varying Racial/Ethnic Groups

	AUROC ^{b,c}	PR-AUC ^b	Sensitivity ^d	Specificity ^d
<i>non-Hispanic White</i>				
	0.72		63.0	66.8
Clinical Factor Specific - C-SHBSL ^a	(0.64, 0.80)	0.17	(47.5, 76.8%)	(64.4, 69.1%)
	0.86		82.6	74.2
Full Predictor-SHBSL ^a	(0.82, 0.91)	0.15	(68.6, 92.2%)	(72.0, 76.4%)
	0.64		52.2	76.2
SIRS (2+)	(0.55, 0.72)	0.07	(36.9, 67.1%)	(74.0, 78.3%)
	0.50		6.5	99.5
qSOFA (2+)	(0.44, 0.56)	0.07	(1.4, 17.9%)	(99.0, 99.8%)
	0.60		10.9	99.2
NEWS (4+)	(0.51, 0.69)	0.08	(3.6, 23.6%)	(98.6, 99.6%)
<i>Other</i>				
	0.72		66.7	68.0
Clinical Factor Specific - C-SHBSL ^a	(0.61, 0.83)	0.10	(44.7, 84.4%)	(64.2, 71.6%)
	0.81		70.8	72.8
Full Predictor-SHBSL ^a	(0.71, 0.91)	0.36	(48.9, 87.4%)	(69.1, 76.3%)
	0.63		58.3	75.0
SIRS (2+)	(0.51, 0.76)	0.06	(36.6, 77.9%)	(71.5, 78.4%)
	0.51		8.3	99.4
qSOFA (2+)	(0.42, 0.60)	0.07	(1.0, 27.0%)	(98.4, 99.8%)
	0.56		8.3	97.8
NEWS (4+)	(0.44, 0.69)	0.07	(1.0, 27.0%)	(96.3, 98.8%)

^a Based on Upper-Left Selected Cut-points from primary validation analysis (C-SHBSL: 9.3%; SHBSL: 5.0%)

^b AUROC: area under the receiver operating characteristic curve, PR-AUC: area under the precision recall curve

^c 95% Confidence Interval estimated using 2000 stratified bootstrapped replicates

^d 95% Confidence Intervals estimated using Clopper Pearson methods

Window of Measurement Collection

We developed tools to provide decision support at the time of culture collection and based on our predictor measurement inclusion defined the run time of the tools as 2 hours following culture collection. This time period was selected because predictor measurements were commonly collected within the first few hours following culture collection and we considered these measures to be associated with the collection of the culture. However, in practice, it may be desirable to run these tools closer in time to the culture. For this reason, we estimated the AUC, sensitivity, and specificities of each examined tools using data from the 24 hours and:

1. 1 hours prior
2. 0.5 hours prior
3. 0 hours prior

e.Table 10: Predictive Ability of Sepsis Prediction Tools Under Varying Predictor Measurement Collection Time Windows

	AUROC ^{b,c}	PR-AUC ^b	Sensitivity ^d	Specificity ^d
<i>0 Hours</i>				
Clinical Factor Specific				
- C-SHBSL ^a	0.7 (0.63, 0.76)	0.11	61.4 (49.0, 72.8%)	67.6 (65.7, 69.6%)
Full Predictor-SHBSL ^a	0.85 (0.81, 0.9)	0.20	81.4 (70.3, 89.7%)	72.4 (70.5, 74.3%)
SIRS (2+)	0.63 (0.56, 0.7)	0.07	41.4 (29.8, 53.8%)	81.5 (79.8, 83.1%)
qSOFA (2+)	0.56 (0.5, 0.61)	0.10	5.7 (1.6, 14.0%)	99.7 (99.4, 99.9%)
NEWS (4+)	0.63 (0.56, 0.7)	0.07	5.7 (1.6, 14.0%)	98.9 (98.4, 99.3%)
<i>0.5 Hours</i>				
Clinical Factor Specific				
- C-SHBSL ^a	0.78 (0.75, 0.82)	0.18	73.9 (67.9, 79.4%)	67.0 (65.9, 68.0%)
Full Predictor-SHBSL ^a	0.93 (0.92, 0.95)	0.39	93.3 (89.3, 96.1%)	75.3 (74.4, 76.3%)
SIRS (2+)	0.61 (0.58, 0.65)	0.05	42.4 (36.1, 49%)	78.6 (77.7, 79.5%)
qSOFA (2+)	0.50 (0.48, 0.53)	0.05	2.5 (0.9, 5.4%)	99.5 (99.4, 99.7%)
NEWS (4+)	0.60 (0.57, 0.64)	0.05	3.8 (1.7, 7.1%)	98.9 (98.6, 99.1%)
<i>1 Hour</i>				
Clinical Factor Specific				
- C-SHBSL ^a	0.74 (0.68, 0.80)	0.13	65.7 (53.4, 76.7%)	67.0 (65.0, 68.9%)
Full Predictor-SHBSL ^a	0.85 (0.81, 0.90)	0.20	78.6 (67.1, 87.5%)	73.4 (71.5, 75.3%)
SIRS (2+)	0.62 (0.55, 0.69)	0.06	45.7 (33.7, 58.1%)	77.3 (75.5, 79.1%)
qSOFA (2+)	0.54 (0.49, 0.60)	0.09	7.1 (2.4, 15.9%)	99.5 (99.1, 99.7%)
NEWS (4+)	0.61 (0.54, 0.69)	0.07	7.1 (2.4, 15.9%)	98.6 (98.1, 99.1%)

^a Based on Upper-Left Selected Cut-points from primary validation analysis (C-SHBSL: 9.3%; SHBSL: 5.0%)

^b AUROC: area under the receiver operating characteristic curve, PR-AUC: area under the precision recall curve

^c 95% Confidence Interval estimated using 2000 stratified bootstrapped replicates

^d 95% Confidence Intervals estimated using Clopper Pearson methods

Culture Collection Practices

Our tools are designed to supply decision support at the time of blood culture collection and are likely biased towards the culture collection practices at our center. While we cannot retrospectively change the culture collection practices of our center to examine the generalizability of our tools, we were able to adjust the cultures we included and include a proportion of non-culture time to simulate a “what if cultures had been collected” scenario.

1. Excluding Surveillance Cultures - Surveillance status was not perfectly recorded, and this sensitivity analysis was performed among cultures with recorded reasons other than “surveillance.” While this reduces the number of surveillance cultures included in our analysis, it may not fully remove them.
2. Percent non-Culture 24 Hours: To the observed PBIs in the validation data (2286), we added additional periods of 24-hour follow-up. The 24-hour (one full calendar day) data collection periods were randomly selected among calendar days between transplant and end of post-transplant follow-up. 24-hour periods within 3 days of a culture collection and during follow-up cultures periods were excluded.
 - a. Added an additional 114 (or 5%) non-Culture 24 hours
 - b. Added an additional 229 (or 10%) non-Culture 24 hours

eTable 11: Predictive Ability of Sepsis Prediction Tools Under Varying Culture Collection and Potential Infection Restriction Definitions

	AUROC ^{b,c}	PR-AUC ^b	Sensitivity ^d	Specificity ^d
<i>Cultures Excluding Surveillance^b</i>				
Clinical Factor Specific - C-SHBSL ^a	0.71 (0.64, 0.78)	0.14	66.2 (53.4, 77.4%)	61.7 (59.2, 64.1%)
Full Predictor-SHBSL ^a	0.83 (0.79, 0.88)	0.21	78.5 (66.5, 87.7%)	74.0 (71.7, 76.1%)
SIRS (2+)	0.64 (0.57, 0.70)	0.08	56.9 (44.0, 69.2%)	71.0 (68.7, 73.3%)
qSOFA (2+)	0.56 (0.51, 0.62)	0.10	9.2 (3.5, 19.0%)	99.1 (98.5, 99.5%)
NEWS (4+)	0.62 (0.55, 0.70)	0.09	10.8 (4.4, 20.9%)	98.1 (97.2, 98.7%)
<i>5% non-Culture 24 hours</i>				
Clinical Factor Specific - C-SHBSL ^a	0.73 (0.67, 0.8)	0.14	68.6 (56.4, 79.1%)	66.9 (64.9, 68.8%)
Full Predictor-SHBSL ^a	0.85 (0.81, 0.9)	0.21	81.4 (70.3, 89.7%)	73.4 (71.5, 75.1%)
SIRS (2+)	0.63 (0.57, 0.7)	0.06	48.6 (36.4, 60.8%)	76.2 (74.4, 77.9%)
qSOFA (2+)	0.55 (0.5, 0.61)	0.09	7.1 (2.4, 15.9%)	99.4 (99.0, 99.7%)
NEWS (4+)	0.63 (0.56, 0.7)	0.08	10.0 (4.1, 19.5%)	98.9 (98.4, 99.3%)
<i>10% non-Culture 24 hours</i>				
Clinical Factor Specific - C-SHBSL ^a	0.86 (0.83, 0.89)	0.38	83.2 (77.8, 87.7%)	66.6 (65.5, 67.6%)
Full Predictor-SHBSL ^a	0.94 (0.93, 0.96)	0.45	94.1 (90.3, 96.7%)	75.8 (74.8, 76.7%)
SIRS (2+)	0.63 (0.60, 0.67)	0.06	49.6 (43.1, 56.1%)	76.6 (75.7, 77.5%)
qSOFA (2+)	0.52 (0.50, 0.55)	0.05	3.4 (1.5, 6.5%)	99.5 (99.3, 99.6%)
NEWS (4+)	0.6 (0.57, 0.64)	0.05	5.9 (3.3, 9.7%)	98.7 (98.5, 99.0%)

^a Based on Upper-Left Selected Cut-points from primary validation analysis (C-SHBSL: 9.3%; SHBSL: 5.0%)

^b AUROC: area under the receiver operating characteristic curve, PR-AUC: area under the precision recall curve

^c 95% Confidence Interval estimated using 2000 stratified bootstrapped replicates

^d 95% Confidence Intervals estimated using Clopper Pearson methods

Antibiotic Influence on Culture Results

We carefully selected our definition of bacterial sepsis for our tool development but understand that culture confirmation of a blood borne infection comes with limitations. A large limitation being that recent antibiotic use may impact culture results. We addressed this in tool development and the primary evaluation by excluding cultures collected within 3 days of antibiotic use but excluded our center's primary prophylaxis, levofloxacin. To further account for the potential impact of recent antibiotic use, we performed the following sensitivity analysis:

1. All Cultures Regardless of Recent Antibiotic Use
2. No Antibiotic Use in Last 3 Days (Including Levofloxacin)
3. No Use in Last 7 Days (Excluding Levofloxacin)

e.Table 12: Predictive Ability of Sepsis Prediction Tools Under Varying Recent Antibiotic Definitions

	AUROC ^{b,c}	PRAUC ^b	Sensitivity ^d	Specificity ^d
<i>All Cultures Regardless of Recent Antibiotic Use</i>				
Clinical Factor Specific - C-SHBSL ^a	0.64 (0.59, 0.68)	0.06	72.3 (64.2, 79.5%)	52.8 (51.2, 54.4%)
Full Predictor-SHBSL ^a	0.81 (0.78, 0.84)	0.15	75.9 (68.0, 82.7%)	69.6 (68.1, 71.1%)
SIRS (2+)	0.56 (0.52, 0.60)	0.04	50.4 (41.8, 58.9%)	61.3 (59.7, 62.8%)
qSOFA (2+)	0.51 (0.47, 0.55)	0.05	7.1 (3.5, 12.7%)	97.2 (96.6, 97.7%)
NEWS (4+)	0.55 (0.50, 0.60)	0.04	9.9 (5.5, 16.1%)	93.7 (92.8, 94.4%)
<i>No Use^b in Last 3 Days (Including Levofloxacin)</i>				
Clinical Factor Specific - C-SHBSL ^a	0.67 (0.59, 0.75)	0.10	56.6 (42.3, 70.2%)	68.4 (66.4, 70.4%)
Full Predictor-SHBSL ^a	0.83 (0.78, 0.88)	0.13	69.8 (55.7, 81.7%)	76.2 (74.3, 78.1%)
SIRS (2+)	0.60 (0.52, 0.69)	0.07	41.5 (28.1, 55.9%)	78.2 (76.4, 80.0%)
qSOFA (2+)	0.57 (0.50, 0.63)	0.09	7.5 (2.1, 18.2%)	99.6 (99.2, 99.8%)
NEWS (4+)	0.59 (0.50, 0.68)	0.06	7.5 (2.1, 18.2%)	98.6 (98.0, 99.1%)
<i>No Use^b in Last 7 Days (Excluding Levofloxacin)</i>				
Clinical Factor Specific - C-SHBSL ^a	0.77 (0.70, 0.83)	0.14	70.0 (56.8, 81.2%)	68.1 (66.0, 70.2%)
Full Predictor-SHBSL ^a	0.86 (0.82, 0.91)	0.24	85.0 (73.4, 92.9%)	71.5 (69.5, 73.5%)
SIRS (2+)	0.66 (0.59, 0.73)	0.07	55.0 (41.6, 67.9%)	76.7 (74.7, 78.5%)
qSOFA (2+)	0.54 (0.49, 0.60)	0.08	5.0 (1.0, 13.9%)	99.6 (99.2, 99.8%)
NEWS (4+)	0.62 (0.54, 0.70)	0.07	6.7 (1.8, 16.2%)	99.2 (98.7, 99.6%)

^a Based on Upper-Left Selected Cut-points from primary validation analysis (C-SHBSL: 9.3%; SHBSL: 5.0%)

^b AUROC: area under the receiver operating characteristic curve, PR-AUC: area under the precision recall curve

^c 95% Confidence Interval estimated using 2000 stratified bootstrapped replicates

^d 95% Confidence Intervals estimated using Clopper Pearson methods

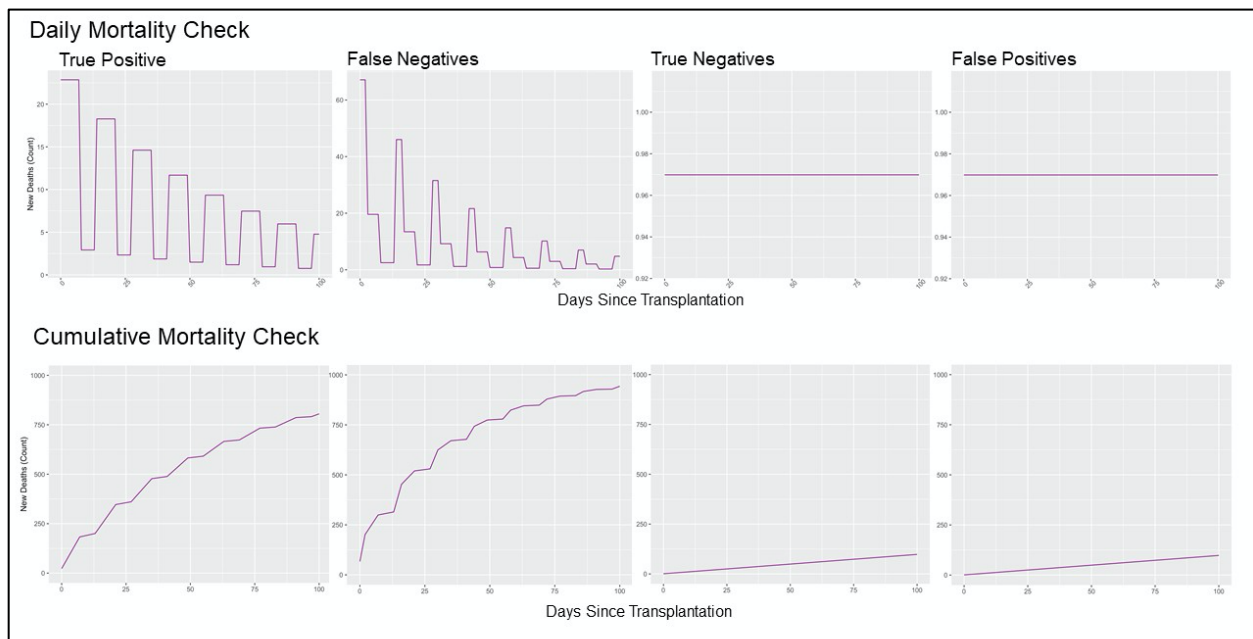
References

1. Hämäläinen S, Kuittinen T, Matinlauri I, Nousiainen T, Koivula I, Jantunen E. Severe sepsis in autologous stem cell transplant recipients: Microbiological aetiology, risk factors and outcome. *Scandinavian Journal of Infectious Diseases*. 2009;41(1):14-20. doi:10.1080/00365540802454706
2. Póvoa P, Almeida E, Moreira P, et al. C-reactive protein as an indicator of sepsis. *Intensive Care Med*. 1998;24(10):1052-1056.
3. Abe R, Oda S, Sadahiro T, et al. Gram-negative bacteremia induces greater magnitude of inflammatory response than Gram-positive bacteremia. *Crit Care*. 2010;14(2):R27. doi:10.1186/cc8898
4. Laan MJ van der, Rose S. *Targeted Learning: Causal Inference for Observational and Experimental Data*. 2011 edition. Springer; 2011.
5. Majhail NS, Nayyar S, Santibañez MEB, Murphy EA, Denzen EM. Racial disparities in hematopoietic cell transplantation in the United States. *Bone Marrow Transplant*. 2012;47(11). doi:10.1038/bmt.2011.214
6. Jones JM, Fingar KR, Miller MA, et al. Racial Disparities in Sepsis-Related In-Hospital Mortality: Using a Broad Case Capture Method and Multivariate Controls for Clinical and Hospital Variables, 2004-2013. *Crit Care Med*. 2017;45(12):e1209-e1217. doi:10.1097/CCM.0000000000002699
7. Moore JX, Donnelly JP, Griffin R, et al. Black-white racial disparities in sepsis: a prospective analysis of the REasons for Geographic And Racial Differences in Stroke (REGARDS) cohort. *Crit Care*. 2015;19(1). doi:10.1186/s13054-015-0992-8
8. DiMeglio M, Dubensky J, Schadt S, Potdar R, Laudanski K. Factors Underlying Racial Disparities in Sepsis Management. *Healthcare (Basel)*. 2018;6(4). doi:10.3390/healthcare6040133

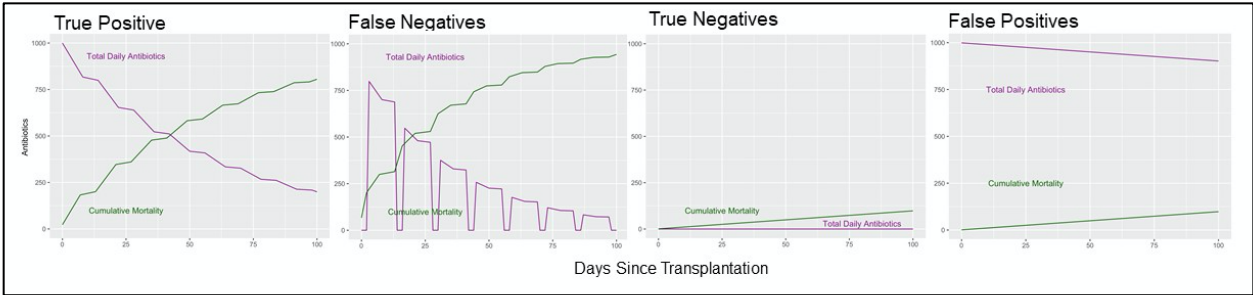
Supplemental Paper 3: Worth the switch? The impact of a novel bacterial sepsis prediction tool on mortality among allogeneic HCT recipients with potential infections

Model Validation

e.Figure 1: Stochastic equation-based Model Daily and Cumulative Mortality Checks



e.Figure 2: Stochastic equation-based Model Daily Antibiotic Check



Modeling Assumptions

1. All patients are at risk for an infection at time zero (transplant)
2. All infected patients are cultured
3. All culture results are accurate
4. Tool classification (test negative or positive) equals targeted antibiotic therapy
5. Antibiotics were appropriate (no resistance)
6. Antibiotics were terminated when negative culture results were returned
7. Infection duration, time to culture results, mortality by infection and treatment status is fixed (see Methods Parameters)
 - a. Aggregated parameter estimates are appropriate – ignoring complexities such as GVHD and neutropenia
8. Mortality risk difference is limited to high-risk bacteremia – on average there is no mortality-risk related difference between infected patients whose prognostic results are positive or negative or non-infected patients whose prognostic results are positive or negative
9. Infection risk is independent of previous antibiotic treatments or infections
10. Mortality risk is dependent upon current infection and treatment status and ignores lingering risks associated with previous infections (treated or untreated)
11. Non-antibiotic related treatments (ex. fluids) were not considered

Observed Mortality Proportions

Methods

Tool Retrospective Implementation

Each tool was retrospectively evaluated using data from the 24 hours prior and 2 hours following culture collection. If more than one measurement per predictor was identified, the measurement taken closest in time to that of the culture collection was used. Missing measurements were accounted for using a single normal value imputation approach, an imputation method common within sepsis prediction research. For this analysis, we limited the examined PBIs to those not used in tool development.(Paper 2)

PBIs were considered positive for a tool if the score (NEWS, qSOFA, SIRS, HBS2) or probability (SHBSL, C-SHBSL, HBST) met the following thresholds: SIRS and qSOFA 2+, NEWS 4+, HBS2 2+, HBST 3.3%+, SHBSL 5.5%+, and C-SHBSL 11.1%+. All PBIs with scores or probabilities below a tool's threshold were considered negative. Correctness was defined in relation to high-risk bacteremia. This analysis was done ignoring antibiotic initiation at time of culture collection or directly following culture collection. However, to examine the impact of antibiotic initiation in relation tool classification, we further stratified the above analysis by presence or absence of antibiotic initiation (initiation at or within the 3 days following collection).

Outcome Definition

Our primary outcome was proportion of culture that occurred within 10 days of death.

Statistical Analysis

We estimated the 10-day mortality proportion of each tool's predictions using a Poisson regression with an independent model covariate and culture count offset. Using pair-wise proportion ratios, we compared the sepsis related mortality proportions between each tool's positive and negative prediction groups.

Results

Among the examined tools, the largest proportion of true positive PBIs identified by qSOFA occurred within 10 days of death (40.0%), followed by NEWS (33.3%). Of the high-risk bacteremia PBIs incorrectly classified as negative (false negatives) by SHBSL and HBS2, zero occurred within 10 days of death. A greater proportion of the true positive events identified by SBHSL, HBS2, NEWS, and qSOFA occurred within 10 days of death than of the false negatives identified by these tools (eTable 1).

eTable 2 displays the pairwise 10-day mortality comparison between tool classified positive to negative PBIs. Patients with PBIs classified as positive (being high-risk for bacteremia) by qSOFA (the tool with the highest pairwise ratio) were 38.9 (95% CI: 15.5-85.8) times more likely to die within 10 days of culture collection than patients classified as negative. SIRS had the lowest pairwise ratio and patients classified as positive were 2.46 (95% CI: 1.19-4.99) times more likely to die than patients classified as negative (eTable 2).

Of the examined PBIs ($n = 2286$), 500 (21.8%) were accompanied with an antibiotic at the time of or within the 3 days following culture collection. Of the 500 PBIs with accompanying antibiotic initiation, 63 (12.6%) were defined as high-risk bacteremia and of the 1786 PBIs without accompanying antibiotic initiation, 7 (0.4%) were defined as high-risk bacteremia. Following stratification by accompanying antibiotic initiation, SIRS maintained having the highest proportion of 10-day mortality events among true positive and false positive classified PBIs (eTable 3, 5). The pairwise proportions for each tool were higher among PBIs without accompanying antibiotic initiation than with and under both scenarios, qSOFA had the highest pairwise ratio (with antibiotics: 22.59; without antibiotics: 47.49; eTable 4, 6).

Discussion

Through retrospective evaluation, we examined the association between 10-day mortality and

test classification. From this analysis, we identified two key take-a-ways. First, we found that, qSOFA classification had the strongest association with 10-day mortality and among developed tools, SHBSL had the strongest association. While the relative strength of qSOFA's association held following stratification by presence of accompanying antibiotic initiation events, among PBIs without accompanying antibiotic initiations, neither SIRs nor NEWS (the other tool with a strong association) properly classified any PBIs as positive (true positive count of zero). This suggests that, while PBIs classified as positive by qSOFA or NEWS were more likely to occur within 10-days of death than PBIs classified as positive by the other tools, this association, is at least partially, driven by deaths due to non-high-risk bacteremia associated issues and thus may not benefit from antibiotic initiation. Further, we observed a strong association between qSOFA and NEWS classification and mortality among PBIs with accompanying antibiotic initiation (Mortality proportion among true positives qSOFA: 40% and NEWS 33.3%). These findings indicate that being classified as positive by qSOFA or NEWS is highly associated with mortality regardless of antibiotic initiation, calling into question the utility of these tools' positive classification.

Second, we found that for PBIs without accompanying antibiotic initiation, the mortality proportion among true positives and false negatives for each tool was either zero or, for NEWS and SIRS, uncalculatable due to a lack of PBIs of that classification. This likely reflects the limited number of high-risk bacteremia events among this population and the fact that, regardless of tool classification, antibiotic initiation within our observed population aligned strongly with the presence of high-risk bacteremia. However, these results have limited biological plausibility. Given the well-established associations between our selected high-risk bacteremia events, mortality, and antibiotic use, the chance that mortality among patients with high-risk bacteremia

(both true positives and false negatives) in the absence of antibiotic initiation would be zero seems extremely small (eTable 6).

Our results point to the strong association between mortality events and SIRS, NEWS, and SHBSL. While these analysis shine light onto the potential connection between tool classifications and mortality risk, they suffer from an inability to accurately examine the true link between tool classification, treatment initiation, and mortality. In the absence of an adequate sample of patients with tool classification linked antibiotic initiation, we feel that the usefulness of analyses, such as these, is narrow.

eTable 1: 10-day Mortality Probabilities by High-Risk Bacteremia Events Prediction Categories

	True Positives	False Negatives	True Negatives	False Positives
SHBSL 5.5%+	7.8 (2.2, 18.9%)	0 (0, 17.6%)	0.2 (0.1, 0.6%)	4.5 (2.9, 6.7%)
C-SHBSL 11.1%+	2.6 (0.1, 13.8%)	9.4 (2, 25%)	0.5 (0.2, 1%)	3.4 (2, 5.3%)
HBS2 2+	5.0 (0.6, 16.9%)	6.7 (0.8, 22.1%)	0.8 (0.4, 1.3%)	2.3 (1.3, 3.7%)
HBST 3.3%	5.6 (0.7, 18.7%)	5.9 (0.7, 19.7%)	0.9 (0.5, 1.5%)	2.3 (1.2, 4.1%)
SIRS 2+	5.3 (0.6, 17.7%)	6.2 (0.8, 20.8%)	0.9 (0.5, 1.5%)	2.2 (1.2, 3.9%)
qSOFA 2+	40 (5.3, 85.3%)	3.1 (0.4, 10.7%)	1 (0.6, 1.5%)	41.7 (15.2, 72.3%)
NEWS 4+	33.3 (4.3, 77.7%)	3.1 (0.4, 10.8%)	1.1 (0.7, 1.6%)	10.3 (2.2, 27.4%)

eTable 2: Pairwise Comparison of 10-day Mortality by Tool Identified Positive to Negative Classifications

	Proportion Ratios	95% Confidence Intervals
SHBSL 5.5%+	20.85	(8.16, 70.55)
C-SHBSL 11.1%+	4.77	(2.34, 10.10)
HBS2 2+	2.73	(1.35, 5.63)
HBST 3.3%	2.51	(1.2, 5.09)
SIRS 2+	2.46	(1.19, 4.99)
qSOFA 2+	38.93	(15.49, 85.75)
NEWS 4+	12.37	(4.18, 29.6)

eTable 3: 10-day Mortality Probabilities by High-Risk Bacteremia Events Prediction Categories Among PBIs with Accompanying Antibiotic Initiation

	True Positives	False Negatives	True Negatives	False Positives
SHBSL 5.5%+	8.3 (2.3, 20%)	0 (0, 21.8%)	0.5 (0, 2.9%)	3.6 (1.7, 6.8%)
C-SHBSL 11.1%+	2.7 (0.1, 14.2%)	11.5 (2.4, 30.2%)	1.8 (0.4, 5.1%)	2.6 (1.1, 5.3%)
HBS2 2+	5.3 (0.6, 17.7%)	8.0 (1.0, 26%)	2.4 (0.6, 5.9%)	2.2 (0.8, 4.8%)
HBST 3.3%	5.7 (0.7, 19.2%)	7.1 (0.9, 23.5%)	1.7 (0.5, 4.3%)	3 (1.1, 6.4%)
SIRS 2+	5.9 (0.7, 19.7%)	6.9 (0.8, 22.8%)	1.8 (0.5, 4.6%)	2.8 (1, 5.9%)
qSOFA 2+	40 (5.3, 85.3%)	3.4 (0.4, 11.9%)	1.6 (0.7, 3.3%)	42.9 (9.9, 81.6%)
NEWS 4+	33.3 (4.3, 77.7%)	3.5 (0.4, 12.1%)	1.9 (0.8, 3.7%)	10 (1.2, 31.7%)

eTable 4: Pairwise Comparison of 10-day Mortality by Tool Identified Positive to Negative Classifications Among PBIs with Accompanying Antibiotic Initiation

	Proportion Ratios	95% Confidence Intervals
SHBSL 5.5%+	9.03	(1.80, 164.12)
C-SHBSL 11.1%+	0.85	(0.30, 2.59)
HBS2 2+	0.85	(0.29, 2.57)
HBST 3.3%	1.49	(0.52, 4.53)
SIRS 2+	1.32	(0.46, 4.02)
qSOFA 2+	22.59	(6.94, 65.40)
NEWS 4+	7.29	(2.00, 21.80)

eTable 5: 10-day Mortality Probabilities by High-Risk Bacteremia Events Prediction Categories Among PBIs without Accompanying Antibiotic Initiation

	True Positives	False Negatives	True Negatives	False Positives
SHBSL 5.5%+	0 (0, 70.8%)	0 (0, 60.2%)	0.2 (0, 0.6%)	5.4 (3, 8.8%)
C-SHBSL 11.1%+	0 (0, 97.5%)	0 (0, 45.9%)	0.4 (0.1, 0.9%)	4.2 (2.1, 7.3%)
HBS2 2+	0 (0, 84.2%)	0 (0, 52.2%)	0.6 (0.2, 1.1%)	2.3 (1, 4.3%)
HBST 3.3%	0 (0, 97.5%)	0 (0, 45.9%)	0.8 (0.4, 1.4%)	1.8 (0.6, 4.2%)
SIRS 2+	0 (0, 60.2%)	0 (0, 70.8%)	0.8 (0.4, 1.3%)	1.9 (0.7, 4.1%)
qSOFA 2+	NA ^a	0 (0, 41%)	0.8 (0.5, 1.4%)	40 (5.3, 85.3%)
NEWS 4+	NA ^a	0 (0, 41%)	0.9 (0.5, 1.5%)	11.1 (0.3, 48.2%)

^a Unable to calculate due to the absence of PBIs classified as such by the given tool

eTable 6: Pairwise Comparison of 10-day Mortality by Tool Identified Positive to Negative Classifications Among PBIs without Accompanying Antibiotic Initiation

	Proportion Ratios	95% Confidence Intervals
SHBSL 5.5%+	26.9	8.78, 116.71
C-SHBSL 11.1%+	10.52	4, 30.55
HBS2 2+	3.94	(1.5, 10.48)
HBST 3.3%	2.29	0.73, 6.17
SIRS 2+	2.48	0.85, 6.52
qSOFA 2+	47.49	7.48, 168.3
NEWS 4+	12.34	0.68, 60.43

Sensitivity Analyses

Varying Time till Culture Result and Infection Duration

eTable 7: Simulated Mortality For 2000 Simulated Allogeneic HCT Recipients (Culture Result and Duration Lengths)

Prognostic Tool	Time to Result, Infection Duration ^a	Mortality Count	Prediction Intervals ^b	Mortality Difference
SHBSL	Results 3; Duration 5	213	(211, 215)	0 (0, 0)
SHBSL	Results 3; Duration 7	219	(217, 220)	0 (0, 0)
SHBSL	Results 2; Duration 5	212	(211, 213)	0 (0, 0)
SHBSL	Results 2; Duration 7	218	(216, 218)	0 (0, 0)
SHBSL	Results 2; Duration 14	221	(220, 222)	0 (0, 0)
SHBSL	Results 5; Duration 30	227	(224, 229)	0 (0, 0)
SHBSL	Results 5; Duration 50	228	(226, 231)	0 (0, 0)
SHBSL	Results 10; Duration 30	232	(227, 236)	0 (0, 0)
SHBSL	Results 10; Duration 50	233	(228, 237)	0 (0, 0)
C-SHBSL	Results 3; Duration 5	216	(214, 218)	3 (1, 6)
C-SHBSL	Results 3; Duration 7	222	(220, 223)	3 (1, 5)
C-SHBSL	Results 2; Duration 5	214	(212, 215)	2 (0, 4)
C-SHBSL	Results 2; Duration 7	219	(218, 220)	2 (0, 4)
C-SHBSL	Results 2; Duration 14	223	(221, 224)	2 (0, 3)
C-SHBSL	Results 5; Duration 30	231	(228, 234)	4 (1, 9)
C-SHBSL	Results 5; Duration 50	233	(230, 235)	4 (1, 9)
C-SHBSL	Results 10; Duration 30	239	(234, 244)	7 (2, 16)
C-SHBSL	Results 10; Duration 50	241	(235, 245)	7 (2, 15)
HBS2	Results 3; Duration 5	215	(213, 216)	1 (-1, 4)
HBS2	Results 3; Duration 7	220	(219, 222)	1 (-1, 4)
HBS2	Results 2; Duration 5	213	(212, 214)	1 (-1, 3)
HBS2	Results 2; Duration 7	218	(217, 219)	1 (-1, 3)
HBS2	Results 2; Duration 14	222	(221, 223)	1 (-1, 2)
HBS2	Results 5; Duration 30	229	(227, 232)	2 (-1, 6)
HBS2	Results 5; Duration 50	230	(228, 233)	2 (-1, 6)
HBS2	Results 10; Duration 30	235	(232, 240)	3 (-2, 11)
HBS2	Results 10; Duration 50	237	(233, 241)	3 (-2, 11)
HBST	Results 3; Duration 5	216	(215, 218)	3 (1, 5)
HBST	Results 3; Duration 7	222	(220, 223)	3 (1, 5)
HBST	Results 2; Duration 5	214	(213, 215)	2 (1, 4)
HBST	Results 2; Duration 7	219	(218, 220)	2 (0, 3)
HBST	Results 2; Duration 14	223	(222, 224)	2 (0, 3)

Prognostic Tool	Time to Result, Infection Duration^a	Mortality Count	Prediction Intervals^b	Mortality Difference
HBST	Results 5; Duration 30	232	(229, 234)	5 (1, 9)
HBST	Results 5; Duration 50	233	(231, 235)	5 (1, 8)
HBST	Results 10; Duration 30	240	(236, 244)	8 (2, 15)
HBST	Results 10; Duration 50	241	(237, 246)	8 (2, 15)
SIRS	Results 3; Duration 5	216	(215, 218)	3 (1, 5)
SIRS	Results 3; Duration 7	222	(221, 224)	3 (1, 5)
SIRS	Results 2; Duration 5	214	(213, 215)	2 (1, 4)
SIRS	Results 2; Duration 7	219	(219, 221)	2 (1, 3)
SIRS	Results 2; Duration 14	223	(222, 224)	2 (1, 3)
SIRS	Results 5; Duration 30	232	(229, 235)	5 (2, 8)
SIRS	Results 5; Duration 50	233	(231, 236)	5 (2, 8)
SIRS	Results 10; Duration 30	241	(236, 246)	9 (3, 15)
SIRS	Results 10; Duration 50	242	(238, 247)	8 (3, 14)
qSOFA	Results 3; Duration 5	225	(224, 226)	11 (10, 14)
qSOFA	Results 3; Duration 7	230	(229, 231)	11 (9, 13)
qSOFA	Results 2; Duration 5	219	(219, 220)	8 (6, 9)
qSOFA	Results 2; Duration 7	225	(224, 225)	7 (6, 9)
qSOFA	Results 2; Duration 14	228	(228, 229)	7 (6, 8)
qSOFA	Results 5; Duration 30	245	(243, 246)	18 (15, 21)
qSOFA	Results 5; Duration 50	246	(245, 247)	18 (15, 21)
qSOFA	Results 10; Duration 30	263	(260, 265)	31 (26, 37)
qSOFA	Results 10; Duration 50	264	(261, 266)	30 (25, 36)
NEWS	Results 3; Duration 5	221	(219, 222)	7 (5, 10)
NEWS	Results 3; Duration 7	226	(224, 227)	7 (5, 9)
NEWS	Results 2; Duration 5	217	(215, 218)	5 (3, 6)
NEWS	Results 2; Duration 7	222	(221, 223)	5 (3, 6)
NEWS	Results 2; Duration 14	225	(225, 226)	4 (3, 6)
NEWS	Results 5; Duration 30	238	(236, 241)	11 (8, 15)
NEWS	Results 5; Duration 50	239	(237, 242)	11 (8, 15)
NEWS	Results 10; Duration 30	251	(247, 256)	19 (14, 26)
NEWS	Results 10; Duration 50	252	(248, 257)	19 (14, 26)
Base: No Tool ^c	Results 3; Duration 5	226	(226, 226)	13 (11, 15)
Base: No Tool ^c	Results 3; Duration 7	231	(231, 231)	12 (11, 14)
Base: No Tool ^c	Results 2; Duration 5	220	(220, 220)	8 (7, 10)
Base: No Tool ^c	Results 2; Duration 7	226	(226, 226)	8 (7, 9)
Base: No Tool ^c	Results 2; Duration 14	229	(229, 229)	8 (7, 9)
Base: No Tool ^c	Results 5; Duration 30	247	(247, 247)	20 (18, 23)
Base: No Tool ^c	Results 5; Duration 50	248	(248, 248)	19 (17, 22)

Prognostic Tool	Time to Result, Infection Duration^a	Mortality Count	Prediction Intervals^b	Mortality Difference
Base: No Tool ^c	Results 10; Duration 30	266	(266, 266)	34 (30, 39)
Base: No Tool ^c	Results 10; Duration 50	267	(267, 267)	33 (30, 38)

Systemic Inflammatory Response Syndrome (SIRS), quick Sequential Organ Failure Assessment (qSOFA), National Early Warning Score (NEWS) Super HCT Bacterial Sepsis Learner (SHBSL), Clinical Factor-specific Super HCT Bacterial Sepsis Learner (C-SHBSL), HCT Bacteremia Sepsis Score (HBS2), HCT Bacteremia Sepsis Tree (HBST)

^a Duration of 50 analyses: mortality proportions extrapolated from 30-day constant estimates

^b 2.5-97.5% Percentiles for Mortality and Antibiotic Days

^c Base scenario: simulation run without a prediction tool (sensitivity: 0%, specificity 100%)

eTable 8: Simulated Antibiotic Use For 2000 Simulated Allogeneic HCT Recipients (Culture Result and Duration Lengths)

Prognostic Tool	Time to Result, Infection Duration^a	Antibiotic Count	Prediction Intervals^b
<i>All Antibiotic Days</i>			
SHBSL	Results 3; Duration 5	6217	(5844, 6657)
SHBSL	Results 3; Duration 7	6440	(6069, 6879)
SHBSL	Results 2; Duration 5	6257	(5881, 6696)
SHBSL	Results 2; Duration 7	6483	(6109, 6920)
SHBSL	Results 2; Duration 14	7212	(6842, 7647)
SHBSL	Results 5; Duration 30	8437	(8047, 8875)
SHBSL	Results 5; Duration 50	9682	(9286, 10145)
SHBSL	Results 10; Duration 30	8219	(7810, 8739)
SHBSL	Results 10; Duration 50	9402	(8962, 9973)
C-SHBSL	Results 3; Duration 5	6422	(6017, 6800)
C-SHBSL	Results 3; Duration 7	6639	(6235, 7013)
C-SHBSL	Results 2; Duration 5	6488	(6087, 6878)
C-SHBSL	Results 2; Duration 7	6709	(6309, 7096)
C-SHBSL	Results 2; Duration 14	7426	(7026, 7804)
C-SHBSL	Results 5; Duration 30	8498	(8064, 8843)
C-SHBSL	Results 5; Duration 50	9688	(9253, 10049)
C-SHBSL	Results 10; Duration 30	8141	(7689, 8538)
C-SHBSL	Results 10; Duration 50	9229	(8733, 9674)
HBS2	Results 3; Duration 5	7579	(7117, 8025)
HBS2	Results 3; Duration 7	7794	(7335, 8241)
HBS2	Results 2; Duration 5	7632	(7170, 8069)
HBS2	Results 2; Duration 7	7851	(7391, 8287)
HBS2	Results 2; Duration 14	8565	(8108, 9002)
HBS2	Results 5; Duration 30	9700	(9221, 10184)
HBS2	Results 5; Duration 50	10903	(10395, 11399)
HBS2	Results 10; Duration 30	9414	(8909, 9946)
HBS2	Results 10; Duration 50	10538	(10004, 11088)
HBST	Results 3; Duration 5	5811	(5414, 6156)
HBST	Results 3; Duration 7	6030	(5634, 6373)
HBST	Results 2; Duration 5	5879	(5488, 6225)
HBST	Results 2; Duration 7	6102	(5711, 6447)
HBST	Results 2; Duration 14	6822	(6430, 7163)
HBST	Results 5; Duration 30	7889	(7454, 8259)
HBST	Results 5; Duration 50	9081	(8632, 9475)
HBST	Results 10; Duration 30	7520	(7035, 7961)
HBST	Results 10; Duration 50	8607	(8092, 9092)
SIRS	Results 3; Duration 5	6099	(5684, 6534)
SIRS	Results 3; Duration 7	6316	(5902, 6750)
SIRS	Results 2; Duration 5	6168	(5754, 6601)
SIRS	Results 2; Duration 7	6390.0	(5977.4, 6821.4)

Prognostic Tool	Time to Result, Infection Duration^a	Antibiotic Count	Prediction Intervals^a
SIRS	Results 2; Duration 14	7107	(6697, 7536)
SIRS	Results 5; Duration 30	8161	(7773, 8598)
SIRS	Results 5; Duration 50	9347	(8956, 9784)
SIRS	Results 10; Duration 30	7783	(7375, 8233)
SIRS	Results 10; Duration 50	8862	(8444, 9338)
qSOFA	Results 3; Duration 5	385	(310, 470)
qSOFA	Results 3; Duration 7	607	(530, 693)
qSOFA	Results 2; Duration 5	519	(446, 597)
qSOFA	Results 2; Duration 7	751	(677, 829)
qSOFA	Results 2; Duration 14	1475	(1401, 1557)
qSOFA	Results 5; Duration 30	2185	(2072, 2304)
qSOFA	Results 5; Duration 50	3285	(3160, 3422)
qSOFA	Results 10; Duration 30	1426	(1282, 1591)
qSOFA	Results 10; Duration 50	2295	(2129, 2492)
NEWS	Results 3; Duration 5	4601	(4270, 4997)
NEWS	Results 3; Duration 7	4816	(4483, 5215)
NEWS	Results 2; Duration 5	4705	(4386, 5076)
NEWS	Results 2; Duration 7	4927	(4608, 5300)
NEWS	Results 2; Duration 14	5638	(5315, 6020)
NEWS	Results 5; Duration 30	6500	(6178, 6914)
NEWS	Results 5; Duration 50	7629	(7280, 8057)
NEWS	Results 10; Duration 30	5928	(5522, 6392)
NEWS	Results 10; Duration 50	6891	(6428, 7386)
Base: No Tool ^c	Results 3; Duration 5	236	(236, 236)
Base: No Tool ^c	Results 3; Duration 7	455	(455, 455)
Base: No Tool ^c	Results 2; Duration 5	379	(379, 379)
Base: No Tool ^c	Results 2; Duration 7	610	(610, 610)
Base: No Tool ^c	Results 2; Duration 14	1330	(1330, 1330)
Base: No Tool ^c	Results 5; Duration 30	1978	(1978, 1978)
Base: No Tool ^c	Results 5; Duration 50	3056	(3056, 3056)
Base: No Tool ^c	Results 10; Duration 30	1161	(1161, 1161)
Base: No Tool ^c	Results 10; Duration 50	1989	(1989, 1989)
<i>Antibiotic Days - Infected Individuals</i>			
SHBSL	Results 3; Duration 5	576	(533, 628)
SHBSL	Results 3; Duration 7	815	(769, 870)
SHBSL	Results 2; Duration 5	613	(583, 649)
SHBSL	Results 2; Duration 7	855	(824, 893)
SHBSL	Results 2; Duration 14	1611	(1573, 1655)
SHBSL	Results 5; Duration 30	2888	(2771, 3029)
SHBSL	Results 5; Duration 50	4180	(4034, 4355)
SHBSL	Results 10; Duration 30	2675	(2484, 2906)
SHBSL	Results 10; Duration 50	3902	(3660, 4195)
C-SHBSL	Results 3; Duration 5	496	(436, 555)

Prognostic Tool	Time to Result, Infection Duration^a	Antibiotic Count	Prediction Intervals^a
C-SHBSL	Results 3; Duration 7	729	(665, 792)
C-SHBSL	Results 2; Duration 5	557	(515, 598)
C-SHBSL	Results 2; Duration 7	795	(750, 839)
C-SHBSL	Results 2; Duration 14	1539	(1486, 1591)
C-SHBSL	Results 5; Duration 30	2668	(2501, 2829)
C-SHBSL	Results 5; Duration 50	3905	(3698, 4106)
C-SHBSL	Results 10; Duration 30	2319	(2049, 2580)
C-SHBSL	Results 10; Duration 50	3451	(3108, 3782)
HBS2	Results 3; Duration 5	535	(484, 574)
HBS2	Results 3; Duration 7	770	(715, 812)
HBS2	Results 2; Duration 5	584	(548, 611)
HBS2	Results 2; Duration 7	823	(784, 851)
HBS2	Results 2; Duration 14	1569	(1523, 1603)
HBS2	Results 5; Duration 30	2770	(2627, 2876)
HBS2	Results 5; Duration 50	4031	(3853, 4162)
HBS2	Results 10; Duration 30	2493	(2263, 2665)
HBS2	Results 10; Duration 50	3670	(3378, 3888)
HBST	Results 3; Duration 5	490	(445, 539)
HBST	Results 3; Duration 7	724	(675, 775)
HBST	Results 2; Duration 5	554	(522, 587)
HBST	Results 2; Duration 7	792	(758, 828)
HBST	Results 2; Duration 14	1536	(1496, 1580)
HBST	Results 5; Duration 30	2654	(2529, 2787)
HBST	Results 5; Duration 50	3889	(3734, 4055)
HBST	Results 10; Duration 30	2293	(2089, 2508)
HBST	Results 10; Duration 50	3419	(3161, 3691)
SIRS	Results 3; Duration 5	485	(425, 534)
SIRS	Results 3; Duration 7	718	(653, 770)
SIRS	Results 2; Duration 5	550	(507, 584)
SIRS	Results 2; Duration 7	788	(743, 824)
SIRS	Results 2; Duration 14	1530	(1477, 1574)
SIRS	Results 5; Duration 30	2638	(2472, 2773)
SIRS	Results 5; Duration 50	3869	(3663, 4036)
SIRS	Results 10; Duration 30	2269	(2000, 2487)
SIRS	Results 10; Duration 50	3388	(3047, 3664)
qSOFA	Results 3; Duration 5	270	(245, 297)
qSOFA	Results 3; Duration 7	492	(465, 521)
qSOFA	Results 2; Duration 5	403	(385, 422)
qSOFA	Results 2; Duration 7	635	(617, 656)
qSOFA	Results 2; Duration 14	1360	(1338, 1385)
qSOFA	Results 5; Duration 30	2071	(2002, 2146)
qSOFA	Results 5; Duration 50	3172	(3086, 3266)
qSOFA	Results 10; Duration 30	1313	(1201, 1435)

Prognostic Tool	Time to Result, Infection Duration^a	Antibiotic Count	Prediction Intervals^a
qSOFA	Results 10; Duration 50	2183	(2040, 2337)
NEWS	Results 3; Duration 5	375	(326, 424)
NEWS	Results 3; Duration 7	601	(548, 653)
NEWS	Results 2; Duration 5	474	(439, 508)
NEWS	Results 2; Duration 7	708	(671, 744)
NEWS	Results 2; Duration 14	1438	(1394, 1481)
NEWS	Results 5; Duration 30	2343	(2206, 2475)
NEWS	Results 5; Duration 50	3504	(3333, 3668)
NEWS	Results 10; Duration 30	1781	(1561, 1996)
NEWS	Results 10; Duration 50	2771	(2491, 3043)
Base: No Tool ^c	Results 3; Duration 5	236	(236, 236)
Base: No Tool ^c	Results 3; Duration 7	455	(455, 455)
Base: No Tool ^c	Results 2; Duration 5	379	(379, 379)
Base: No Tool ^c	Results 2; Duration 7	610	(610, 610)
Base: No Tool ^c	Results 2; Duration 14	1330	(1330, 1330)
Base: No Tool ^c	Results 5; Duration 30	1978	(1978, 1978)
Base: No Tool ^c	Results 5; Duration 50	3056	(3056, 3056)
Base: No Tool ^c	Results 10; Duration 30	1161	(1161, 1161)
Base: No Tool ^c	Results 10; Duration 50	1989	(1989, 1989)
<i>Antibiotic Days - Uninfected Individuals</i>			
SHBSL	Results 3; Duration 5	5641	(5245, 6078)
SHBSL	Results 3; Duration 7	5625	(5230, 6061)
SHBSL	Results 2; Duration 5	5644	(5246, 6081)
SHBSL	Results 2; Duration 7	5627	(5231, 6063)
SHBSL	Results 2; Duration 14	5601	(5206, 6035)
SHBSL	Results 5; Duration 30	5549	(5159, 5980)
SHBSL	Results 5; Duration 50	5502	(5114, 5929)
SHBSL	Results 10; Duration 30	5544	(5156, 5974)
SHBSL	Results 10; Duration 50	5499	(5113, 5926)
C-SHBSL	Results 3; Duration 5	5926	(5535, 6344)
C-SHBSL	Results 3; Duration 7	5910	(5519, 6326)
C-SHBSL	Results 2; Duration 5	5930	(5539, 6349)
C-SHBSL	Results 2; Duration 7	5914	(5523, 6331)
C-SHBSL	Results 2; Duration 14	5887	(5498, 6302)
C-SHBSL	Results 5; Duration 30	5830	(5445, 6241)
C-SHBSL	Results 5; Duration 50	5783	(5401, 6191)
C-SHBSL	Results 10; Duration 30	5821	(5436, 6230)
C-SHBSL	Results 10; Duration 50	5778	(5396, 6185)
HBS2	Results 3; Duration 5	7044	(6581, 7484)
HBS2	Results 3; Duration 7	7024	(6563, 7463)
HBS2	Results 2; Duration 5	7048	(6585, 7488)
HBS2	Results 2; Duration 7	7028	(6566, 7467)
HBS2	Results 2; Duration 14	6995	(6536, 7433)

Prognostic Tool	Time to Result, Infection Duration^a	Antibiotic Count	Prediction Intervals^a
HBS2	Results 5; Duration 30	6930	(6474, 7363)
HBS2	Results 5; Duration 50	6872	(6421, 7302)
HBS2	Results 10; Duration 30	6922	(6467, 7354)
HBS2	Results 10; Duration 50	6868	(6416, 7297)
HBST	Results 3; Duration 5	5321	(4925, 5702)
HBST	Results 3; Duration 7	5306	(4911, 5687)
HBST	Results 2; Duration 5	5325	(4929, 5708)
HBST	Results 2; Duration 7	5310	(4915, 5692)
HBST	Results 2; Duration 14	5286	(4892, 5666)
HBST	Results 5; Duration 30	5235	(4845, 5610)
HBST	Results 5; Duration 50	5192	(4805, 5566)
HBST	Results 10; Duration 30	5227	(4838, 5598)
HBST	Results 10; Duration 50	5188	(4801, 5560)
SIRS	Results 3; Duration 5	5614	(5203, 6042)
SIRS	Results 3; Duration 7	5599	(5188, 6026)
SIRS	Results 2; Duration 5	5618	(5207, 6047)
SIRS	Results 2; Duration 7	5602	(5192, 6030)
SIRS	Results 2; Duration 14	5577	(5168, 6002)
SIRS	Results 5; Duration 30	5523	(5118, 5944)
SIRS	Results 5; Duration 50	5478	(5077, 5896)
SIRS	Results 10; Duration 30	5514	(5110, 5935)
SIRS	Results 10; Duration 50	5474	(5072, 5891)
qSOFA	Results 3; Duration 5	115	(52, 188)
qSOFA	Results 3; Duration 7	115	(52, 188)
qSOFA	Results 2; Duration 5	116	(53, 188)
qSOFA	Results 2; Duration 7	115	(52, 188)
qSOFA	Results 2; Duration 14	115	(52, 187)
qSOFA	Results 5; Duration 30	113	(52, 185)
qSOFA	Results 5; Duration 50	113	(51, 184)
qSOFA	Results 10; Duration 30	113	(51, 184)
qSOFA	Results 10; Duration 50	112	(51, 183)
NEWS	Results 3; Duration 5	4226	(3923, 4566)
NEWS	Results 3; Duration 7	4215	(3912, 4554)
NEWS	Results 2; Duration 5	4231	(3927, 4571)
NEWS	Results 2; Duration 7	4219	(3916, 4558)
NEWS	Results 2; Duration 14	4200	(3899, 4537)
NEWS	Results 5; Duration 30	4157	(3859, 4492)
NEWS	Results 5; Duration 50	4126	(3829, 4457)
NEWS	Results 10; Duration 30	4147	(3850, 4482)
NEWS	Results 10; Duration 50	4120	(3825, 4452)
Base: No Tool ^c	Results 3; Duration 5	0	(0, 0)
Base: No Tool ^c	Results 3; Duration 7	0	(0, 0)
Base: No Tool ^c	Results 2; Duration 5	0	(0, 0)

Prognostic Tool	Time to Result, Infection Duration^a	Antibiotic Count	Prediction Intervals^a
Base: No Tool ^c	Results 2; Duration 7	0	(0, 0)
Base: No Tool ^c	Results 2; Duration 14	0	(0, 0)
Base: No Tool ^c	Results 5; Duration 30	0	(0, 0)
Base: No Tool ^c	Results 5; Duration 50	0	(0, 0)
Base: No Tool ^c	Results 10; Duration 30	0	(0, 0)
Base: No Tool ^c	Results 10; Duration 50	0	(0, 0)

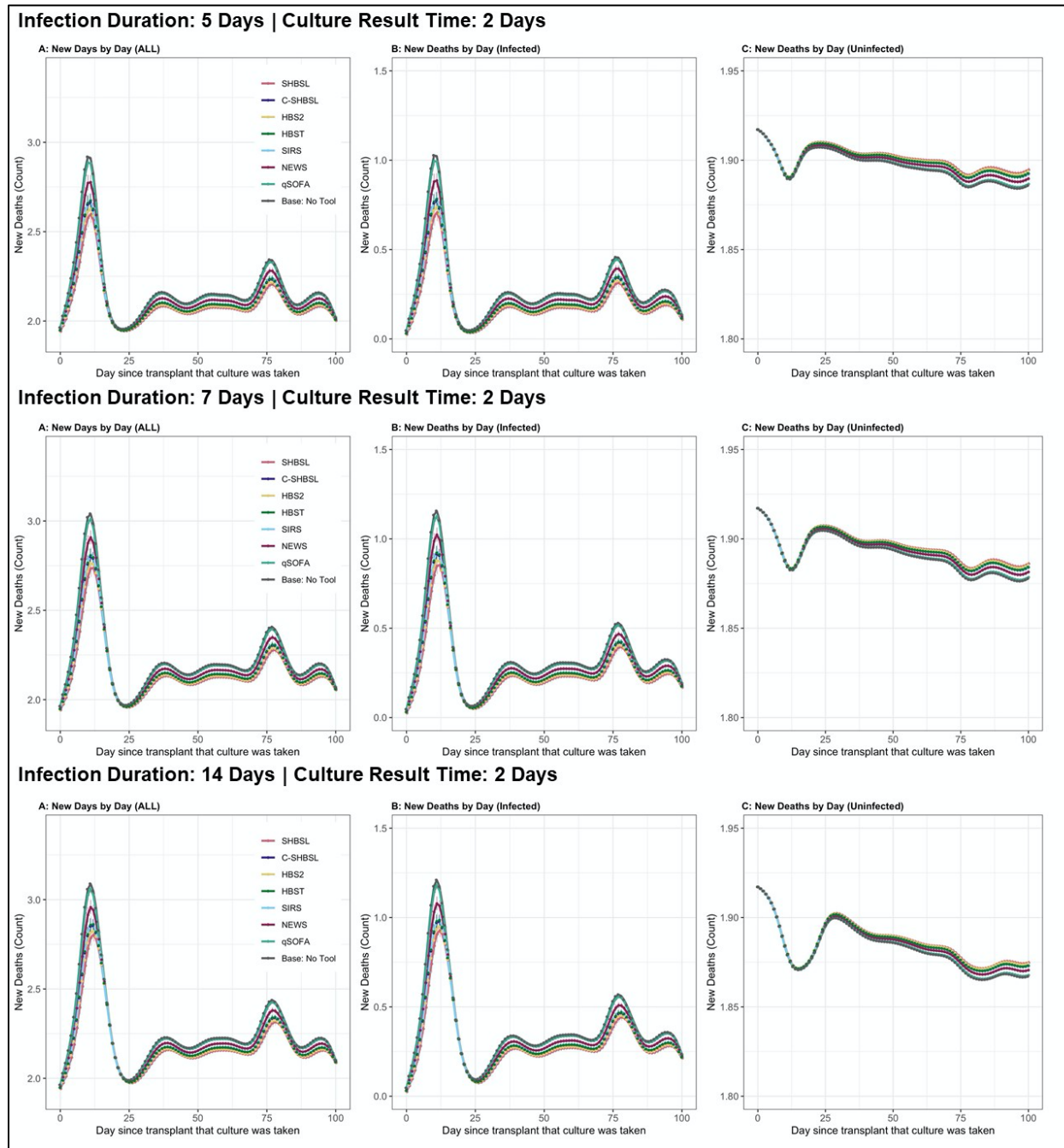
Systemic Inflammatory Response Syndrome (SIRS), quick Sequential Organ Failure Assessment (qSOFA), National Early Warning Score (NEWS) Super HCT Bacterial Sepsis Learner (SHBSL), Clinical Factor-specific Super HCT Bacterial Sepsis Learner (C-SHBSL), HCT Bacteremia Sepsis Score (HBS2), HCT Bacteremia Sepsis Tree (HBST)

^a Duration of 50 analyses: morality proportions extrapolated from 30-day constant estimates

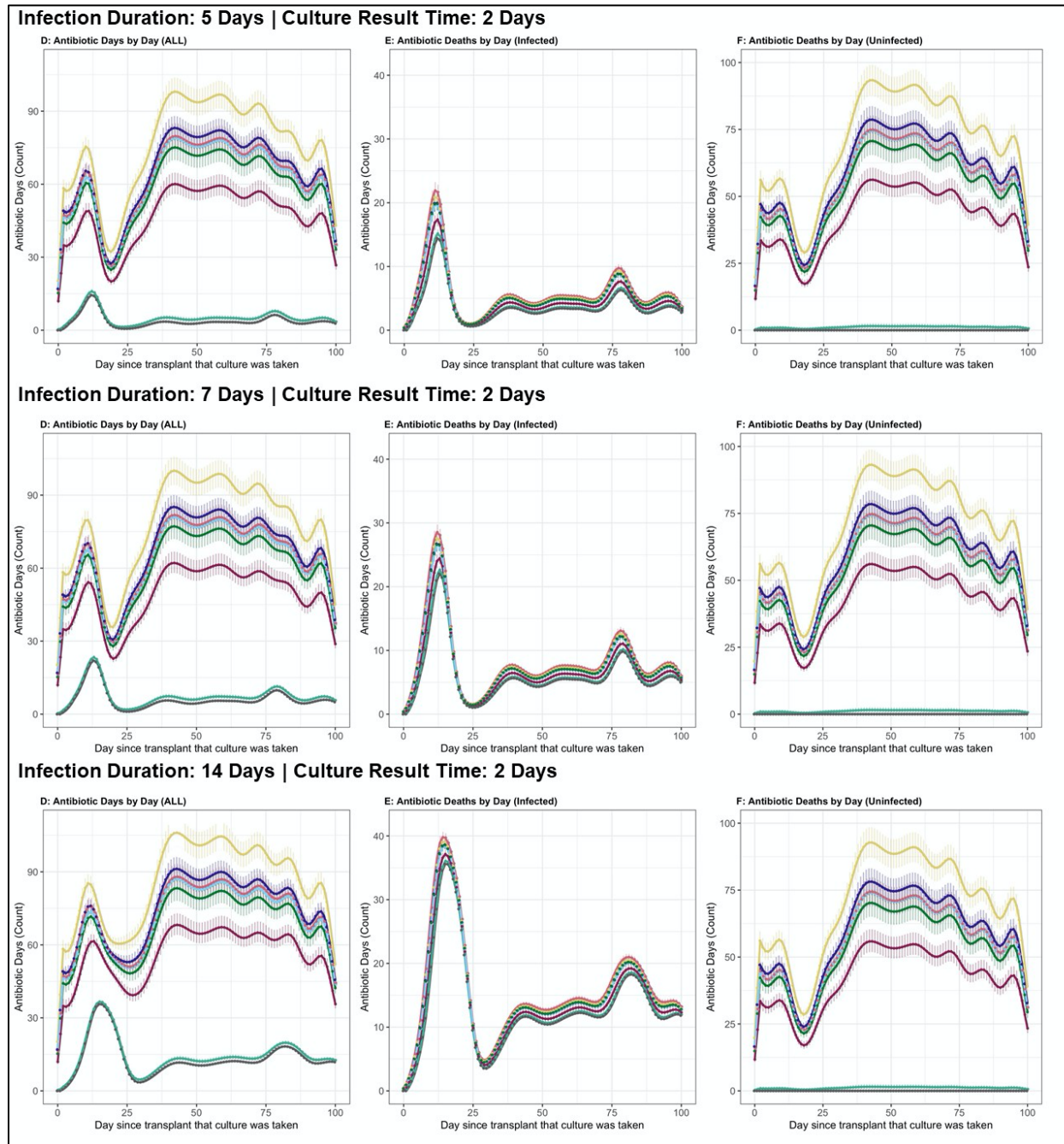
^b 2.5-97.5% Percentiles for Mortality and Antibiotic Days

^c Base scenario: simulation run without a prediction tool (sensitivity: 0%, specificity 100%)

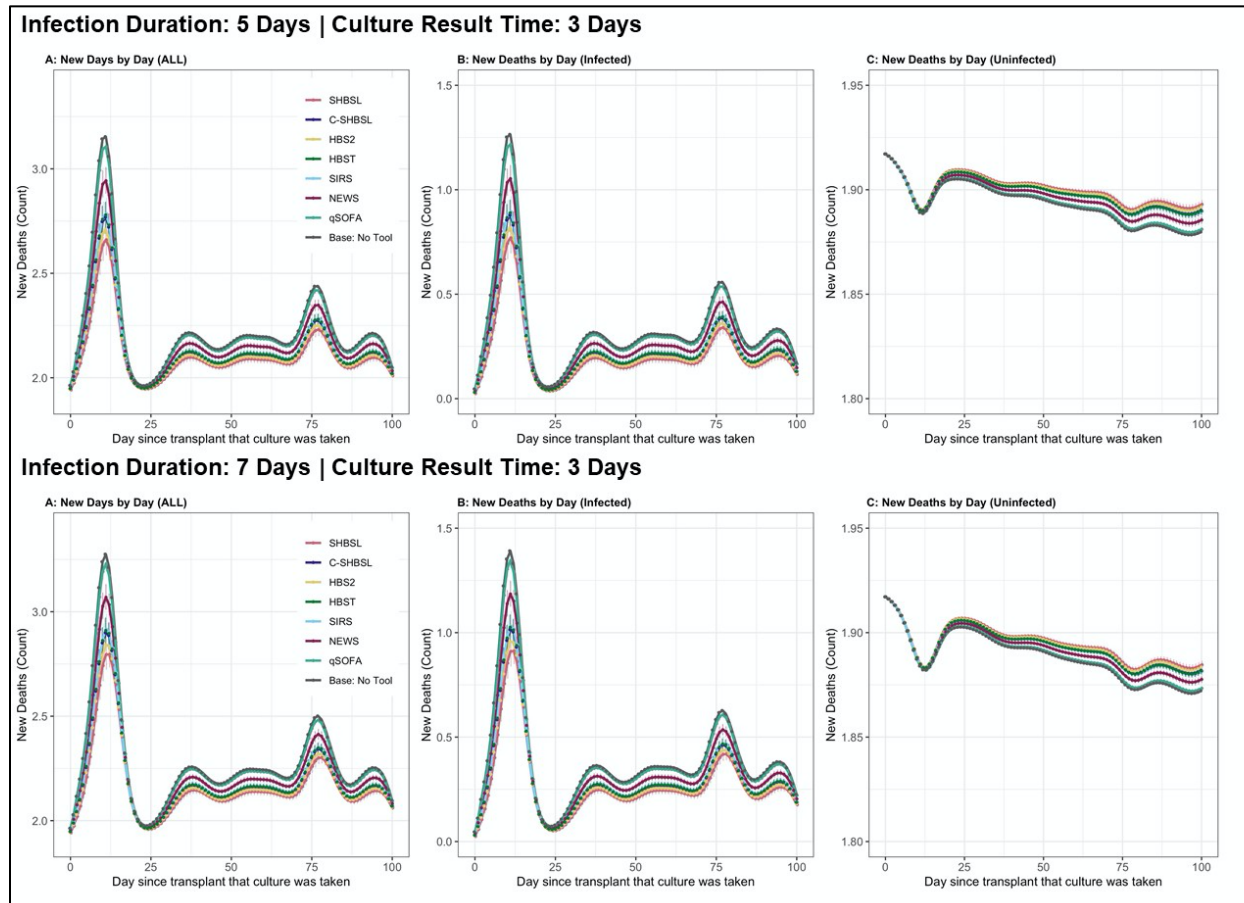
e.Figure 3: Mortality Events from Semi-Stochastic equation-based Model Results for Patients with Potential Bloodstream Infections (Primary Analysis Cohort) with 2 Day Culture Results and Varying Infection Duration



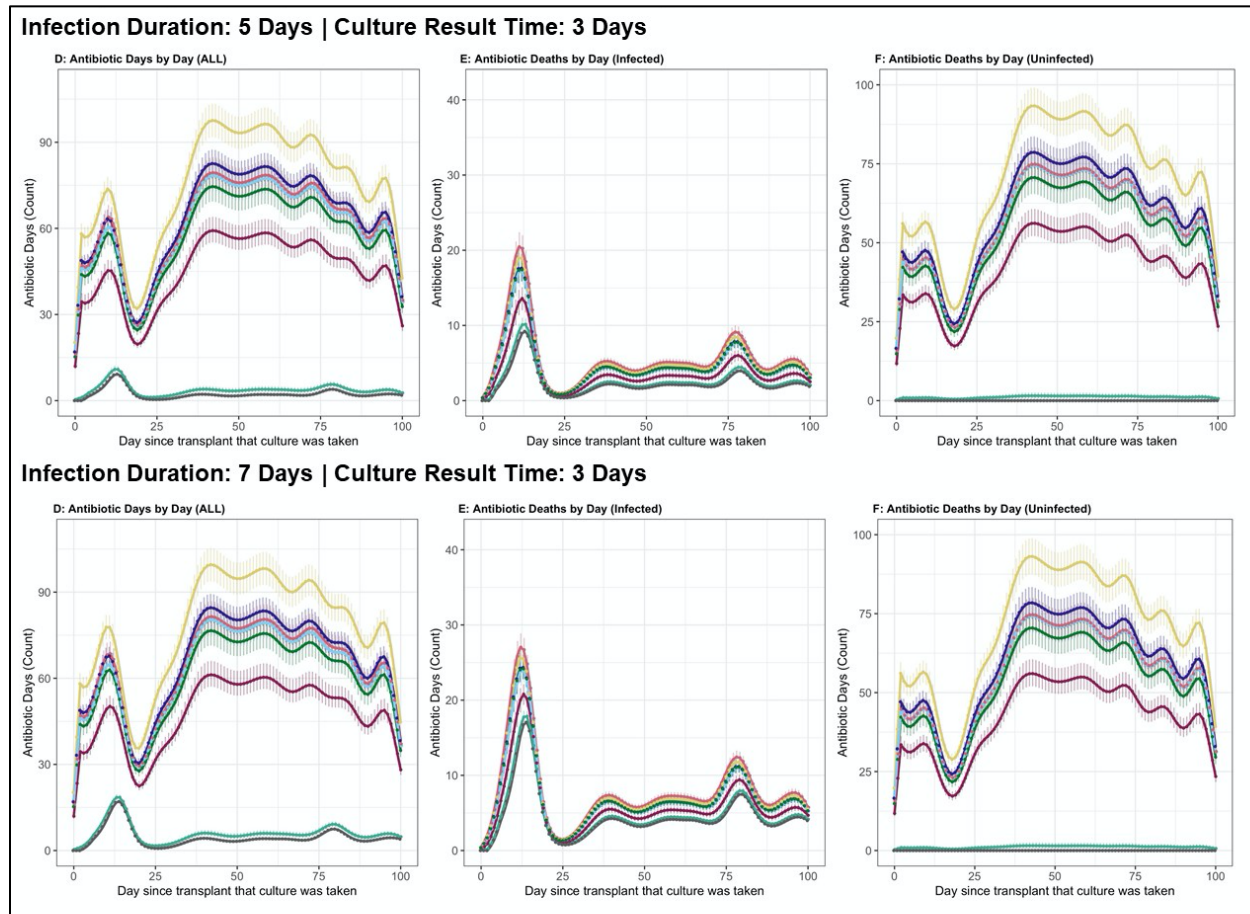
e.Figure 4: Antibiotic Use from Semi-Stochastic equation-based Model Results for Patients with Potential Bloodstream Infections (Primary Analysis Cohort) with 2 Day Culture Results and Varying Infection Duration



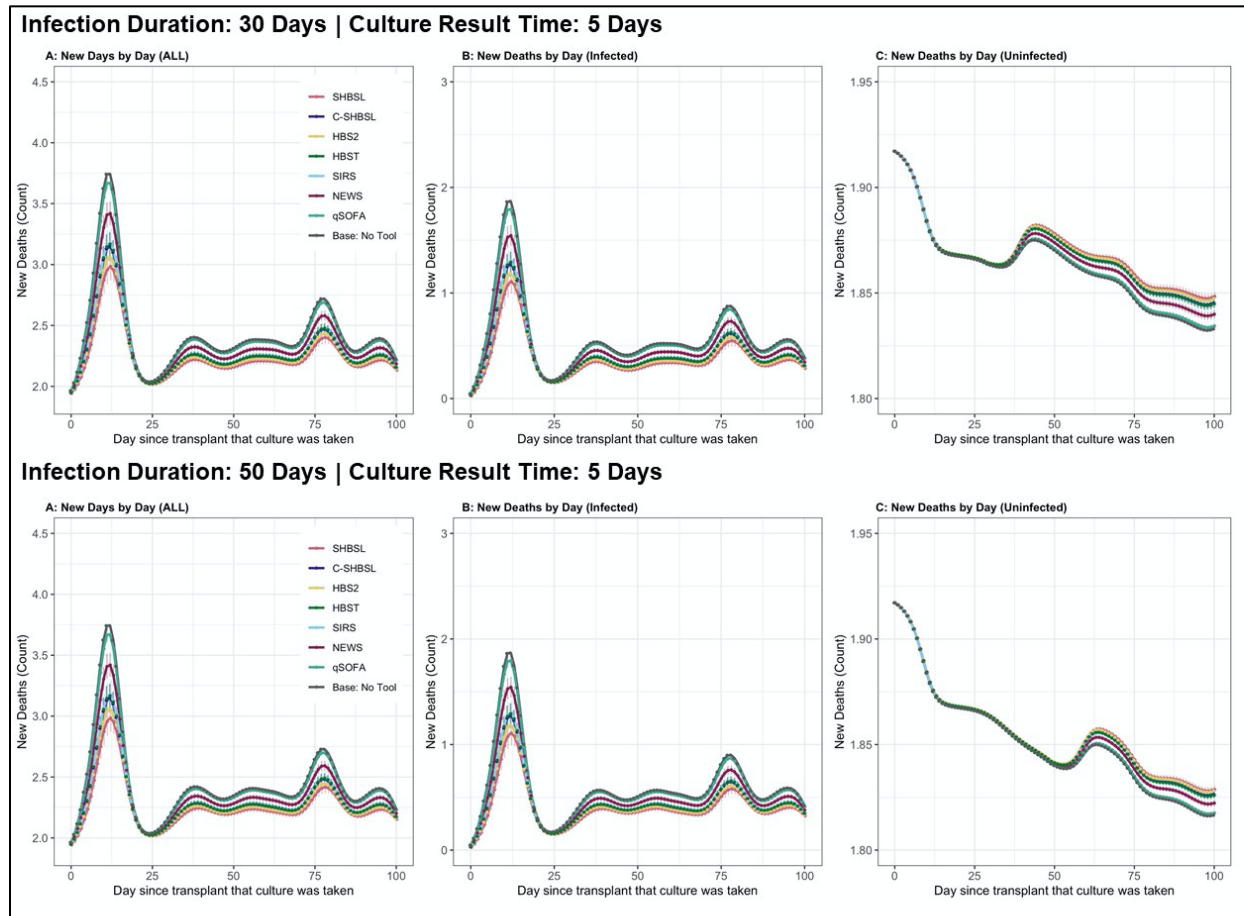
e.Figure 5: Mortality Events from Semi-Stochastic equation-based Model Results for Patients with Potential Bloodstream Infections (Primary Analysis Cohort) with 3 Day Culture Results and Varying Infection Duration



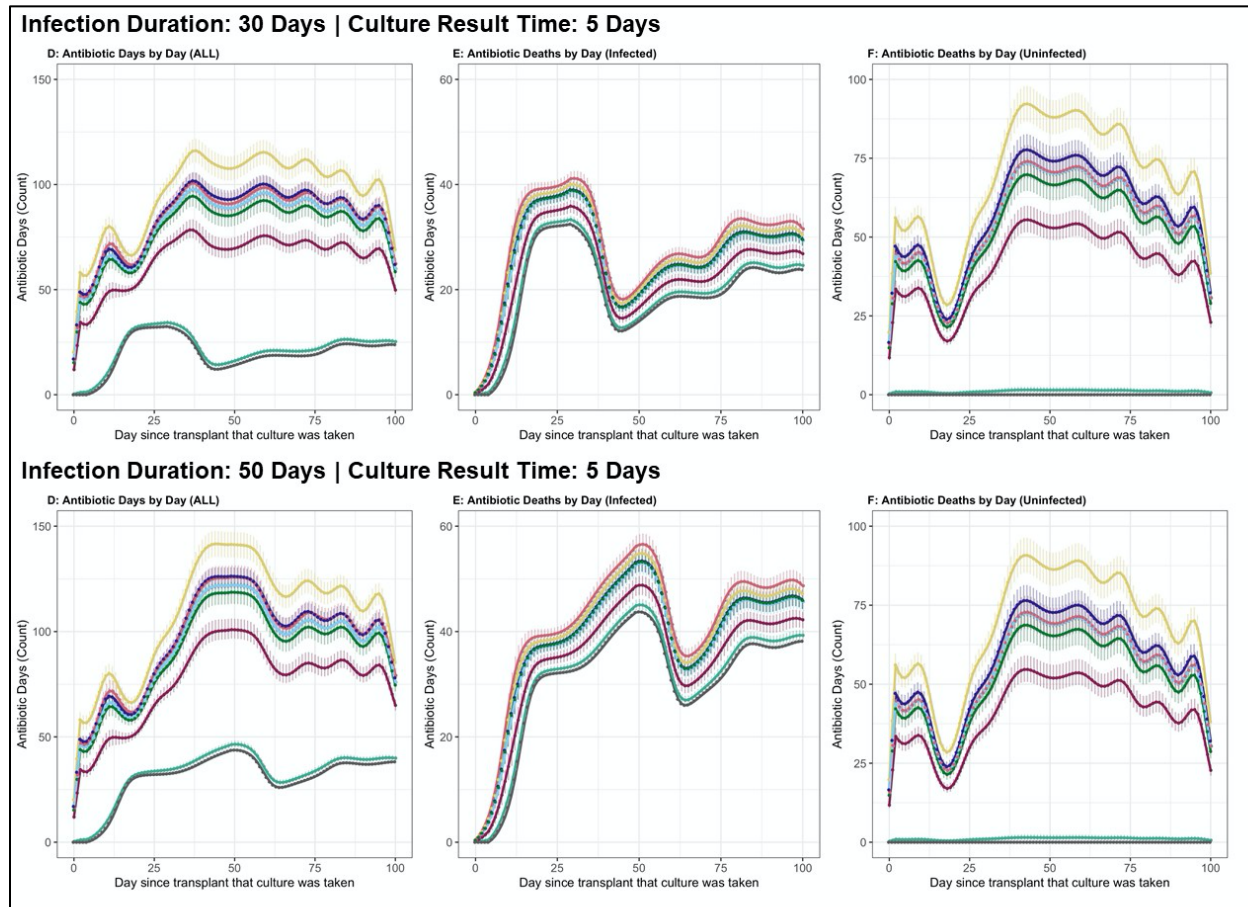
e.Figure 6: Antibiotic Use from Semi-Stochastic equation-based Model Results for Patients with Potential Bloodstream Infections (Primary Analysis Cohort) with 3 Day Culture Results and Varying Infection Duration



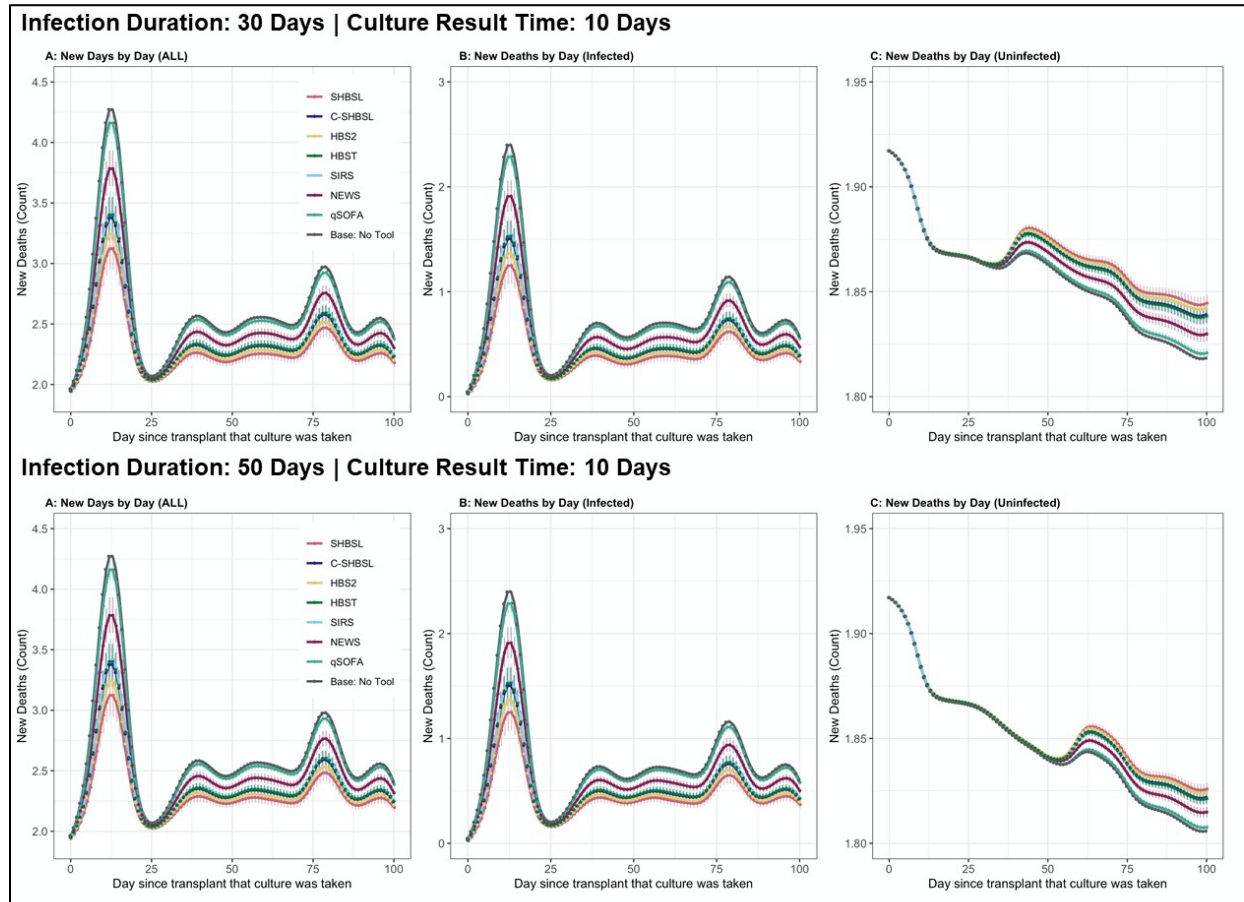
e.Figure 7: Mortality Events from Semi-Stochastic equation-based Model Results for Patients with Potential Bloodstream Infections (Primary Analysis Cohort) with 5 Day Culture Results and Varying Infection Duration



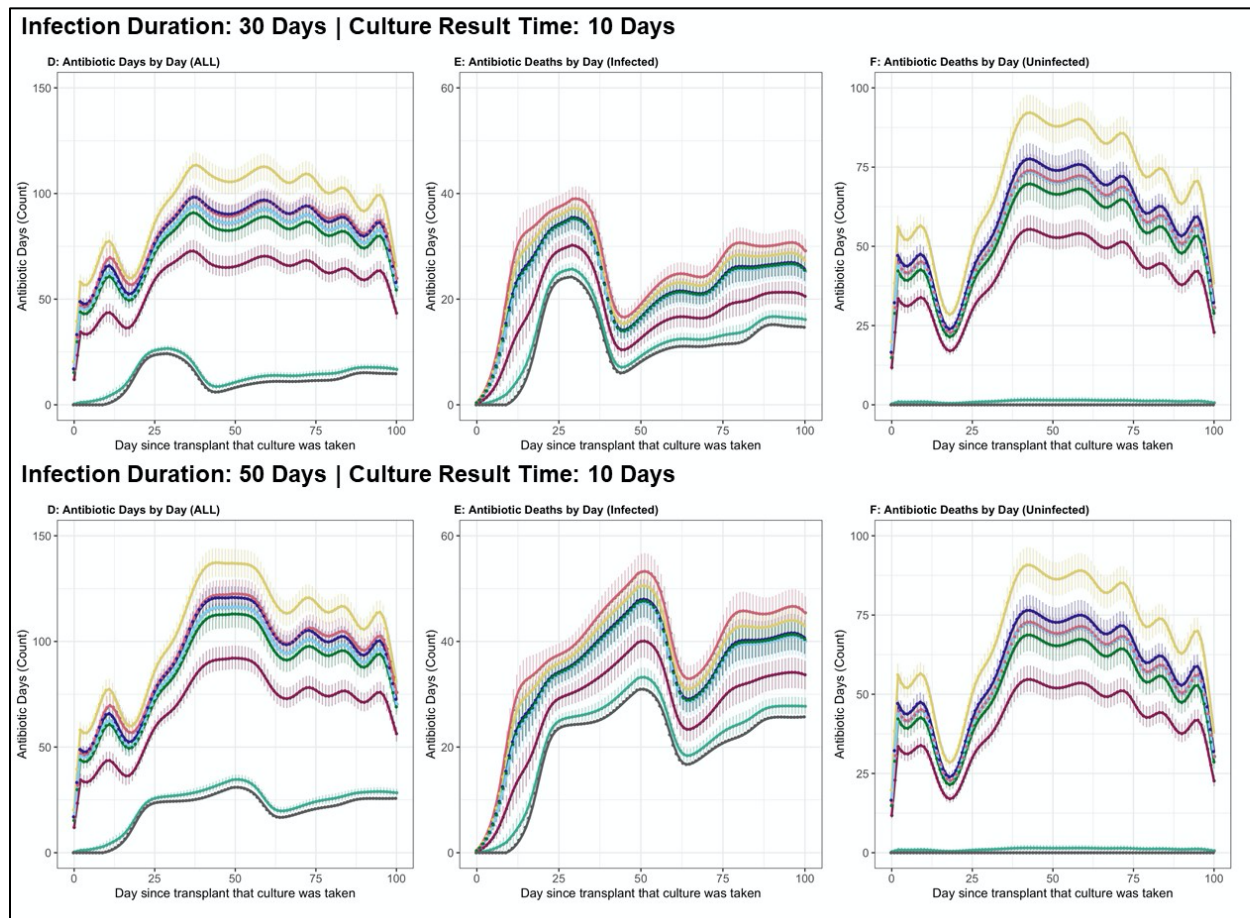
e.Figure 8: Antibiotic Use from Semi-Stochastic equation-based Model Results for Patients with Potential Bloodstream Infections (Primary Analysis Cohort) with 5 Day Culture Results and Varying Infection Duration



e.Figure 9: Mortality Events from Semi-Stochastic equation-based Model Results for Patients with Potential Bloodstream Infections (Primary Analysis Cohort) with 10 Day Culture Results and Varying Infection Duration



e.Figure 10: Antibiotic Use from Semi-Stochastic equation-based Model Results for Patients with Potential Bloodstream Infections (Primary Analysis Cohort) with 10 Day Culture Results and Varying Infection Duration



Random Disagreement Between Tool Classification and Antibiotic Treatment

eTable 9: Simulated Mortality For 2000 Allogeneic HCT Recipients (Random Disagreement)

Prognostic Tool	Disagreement Percentage	Mortality Count	Prediction Intervals ^a	Mortality Difference
SHBSL	Disagreement 100%	218	(217, 219)	0 (0, 0)
SHBSL	Disagreement 70%	217	(216, 217)	0 (0, 0)
SHBSL	Disagreement 50%	432	(432, 432)	0 (0, 0)
SHBSL	Disagreement 30%	215	(215, 215)	0 (0, 0)
SHBSL	Disagreement 10%	214	(213, 215)	0 (0, 0)
C-SHBSL	Disagreement 100%	216	(215, 217)	-2 (-4, 0)
C-SHBSL	Disagreement 70%	216	(216, 216)	-1 (-1, 0)
C-SHBSL	Disagreement 50%	432	(432, 432)	0 (0, 0)
C-SHBSL	Disagreement 30%	216	(215, 216)	1 (0, 1)
C-SHBSL	Disagreement 10%	216	(215, 217)	1 (0, 3)
HBS2	Disagreement 100%	217	(216, 218)	-1 (-3, 1)
HBS2	Disagreement 70%	216	(216, 217)	0 (-1, 0)
HBS2	Disagreement 50%	432	(432, 432)	0 (0, 0)
HBS2	Disagreement 30%	215	(215, 216)	0 (0, 1)
HBS2	Disagreement 10%	215	(214, 216)	1 (0, 2)
HBST	Disagreement 100%	216	(215, 217)	-2 (-3, 0)
HBST	Disagreement 70%	216	(215, 216)	-1 (-1, 0)
HBST	Disagreement 50%	432	(432, 432)	0 (0, 0)
HBST	Disagreement 30%	216	(215, 216)	1 (0, 1)
HBST	Disagreement 10%	216	(215, 217)	2 (0, 3)
SIRS	Disagreement 100%	216	(215, 217)	-2 (-3, -1)
SIRS	Disagreement 70%	216	(215, 216)	-1 (-1, 0)
SIRS	Disagreement 50%	432	(432, 432)	0 (0, 0)
SIRS	Disagreement 30%	216	(215, 216)	1 (0, 1)
SIRS	Disagreement 10%	216	(215, 217)	2 (1, 3)
qSOFA	Disagreement 100%	211	(210, 211)	-7 (-8, -6)
qSOFA	Disagreement 70%	214	(214, 214)	-3 (-3, -2)
qSOFA	Disagreement 50%	432	(432, 432)	0 (0, 0)
qSOFA	Disagreement 30%	218	(218, 218)	3 (2, 3)
qSOFA	Disagreement 10%	220	(220, 221)	6 (5, 7)
NEWS	Disagreement 100%	213	(212, 214)	-4 (-6, -3)
NEWS	Disagreement 70%	215	(214, 215)	-2 (-2, -1)
NEWS	Disagreement 50%	432	(432, 432)	0 (0, 0)
NEWS	Disagreement 30%	217	(216, 217)	2 (1, 2)
NEWS	Disagreement 10%	218	(217, 219)	4 (3, 5)

Prognostic Tool	Disagreement Percentage	Mortality Count	Prediction Intervals^a	Mortality Difference
Base: No Tool ^b	Disagreement 100%	210	(210, 210)	-8 (-9, -7)
Base: No Tool ^b	Disagreement 70%	213	(213, 213)	-3 (-4, -3)
Base: No Tool ^b	Disagreement 50%	432	(432, 432)	0 (0, 0)
Base: No Tool ^b	Disagreement 30%	218	(218, 218)	3 (3, 4)
Base: No Tool ^b	Disagreement 10%	221	(221, 221)	7 (6, 7)
Base: No Tool ^b	Disagreement 10%	210	(210, 210)	-8 (-9, -7)

Systemic Inflammatory Response Syndrome (SIRS), quick Sequential Organ Failure Assessment (qSOFA), National Early Warning Score (NEWS) Super HCT Bacterial Sepsis Learner (SHBSL), Clinical Factor-specific Super HCT Bacterial Sepsis Learner (C-SHBSL), HCT Bacteremia Sepsis Score (HBS2), HCT Bacteremia Sepsis Tree (HBST)

^a 2.5-97.5% Percentiles for Mortality and Antibiotic Days

^b Base scenario: simulation run without a prediction tool (sensitivity: 0%, specificity 100%)

eTable 10: Simulated Antibiotic Use For 2000 Simulated Allogeneic HCT Recipients (Random Disagreement)

Prognostic Tool	Disagreement Percentage	Antibiotic Count	Prediction Intervals ^a
<i>All Antibiotic Days</i>			
SHBSL	Disagreement 100%	16215	(15849, 16523)
SHBSL	Disagreement 70%	13289	(13139, 13415)
SHBSL	Disagreement 50%	22595	(22595, 22595)
SHBSL	Disagreement 30%	9273	(9142, 9427)
SHBSL	Disagreement 10%	7214	(6948, 7528)
C-SHBSL	Disagreement 100%	16052	(15740, 16390)
C-SHBSL	Disagreement 70%	13221	(13093, 13359)
C-SHBSL	Disagreement 50%	22595	(22595, 22595)
C-SHBSL	Disagreement 30%	9344	(9202, 9477)
C-SHBSL	Disagreement 10%	7361	(7072, 7630)
HBS2	Disagreement 100%	15077	(14697, 15466)
HBS2	Disagreement 70%	12823	(12669, 12982)
HBS2	Disagreement 50%	22595	(22595, 22595)
HBS2	Disagreement 30%	9752	(9589, 9910)
HBS2	Disagreement 10%	8187	(7857, 8506)
HBST	Disagreement 100%	16559	(16274, 16887)
HBST	Disagreement 70%	13429	(13312, 13563)
HBST	Disagreement 50%	22595	(22595, 22595)
HBST	Disagreement 30%	9130	(8991, 9251)
HBST	Disagreement 10%	6925	(6642, 7171)
SIRS	Disagreement 100%	16322	(15959, 16666)
SIRS	Disagreement 70%	13331	(13183, 13472)
SIRS	Disagreement 50%	22595	(22595, 22595)
SIRS	Disagreement 30%	9231	(9085, 9383)
SIRS	Disagreement 10%	7130	(6834, 7440)
qSOFA	Disagreement 100%	20923	(20853, 20984)
qSOFA	Disagreement 70%	15231	(15202, 15256)
qSOFA	Disagreement 50%	22595	(22595, 22595)
qSOFA	Disagreement 30%	7247	(7220, 7276)
qSOFA	Disagreement 10%	3073	(3019, 3134)
NEWS	Disagreement 100%	17565	(17234, 17840)
NEWS	Disagreement 70%	13840	(13705, 13953)
NEWS	Disagreement 50%	22595	(22595, 22595)
NEWS	Disagreement 30%	8705	(8589, 8844)
NEWS	Disagreement 10%	6062	(5826, 6344)
Base: No Tool ^b	Disagreement 100%	21045	(21045, 21045)
Base: No Tool ^b	Disagreement 70%	15281	(15281, 15281)

Prognostic Tool	Disagreement Percentage	Antibiotic Count	Prediction Intervals^a
Base: No Tool ^b	Disagreement 50%	22595	(22595, 22595)
Base: No Tool ^b	Disagreement 30%	7194	(7194, 7194)
Base: No Tool ^b	Disagreement 10%	2967	(2967, 2967)
<i>Antibiotic Days - Infected Individuals</i>			
SHBSL	Disagreement 100%	871	(826, 909)
SHBSL	Disagreement 70%	939	(921, 954)
SHBSL	Disagreement 50%	1970	(1970, 1970)
SHBSL	Disagreement 30%	1032	(1017, 1050)
SHBSL	Disagreement 10%	1079	(1048, 1116)
C-SHBSL	Disagreement 100%	941	(890, 994)
C-SHBSL	Disagreement 70%	967	(947, 989)
C-SHBSL	Disagreement 50%	1970	(1970, 1970)
C-SHBSL	Disagreement 30%	1003	(981, 1024)
C-SHBSL	Disagreement 10%	1021	(978, 1064)
HBS2	Disagreement 100%	908	(874, 954)
HBS2	Disagreement 70%	954	(940, 973)
HBS2	Disagreement 50%	1970	(1970, 1970)
HBS2	Disagreement 30%	1016	(998, 1030)
HBS2	Disagreement 10%	1048	(1011, 1076)
HBST	Disagreement 100%	946	(903, 985)
HBST	Disagreement 70%	969	(952, 985)
HBST	Disagreement 50%	1970	(1970, 1970)
HBST	Disagreement 30%	1001	(985, 1019)
HBST	Disagreement 10%	1018	(985, 1053)
SIRS	Disagreement 100%	951	(908, 1003)
SIRS	Disagreement 70%	971	(954, 992)
SIRS	Disagreement 50%	1970	(1970, 1970)
SIRS	Disagreement 30%	999	(978, 1017)
SIRS	Disagreement 10%	1014	(970, 1049)
qSOFA	Disagreement 100%	1126	(1102, 1147)
qSOFA	Disagreement 70%	1042	(1033, 1051)
qSOFA	Disagreement 50%	1970	(1970, 1970)
qSOFA	Disagreement 30%	926	(917, 936)
qSOFA	Disagreement 10%	866	(848, 885)
NEWS	Disagreement 100%	1044	(1002, 1087)
NEWS	Disagreement 70%	1009	(992, 1026)
NEWS	Disagreement 50%	1970	(1970, 1970)
NEWS	Disagreement 30%	961	(943, 978)
NEWS	Disagreement 10%	936	(900, 971)
Base: No Tool ^b	Disagreement 100%	1154	(1154, 1154)

Prognostic Tool	Disagreement Percentage	Antibiotic Count	Prediction Intervals^a
Base: No Tool ^b	Disagreement 70%	1054	(1054, 1054)
Base: No Tool ^b	Disagreement 50%	1970	(1970, 1970)
Base: No Tool ^b	Disagreement 30%	914	(914, 914)
Base: No Tool ^b	Disagreement 10%	841	(841, 841)
<i>Antibiotic Days - Uninfected Individuals</i>			
SHBSL	Disagreement 100%	15344	(14980, 15670)
SHBSL	Disagreement 70%	12350	(12201, 12484)
SHBSL	Disagreement 50%	20624	(20624, 20624)
SHBSL	Disagreement 30%	8241	(8102, 8395)
SHBSL	Disagreement 10%	6135	(5851, 6447)
C-SHBSL	Disagreement 100%	15110	(14763, 15437)
C-SHBSL	Disagreement 70%	12253	(12111, 12387)
C-SHBSL	Disagreement 50%	20624	(20624, 20624)
C-SHBSL	Disagreement 30%	8341	(8203, 8488)
C-SHBSL	Disagreement 10%	6339	(6060, 6638)
HBS2	Disagreement 100%	14169	(13795, 14560)
HBS2	Disagreement 70%	11869	(11717, 12029)
HBS2	Disagreement 50%	20624	(20624, 20624)
HBS2	Disagreement 30%	8736	(8572, 8892)
HBS2	Disagreement 10%	7139	(6807, 7455)
HBST	Disagreement 100%	15614	(15301, 15939)
HBST	Disagreement 70%	12459	(12331, 12593)
HBST	Disagreement 50%	20624	(20624, 20624)
HBST	Disagreement 30%	8128	(7990, 8262)
HBST	Disagreement 10%	5907	(5624, 6180)
SIRS	Disagreement 100%	15371	(15014, 15712)
SIRS	Disagreement 70%	12360	(12214, 12500)
SIRS	Disagreement 50%	20624	(20624, 20624)
SIRS	Disagreement 30%	8231	(8087, 8382)
SIRS	Disagreement 10%	6116	(5822, 6422)
qSOFA	Disagreement 100%	19798	(19739, 19849)
qSOFA	Disagreement 70%	14189	(14164, 14210)
qSOFA	Disagreement 50%	20624	(20624, 20624)
qSOFA	Disagreement 30%	6320	(6298, 6346)
qSOFA	Disagreement 10%	2207	(2163, 2259)
NEWS	Disagreement 100%	16521	(16241, 16767)
NEWS	Disagreement 70%	12831	(12716, 12933)
NEWS	Disagreement 50%	20624	(20624, 20624)
NEWS	Disagreement 30%	7744	(7639, 7863)
NEWS	Disagreement 10%	5126	(4910, 5368)

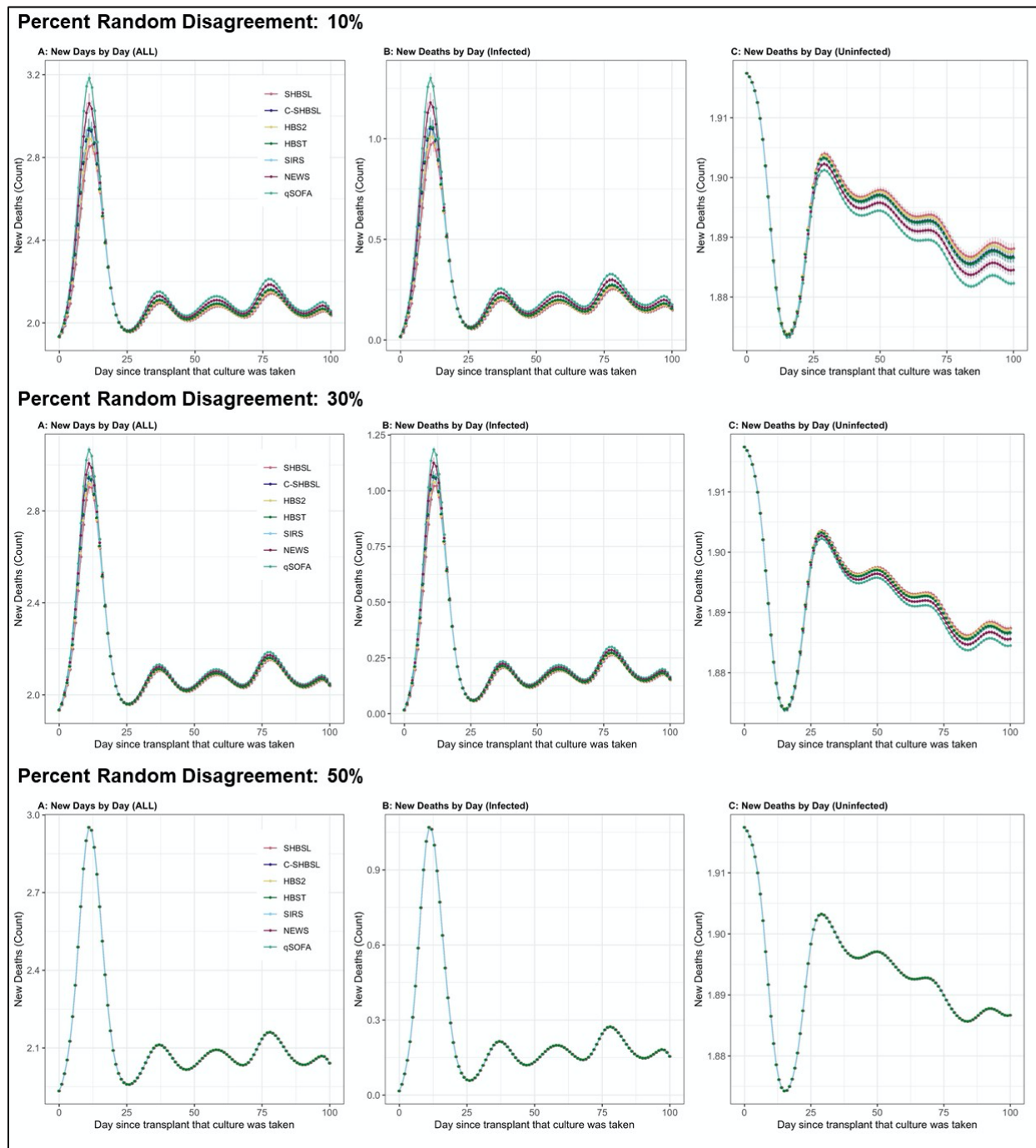
Prognostic Tool	Disagreement Percentage	Antibiotic Count	Prediction Intervals^a
Base: No Tool ^b	Disagreement 100%	19891	(19891, 19891)
Base: No Tool ^b	Disagreement 70%	14227	(14227, 14227)
Base: No Tool ^b	Disagreement 50%	20624	(20624, 20624)
Base: No Tool ^b	Disagreement 30%	6280	(6280, 6280)
Base: No Tool ^b	Disagreement 10%	2125	(2125, 2125)

Systemic Inflammatory Response Syndrome (SIRS), quick Sequential Organ Failure Assessment (qSOFA), National Early Warning Score (NEWS) Super HCT Bacterial Sepsis Learner (SHDSL), Clinical Factor-specific Super HCT Bacterial Sepsis Learner (C-SHDSL), HCT Bacteremia Sepsis Score (HBS2), HCT Bacteremia Sepsis Tree (HBST)

^a 2.5-97.5% Percentiles for Mortality and Antibiotic Days

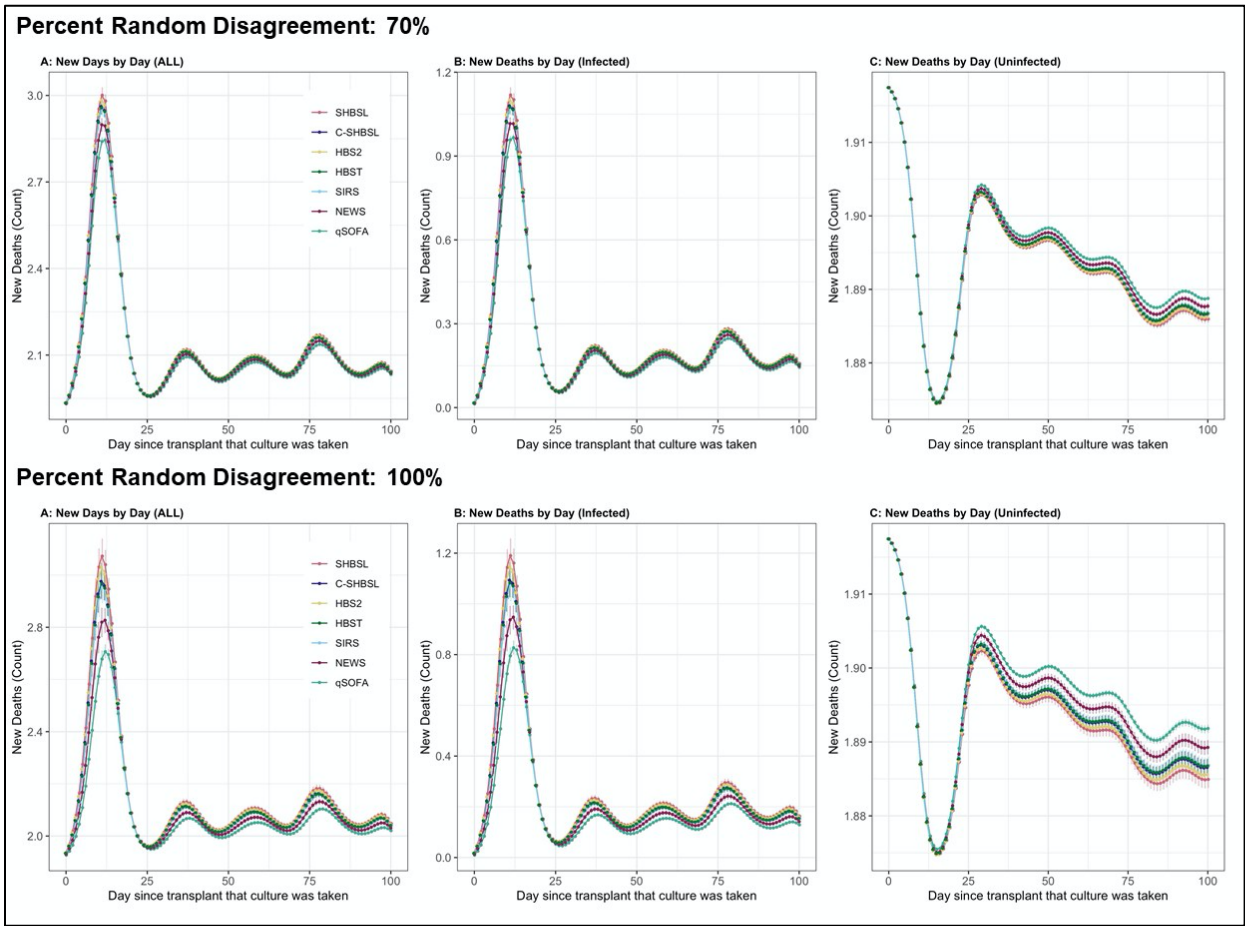
^b Base scenario: simulation run without a prediction tool (sensitivity: 0%, specificity 100%)

e.Figure 11: Morality from Semi-Stochastic equation-based Model Results for Patients with Potential Bloodstream Infections (Primary Analysis Cohort) with Random Disagreement Between Tool Classification and Antibiotic Treatment

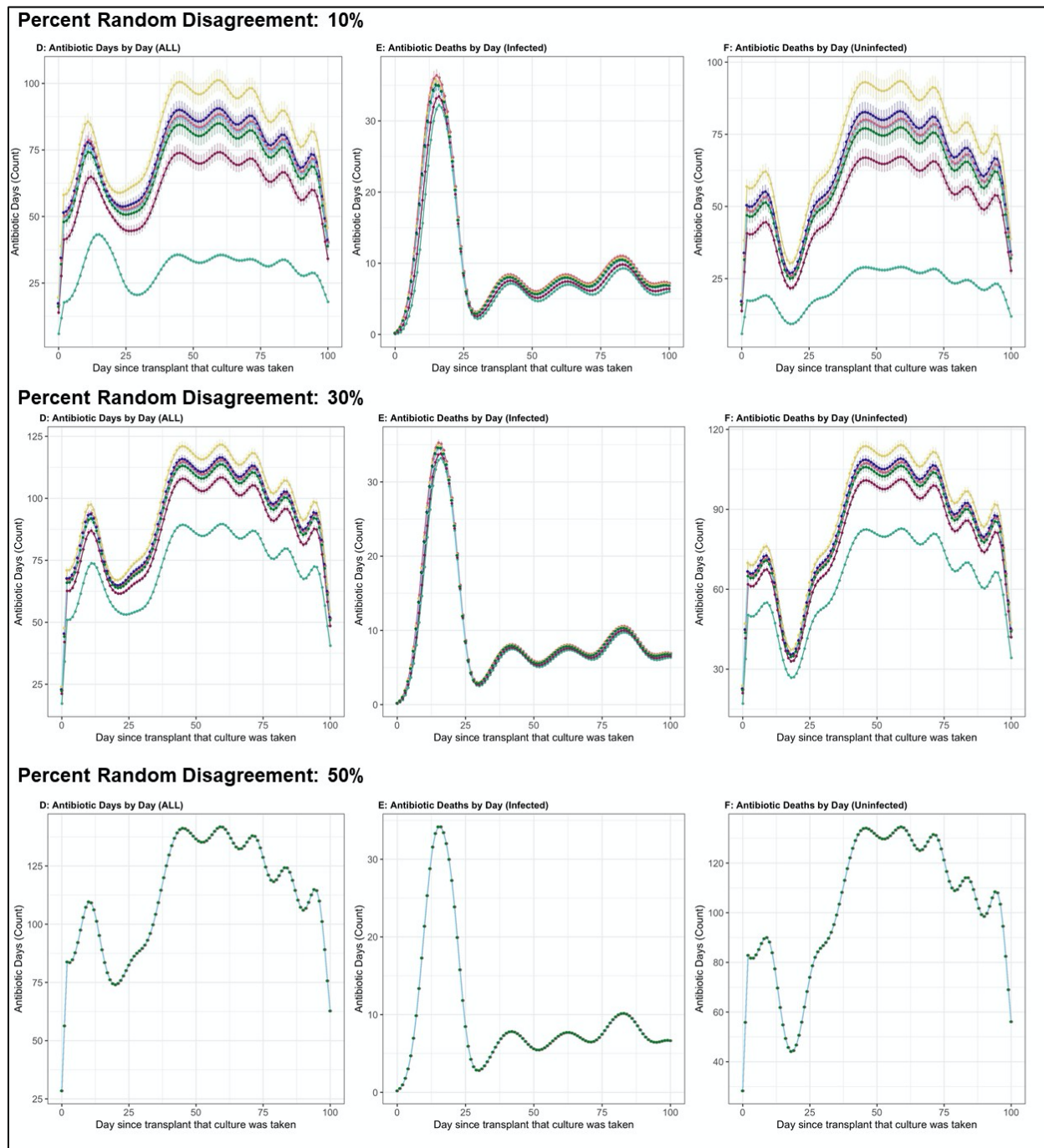


e.Figure 11 (Continued): Morality from Semi-Stochastic equation-based Model Results for Patients with Potential Bloodstream Infections (Primary Analysis Cohort) with Random Disagreement Between Tool

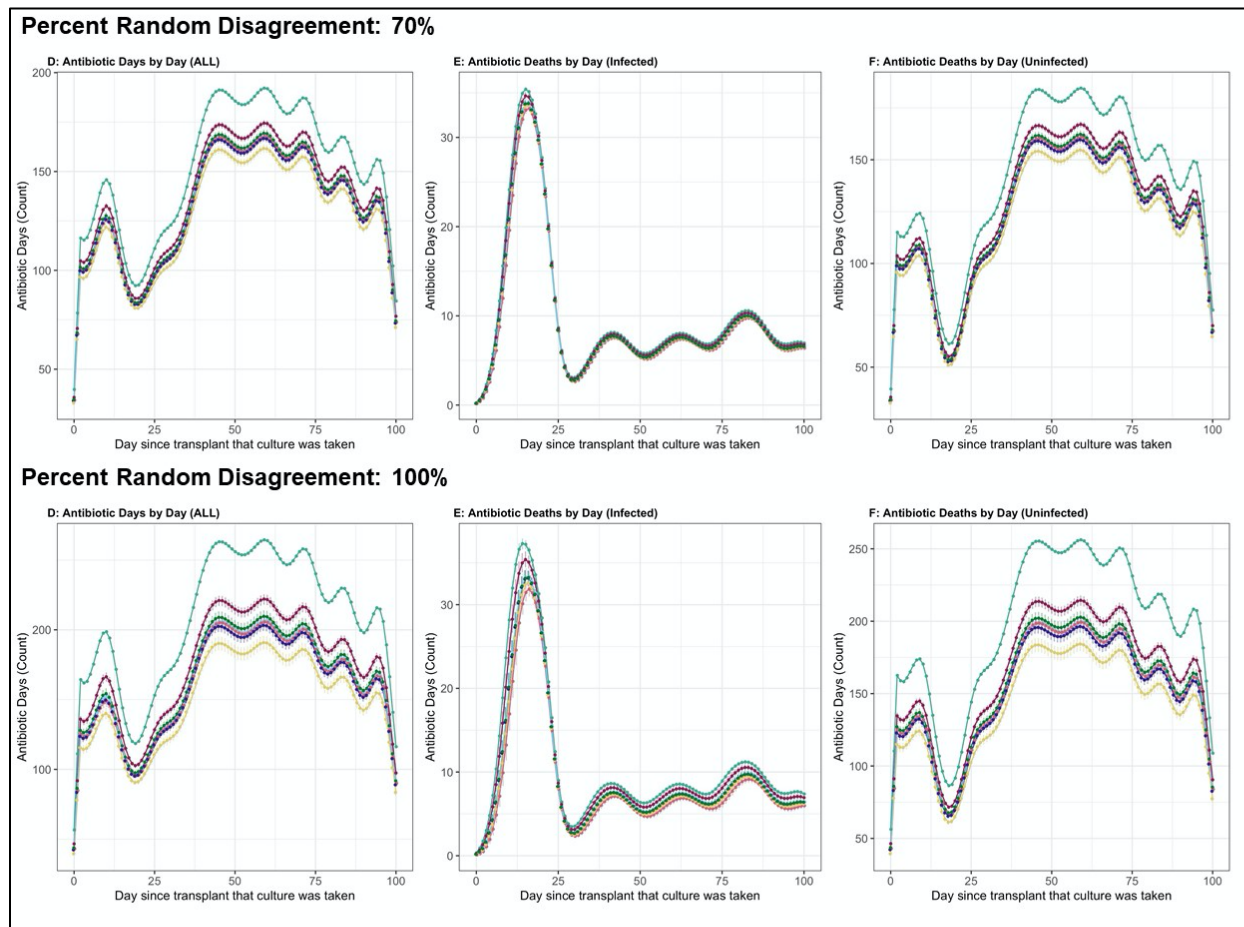
Classification and Antibiotic Treatment



e.Figure 12: Antibiotic Use from Semi-Stochastic equation-based Model Results for Patients with Potential Bloodstream Infections (Primary Analysis Cohort) with Random Disagreement Between Tool Classification and Antibiotic Treatment



e.Figure 12 (Continued): Antibiotic Use from Semi-Stochastic equation-based Model Results for Patients with Potential Bloodstream Infections (Primary Analysis Cohort) with Random Disagreement Between Tool Classification and Antibiotic Treatment



Correct Reclassification

eTable 11: Simulated Mortality For 2000 Allogeneic HCT Recipients (Correct Reclassification)

Prognostic Tool	Reclassification Percentage	Mortality Count	Prediction Intervals ^a	Mortality Difference
SHBSL	Reclassify 5%	214	(213, 215)	0 (0, 0)
SHBSL	Reclassify 10%	214	(212, 214)	0 (0, 0)
SHBSL	Reclassify 15%	213	(212, 214)	0 (0, 0)
SHBSL	Reclassify 20%	213	(212, 214)	0 (0, 0)
C-SHBSL	Reclassify 5%	215	(214, 216)	2 (0, 3)
C-SHBSL	Reclassify 10%	215	(214, 216)	2 (0, 3)
C-SHBSL	Reclassify 15%	215	(214, 216)	1 (0, 3)
C-SHBSL	Reclassify 20%	215	(214, 216)	1 (0, 3)
HBS2	Reclassify 5%	214	(214, 215)	1 (-1, 3)
HBS2	Reclassify 10%	214	(213, 215)	1 (-1, 2)
HBS2	Reclassify 15%	214	(213, 215)	1 (-1, 2)
HBS2	Reclassify 20%	214	(213, 215)	1 (0, 2)
HBST	Reclassify 5%	215	(214, 216)	2 (0, 3)
HBST	Reclassify 10%	215	(214, 216)	2 (0, 3)
HBST	Reclassify 15%	215	(214, 216)	2 (0, 3)
HBST	Reclassify 20%	215	(214, 216)	2 (0, 3)
SIRS	Reclassify 5%	216	(215, 217)	2 (1, 3)
SIRS	Reclassify 10%	215	(214, 216)	2 (1, 3)
SIRS	Reclassify 15%	215	(214, 216)	2 (1, 3)
SIRS	Reclassify 20%	215	(214, 216)	2 (1, 3)
qSOFA	Reclassify 5%	221	(220, 221)	7 (6, 8)
qSOFA	Reclassify 10%	220	(220, 221)	7 (6, 8)
qSOFA	Reclassify 15%	220	(219, 220)	6 (5, 7)
qSOFA	Reclassify 20%	219	(219, 220)	6 (5, 7)
NEWS	Reclassify 5%	218	(217, 219)	4 (3, 6)
NEWS	Reclassify 10%	218	(217, 219)	4 (3, 6)
NEWS	Reclassify 15%	217	(217, 218)	4 (3, 5)
NEWS	Reclassify 20%	217	(216, 218)	4 (3, 5)
Base: No Tool ^b	Reclassify 5%	221	(221, 221)	8 (7, 9)
Base: No Tool ^b	Reclassify 10%	221	(221, 221)	7 (7, 8)
Base: No Tool ^b	Reclassify 15%	220	(220, 220)	7 (6, 8)
Base: No Tool ^b	Reclassify 20%	220	(220, 220)	7 (6, 8)

Systemic Inflammatory Response Syndrome (SIRS), quick Sequential Organ Failure Assessment (qSOFA), National Early Warning Score (NEWS) Super HCT Bacterial Sepsis Learner (SHDSL), Clinical Factor-specific Super HCT Bacterial Sepsis Learner (C-SHDSL), HCT Bacteremia Sepsis Score (HBS2), HCT Bacteremia Sepsis Tree (HBST)

^a 2.5-97.5% Percentiles for Mortality and Antibiotic Days

^b Base scenario: simulation run without a prediction tool (sensitivity: 0%, specificity 100%)

eTable 12: Simulated Antibiotic Use For 2000 Allogeneic HCT Recipients (Correct Reclassification)

Prognostic Tool	Reclassification Percentage	Antibiotic Count	Prediction Intervals ^a
<i>All Antibiotic Days</i>			
SHBSL	Reclassify 5%	5929	(5611, 6306)
SHBSL	Reclassify 10%	5687	(5384, 6045)
SHBSL	Reclassify 15%	5444	(5157, 5782)
SHBSL	Reclassify 20%	5200	(4930, 5520)
C-SHBSL	Reclassify 5%	6106	(5760, 6430)
C-SHBSL	Reclassify 10%	5855	(5526, 6162)
C-SHBSL	Reclassify 15%	5603	(5292, 5894)
C-SHBSL	Reclassify 20%	5351	(5057, 5625)
HBS2	Reclassify 5%	7096	(6701, 7478)
HBS2	Reclassify 10%	6795	(6419, 7157)
HBS2	Reclassify 15%	6493	(6137, 6836)
HBS2	Reclassify 20%	6190	(5854, 6513)
HBST	Reclassify 5%	5584	(5244, 5879)
HBST	Reclassify 10%	5359	(5037, 5639)
HBST	Reclassify 15%	5134	(4829, 5399)
HBST	Reclassify 20%	4908	(4620, 5158)
SIRS	Reclassify 5%	5830	(5475, 6202)
SIRS	Reclassify 10%	5592	(5255, 5946)
SIRS	Reclassify 15%	5355	(5036, 5689)
SIRS	Reclassify 20%	5116	(4816, 5432)
qSOFA	Reclassify 5%	953	(889, 1026)
qSOFA	Reclassify 10%	968	(907, 1037)
qSOFA	Reclassify 15%	983	(925, 1048)
qSOFA	Reclassify 20%	998	(944, 1059)
NEWS	Reclassify 5%	4550	(4267, 4887)
NEWS	Reclassify 10%	4378	(4110, 4698)
NEWS	Reclassify 15%	4206	(3952, 4509)
NEWS	Reclassify 20%	4034	(3795, 4319)
Base: No Tool ^b	Reclassify 5%	826	(826, 826)
Base: No Tool ^b	Reclassify 10%	847	(847, 847)
Base: No Tool ^b	Reclassify 15%	869	(869, 869)
Base: No Tool ^b	Reclassify 20%	891	(891, 891)

Prognostic Tool	Reclassification Percentage	Antibiotic Count	Prediction Intervals ^a
<i>Antibiotic Days - Infected Individuals</i>			
SHBSL	Reclassify 5%	1110	(1073, 1154)
SHBSL	Reclassify 10%	1116	(1081, 1158)
SHBSL	Reclassify 15%	1123	(1090, 1163)
SHBSL	Reclassify 20%	1130	(1099, 1167)
C-SHBSL	Reclassify 5%	1041	(989, 1091)
C-SHBSL	Reclassify 10%	1051	(1001, 1099)
C-SHBSL	Reclassify 15%	1061	(1014, 1106)
C-SHBSL	Reclassify 20%	1071	(1027, 1114)
HBS2	Reclassify 5%	1073	(1028, 1106)
HBS2	Reclassify 10%	1081	(1039, 1112)
HBS2	Reclassify 15%	1090	(1049, 1119)
HBS2	Reclassify 20%	1098	(1060, 1126)
HBST	Reclassify 5%	1037	(997, 1078)
HBST	Reclassify 10%	1047	(1010, 1086)
HBST	Reclassify 15%	1057	(1022, 1095)
HBST	Reclassify 20%	1068	(1035, 1103)
SIRS	Reclassify 5%	1031	(980, 1074)
SIRS	Reclassify 10%	1042	(993, 1082)
SIRS	Reclassify 15%	1053	(1006, 1091)
SIRS	Reclassify 20%	1063	(1020, 1099)
qSOFA	Reclassify 5%	855	(833, 879)
qSOFA	Reclassify 10%	875	(855, 897)
qSOFA	Reclassify 15%	895	(876, 916)
qSOFA	Reclassify 20%	915	(897, 935)
NEWS	Reclassify 5%	939	(896, 981)
NEWS	Reclassify 10%	955	(914, 994)
NEWS	Reclassify 15%	970	(932, 1007)
NEWS	Reclassify 20%	986	(949, 1021)
Base: No Tool ^b	Reclassify 5%	826	(826, 826)
Base: No Tool ^b	Reclassify 10%	847	(847, 847)
Base: No Tool ^b	Reclassify 15%	869	(869, 869)
Base: No Tool ^b	Reclassify 20%	891	(891, 891)
<i>Antibiotic Days - Uninfected Individuals</i>			

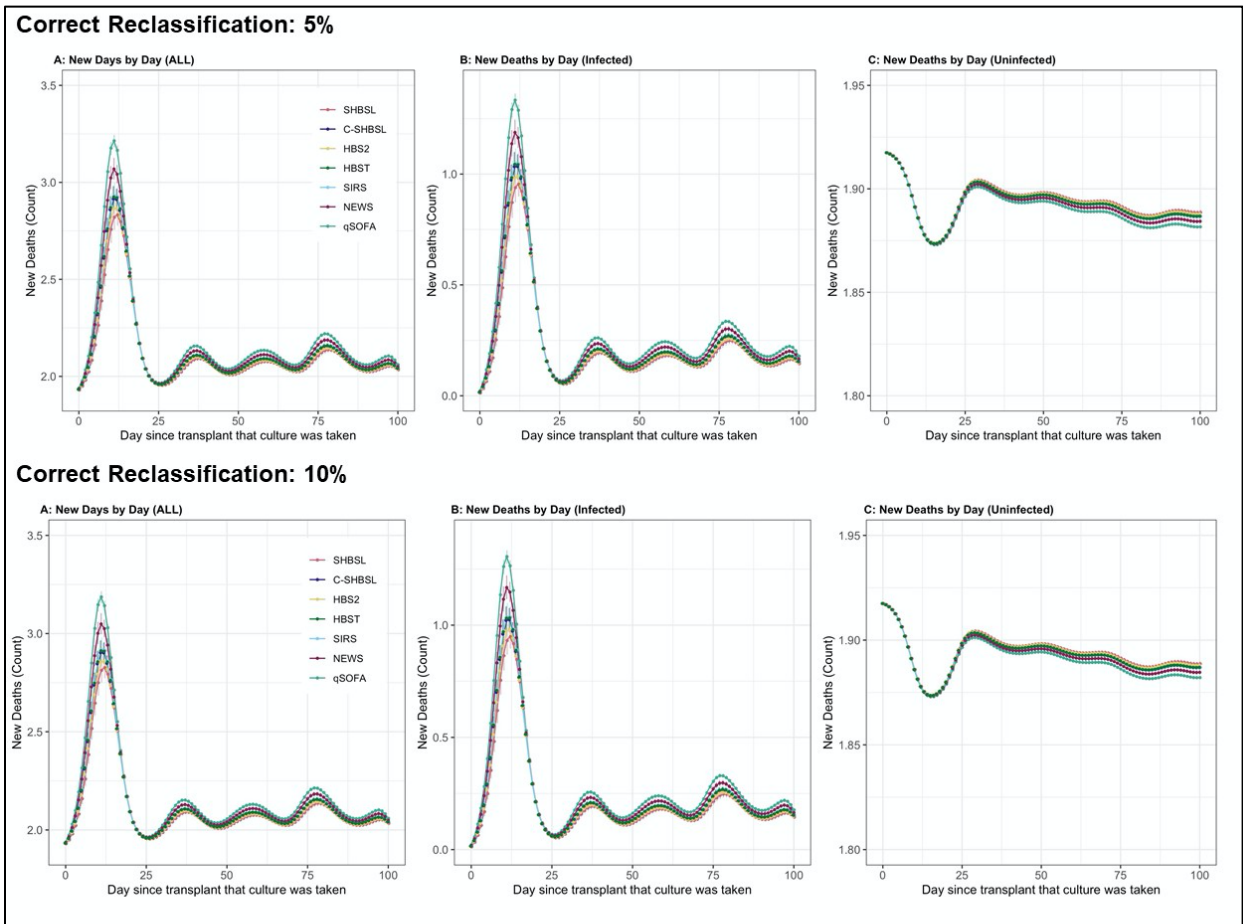
Prognostic Tool	Reclassification Percentage	Antibiotic Count	Prediction Intervals^a
SHBSL	Reclassify 5%	4820	(4479, 5194)
SHBSL	Reclassify 10%	4571	(4247, 4926)
SHBSL	Reclassify 15%	4321	(4015, 4658)
SHBSL	Reclassify 20%	4071	(3782, 4388)
C-SHBSL	Reclassify 5%	5066	(4730, 5425)
C-SHBSL	Reclassify 10%	4804	(4486, 5145)
C-SHBSL	Reclassify 15%	4542	(4241, 4865)
C-SHBSL	Reclassify 20%	4280	(3995, 4584)
HBS2	Reclassify 5%	6024	(5627, 6402)
HBS2	Reclassify 10%	5714	(5337, 6073)
HBS2	Reclassify 15%	5403	(5046, 5743)
HBS2	Reclassify 20%	5092	(4755, 5412)
HBST	Reclassify 5%	4547	(4208, 4875)
HBST	Reclassify 10%	4312	(3990, 4624)
HBST	Reclassify 15%	4076	(3771, 4371)
HBST	Reclassify 20%	3840	(3553, 4119)
SIRS	Reclassify 5%	4798	(4446, 5165)
SIRS	Reclassify 10%	4550	(4216, 4899)
SIRS	Reclassify 15%	4302	(3985, 4632)
SIRS	Reclassify 20%	4053	(3755, 4364)
qSOFA	Reclassify 5%	98	(45, 160)
qSOFA	Reclassify 10%	93	(42, 152)
qSOFA	Reclassify 15%	88	(40, 144)
qSOFA	Reclassify 20%	83	(38, 135)
NEWS	Reclassify 5%	3611	(3351, 3902)
NEWS	Reclassify 10%	3424	(3177, 3699)
NEWS	Reclassify 15%	3236	(3003, 3497)
NEWS	Reclassify 20%	3048	(2828, 3294)
Base: No Tool ^b	Reclassify 5%	0	(0, 0)
Base: No Tool ^b	Reclassify 10%	0	(0, 0)
Base: No Tool ^b	Reclassify 15%	0	(0, 0)
Base: No Tool ^b	Reclassify 20%	0	(0, 0)

Systemic Inflammatory Response Syndrome (SIRS), quick Sequential Organ Failure Assessment (qSOFA), National Early Warning Score (NEWS) Super HCT Bacterial Sepsis Learner (SHDSL), Clinical Factor-specific Super HCT Bacterial Sepsis Learner (C-SHDSL), HCT Bacteremia Sepsis Score (HBS2), HCT Bacteremia Sepsis Tree (HBST)

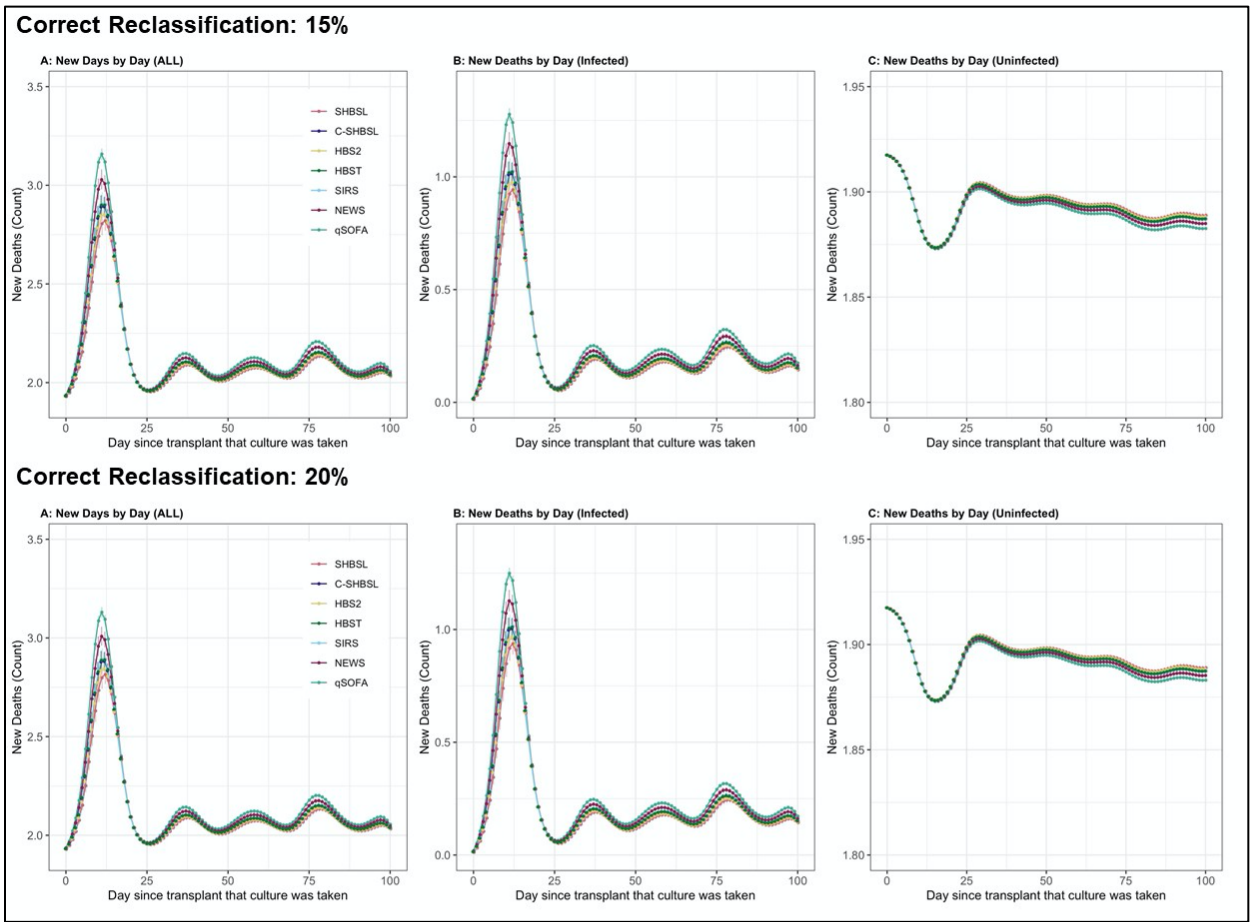
^a 2.5-97.5% Percentiles for Mortality and Antibiotic Days

^b Base scenario: simulation run without a prediction tool (sensitivity: 0%, specificity 100%)

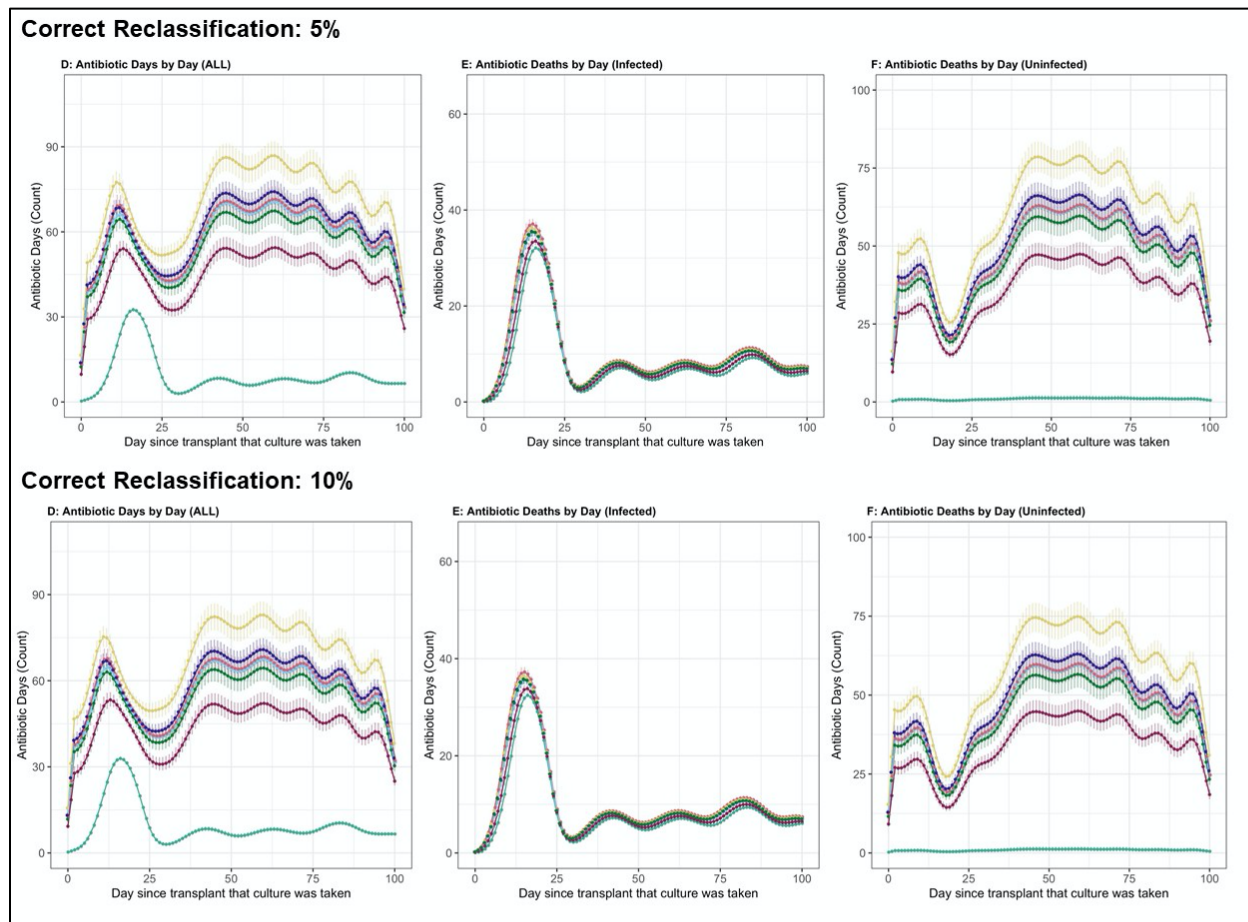
e.Figure 13: Mortality from Semi-Stochastic equation-based Model Results for Patients with Potential Bloodstream Infections (Primary Analysis Cohort) with Correct Reclassification



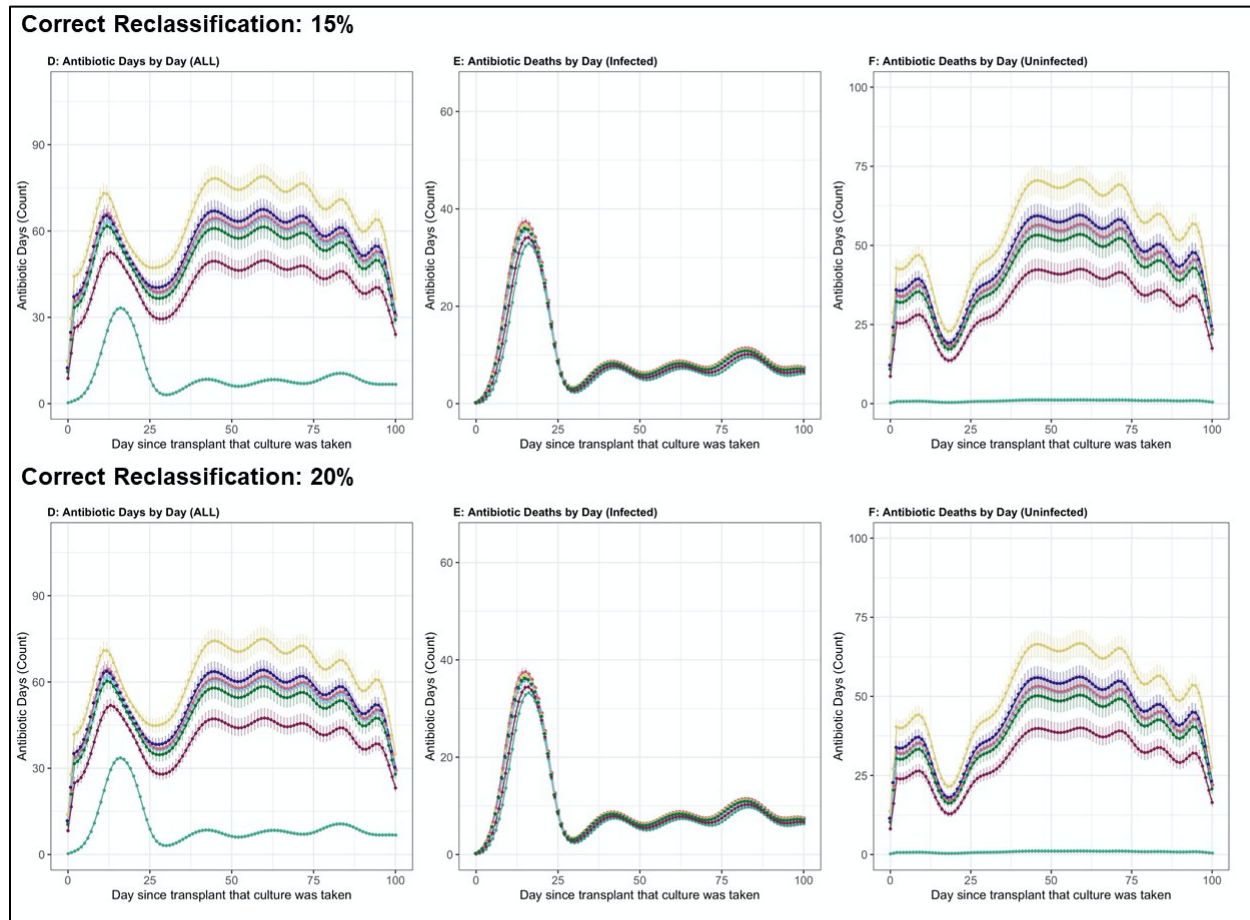
e.Figure 13 (Continued): Mortality from Semi-Stochastic equation-based Model Results for Patients with Potential Bloodstream Infections (Primary Analysis Cohort) with Correct Reclassification



e.Figure 14: Antibiotic Use from Semi-Stochastic equation-based Model Results for Patients with Potential Bloodstream Infections (Primary Analysis Cohort) with Correct Reclassification



e.Figure 14 (Continued): Antibiotic Use from Semi-Stochastic equation-based Model Results for Patients with Potential Bloodstream Infections (Primary Analysis Cohort) with Correct Reclassification



Size of Association Between Treatment and Mortality

eTable 13: Simulated Mortality For 2000 Allogeneic HCT Recipients (Mortality – Treatment Association)

Prognostic Tool	Odds Ratio	Mortality Count	Prediction Intervals ^a	Mortality Difference
SHBSL	OR 1	218	(218, 218)	0 (0, 0)
SHBSL	OR 1.03	223	(221, 224)	0 (0, 0)
SHBSL	OR 1.05	229	(224, 232)	0 (0, 0)
C-SHBSL	OR 1	218	(218, 218)	0 (0, 0)
C-SHBSL	OR 1.03	225	(223, 227)	3 (1, 5)
C-SHBSL	OR 1.05	235	(231, 240)	7 (2, 14)
HBS2	OR 1	218	(218, 218)	0 (0, 0)
HBS2	OR 1.03	224	(222, 225)	1 (-1, 4)
HBS2	OR 1.05	232	(229, 236)	3 (-2, 10)
HBST	OR 1	218	(218, 218)	0 (0, 0)
HBST	OR 1.03	225	(224, 227)	3 (1, 5)
HBST	OR 1.05	236	(232, 240)	7 (2, 13)
SIRS	OR 1	218	(218, 218)	0 (0, 0)
SIRS	OR 1.03	226	(224, 227)	3 (1, 5)
SIRS	OR 1.05	236	(233, 241)	8 (3, 13)
qSOFA	OR 1	219	(219, 219)	0 (0, 0)
qSOFA	OR 1.03	234	(233, 234)	11 (9, 13)
qSOFA	OR 1.05	256	(254, 258)	27 (23, 32)
NEWS	OR 1	218	(218, 218)	0 (0, 0)
NEWS	OR 1.03	229	(228, 231)	7 (5, 9)
NEWS	OR 1.05	246	(242, 250)	17 (12, 23)
Base: No Tool ^b	OR 1	219	(219, 219)	0 (0, 0)
Base: No Tool ^b	OR 1.03	235	(235, 235)	12 (11, 14)
Base: No Tool ^b	OR 1.05	259	(259, 259)	30 (26, 34)

Systemic Inflammatory Response Syndrome (SIRS), quick Sequential Organ Failure Assessment (qSOFA), National Early Warning Score (NEWS) Super HCT Bacterial Sepsis Learner (SHDSL), Clinical Factor-specific Super HCT Bacterial Sepsis Learner (C-SHDSL), HCT Bacteremia Sepsis Score (HBS2), HCT Bacteremia Sepsis Tree (HBST)

^a 2.5-97.5% Percentiles for Mortality and Antibiotic Days

^b Base scenario: simulation run without a prediction tool (sensitivity: 0%, specificity 100%)

eTable 14: Simulated Antibiotic Use For 2000 Allogeneic HCT Recipients (Mortality – Treatment Association)

Prognostic Tool	Odds Ratio	Antibiotic Count	Prediction Intervals ^a
<i>All Antibiotic Days</i>			
SHBSL	OR 1	7213	(6843, 7649)
SHBSL	OR 1.03	7159	(6793, 7596)
SHBSL	OR 1.05	7076	(6711, 7515)
C-SHBSL	OR 1	7428	(7028, 7807)
C-SHBSL	OR 1.03	7338	(6933, 7700)
C-SHBSL	OR 1.05	7203	(6776, 7537)
HBS2	OR 1	8567	(8110, 9005)
HBS2	OR 1.03	8495	(8038, 8944)
HBS2	OR 1.05	8385	(7929, 8852)
HBST	OR 1	6823	(6431, 7165)
HBST	OR 1.03	6732	(6333, 7071)
HBST	OR 1.05	6593	(6177, 6934)
SIRS	OR 1	7109	(6698, 7539)
SIRS	OR 1.03	7015	(6603, 7447)
SIRS	OR 1.05	6872	(6471, 7308)
qSOFA	OR 1	1466	(1391, 1548)
qSOFA	OR 1.03	1292	(1214, 1384)
qSOFA	OR 1.05	1031	(938, 1128)
NEWS	OR 1	5639	(5315, 6021)
NEWS	OR 1.03	5499	(5159, 5914)
NEWS	OR 1.05	5288	(4949, 5723)
Base: No Tool ^b	OR 1	1320	(1320, 1320)
Base: No Tool ^b	OR 1.03	1133	(1133, 1133)
Base: No Tool ^b	OR 1.05	853	(853, 853)
<i>Antibiotic Days - Infected Individuals</i>			
SHBSL	OR 1	1608	(1570, 1654)
SHBSL	OR 1.03	1560	(1504, 1626)
SHBSL	OR 1.05	1486	(1405, 1584)
C-SHBSL	OR 1	1535	(1480, 1588)
C-SHBSL	OR 1.03	1455	(1377, 1531)
C-SHBSL	OR 1.05	1335	(1221, 1446)
HBS2	OR 1	1566	(1518, 1601)
HBS2	OR 1.03	1503	(1435, 1553)
HBS2	OR 1.05	1407	(1310, 1480)
HBST	OR 1	1532	(1490, 1576)
HBST	OR 1.03	1449	(1390, 1512)
HBST	OR 1.05	1324.7	(1238.7, 1415.8)

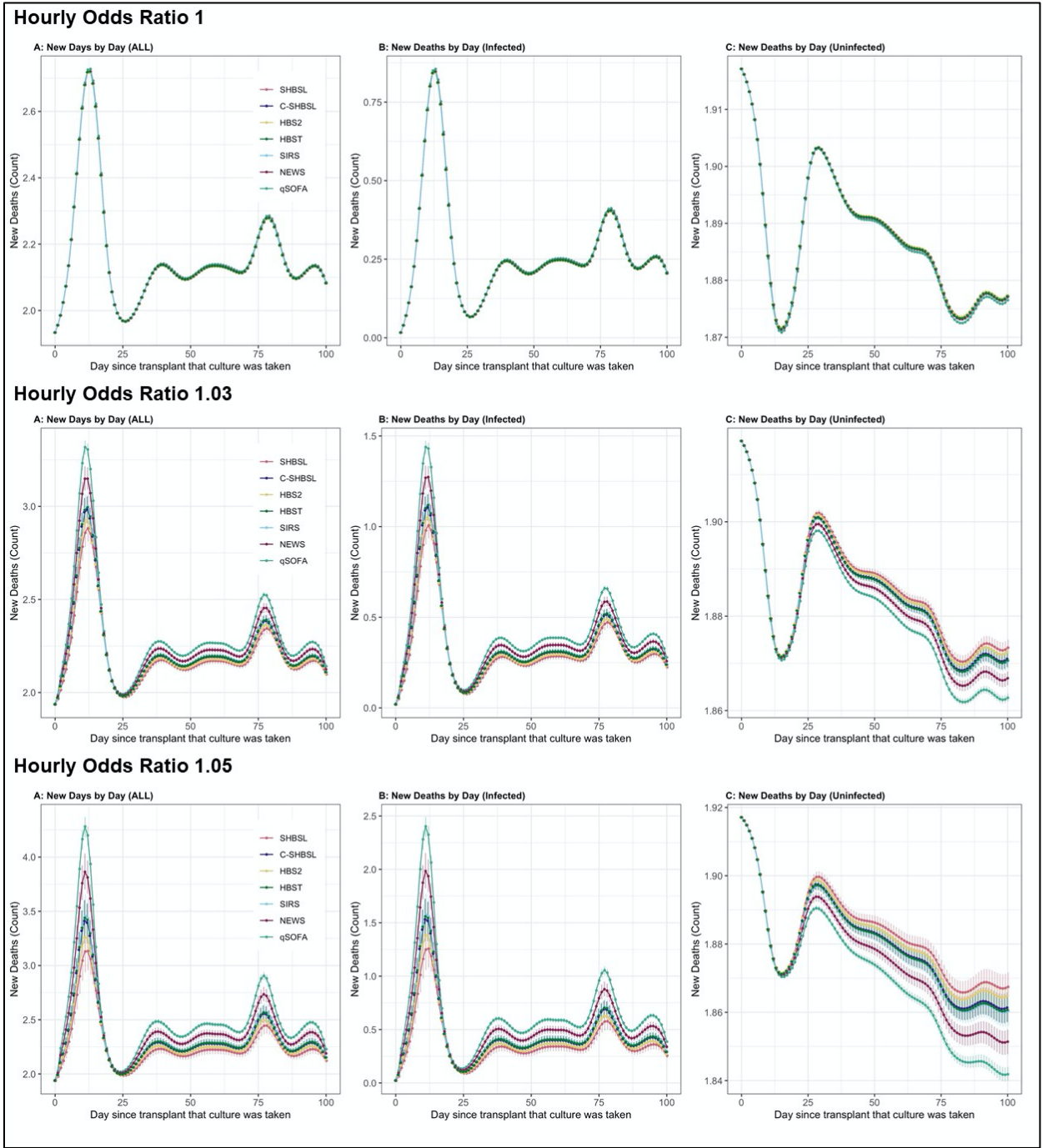
Prognostic Tool	Odds Ratio	Antibiotic Count	Prediction Intervals ^a
SIRS	OR 1	1526	(1471, 1570)
SIRS	OR 1.03	1442	(1363, 1505)
SIRS	OR 1.05	1314	(1200, 1406)
qSOFA	OR 1	1351	(1328, 1376)
qSOFA	OR 1.03	1178	(1145, 1213)
qSOFA	OR 1.05	917	(870, 968)
NEWS	OR 1	1431	(1386, 1475)
NEWS	OR 1.03	1303	(1238, 1365)
NEWS	OR 1.05	1109	(1016, 1200)
Base: No Tool ^b	OR 1	1320	(1320, 1320)
Base: No Tool ^b	OR 1.03	1133	(1133, 1133)
Base: No Tool ^b	OR 1.05	853	(853, 853)
<i>Antibiotic Days - Uninfected Individuals</i>			
SHBSL	OR 1	5605	(5209, 6039)
SHBSL	OR 1.03	5599	(5205, 6033)
SHBSL	OR 1.05	5590	(5199, 6024)
C-SHBSL	OR 1	5893	(5504, 6311)
C-SHBSL	OR 1.03	5883	(5494, 6298)
C-SHBSL	OR 1.05	5868	(5479, 6279)
HBS2	OR 1	7001	(6541, 7439)
HBS2	OR 1.03	6992	(6532, 7429)
HBS2	OR 1.05	6978	(6519, 7414)
HBST	OR 1	5292	(4898, 5675)
HBST	OR 1.03	5282	(4889, 5662)
HBST	OR 1.05	5268	(4877, 5641)
SIRS	OR 1	5583	(5174, 6009)
SIRS	OR 1.03	5573	(5165, 5998)
SIRS	OR 1.05	5558	(5151, 5982)
qSOFA	OR 1	115	(52, 188)
qSOFA	OR 1.03	115	(52, 187)
qSOFA	OR 1.05	114	(52, 186)
NEWS	OR 1	4207	(3905, 4545)
NEWS	OR 1.03	4196	(3895, 4533)
NEWS	OR 1.05	4179	(3880, 4516)
Base: No Tool ^b	OR 1	0	(0, 0)
Base: No Tool ^b	OR 1.03	0	(0, 0)
Base: No Tool ^b	OR 1.05	0	(0, 0)

Systemic Inflammatory Response Syndrome (SIRS), quick Sequential Organ Failure Assessment (qSOFA), National Early Warning Score (NEWS) Super HCT Bacterial Sepsis Learner (SHDSL), Clinical Factor-specific Super HCT Bacterial Sepsis Learner (C-SHDSL), HCT Bacteremia Sepsis Score (HBS2), HCT Bacteremia Sepsis Tree (HBST)

^a 2.5-97.5% Percentiles for Mortality and Antibiotic Days

^b Base scenario: simulation run without a prediction tool (sensitivity: 0%, specificity 100%)

e.Figure 15: Mortality from Semi-Stochastic equation-based Model Results for Patients with Potential Bloodstream Infections (Primary Analysis Cohort) with Varying Odds Ratios



e.Figure 16: Antibiotic Use from Semi-Stochastic equation-based Model Results for Patients with Potential Bloodstream Infections (Primary Analysis Cohort) with Varying Odds Ratios

