OPEN

# Predicting Time to Death After Withdrawal of Life-Sustaining Measures Using Vital Sign Variability: Derivation and Validation

Nathan B. Scales, PhD[1]
Christophe L. Herry, PhD[1]
Amanda van Beinum, PhD[2]
Melanie L. Hogue, MSc[2]
Laura Hornby, MSc[2,3]
Jason Shahin, MD[4,5]
Sonny Dhanani, MD[2,6]
Andrew J. E. Seely, MD, PhD[1,7]

**OBJECTIVES:** To develop a predictive model using vital sign (heart rate and arterial blood pressure) variability to predict time to death after withdrawal of life-supporting measures.

**DESIGN:** Retrospective analysis of observational data prospectively collected as part of the Death Prediction and Physiology after Removal of Therapy study between May 1, 2014, and May 1, 2018.

**SETTING:** Adult ICU.

**PATIENTS:** Adult patients in the ICU with a planned withdrawal of life-supporting measures and an expectation of imminent death.

**INTERVENTIONS:** None.

**MEASUREMENTS AND MAIN RESULTS:** Vital sign waveforms and clinical data were prospectively collected from 429 patients enrolled from 20 ICUs across Canada, the Czech Republic, and the Netherlands. Vital sign variability metrics were calculated during the hour prior to withdrawal. Patients were randomly assigned to the derivation cohort (288 patients) or the validation cohort (141 patients), of which 103 and 54, respectively, were eligible for organ donation after circulatory death. Random survival forest models were developed to predict the probability of death within 30, 60, and 120 minutes following withdrawal using variability metrics, features from existing clinical models, and/or the physician's prediction of rapid death. A model employing variability metrics alone performed similarly to a model employing clinical features, whereas the combination of variability, clinical features, and physician's prediction achieved the highest area under the receiver operating characteristics curve of all models at 0.78 (0.7–0.86), 0.79 (0.71–0.87), and 0.8 (0.72–0.88) for 30-, 60- and 120-minute predictions, respectively.

**CONCLUSIONS:** Machine learning models of vital sign variability data before withdrawal of life-sustaining measures, combined with clinical features and the physician's prediction, are useful to predict time to death. The impact of providing this information for decision support for organ donation merits further investigation.

**KEY WORDS:** clinical decision support systems; donor selection; machine learning; organ donation; organ transplantation; vital signs

Circulatory arrest following the withdrawal of life-sustaining measures (WLSM) is the most common mode of death in patients with a dismal prognosis and irrecoverable injury (1), and organ donation after circulatory death (DCD) has the potential to vastly improve the number of deceased organ donations (2). To minimize the potential for ischemic damage of recovered organs, transplant teams impose strict limits on the allowable time between WLSM and the determination of death, after which the transplant



**DOI: 10.1097/CCE.0000000000000675**

team would "stand down." These limits are specified by organ type, for example, 20–30 minutes for liver, 60 minutes for lungs and pancreas, and 1.5–2 hours for kidneys (3), and vary by jurisdiction (4–6).

Predicting which patients will die within these timelines is often difficult (3, 7), and this uncertainty causes anxiety among healthcare professionals (8, 9) and distress among family members (10–14). The fraction of patients dying within these limits varies by jurisdiction, ranging from 57% to 88% for 2 hours (4, 6, 15–17). Failed donation can lead to disappointment for families (11, 12) and be costly for the medical system due to unnecessary reservation of operating rooms and transportation of transplant teams (18, 19).

The clinician's prediction of time to death has had mixed success (20–22). Several authors have investigated potential predictors of a rapid time to death (7, 15, 23). Most prediction models only considered death within 1 hour of WLSM, a few have been validated (6, 7, 20, 24, 25), and none have seen widespread deployment. Some models, such as the University of Wisconsin tool (26), require a brief trial with reduced ventilation support, which is controversial in some regions and may be a barrier to implementation (20, 22, 27).

Continuous monitoring of cardiovascular variability can enable earlier detection of critical illness (28, 29). Cardiovascular variability includes heart rate and blood pressure variability, which quantifies the fluctuations of beat-by-beat measurements over time. As healthy biological systems possess innate variability, a decrease in variability can be indicative of a stressed system. A variety of variability metrics have been used to predict clinical outcomes such as multiple organ dysfunction syndrome (30), sepsis (31, 32), extubation failure (33), and mortality (34, 35).

As potential DCD patients generally have catastrophic brain injuries and/or severe illness, we hypothesized that vital sign variability could predict rapidity of death after WLSM. We performed a secondary analysis of prospectively collected waveform data from the Death Prediction and Physiology after the Removal of Therapy (DePPaRT) study (36) to derive and validate a random survival forest (RSF) model using heart rate and blood pressure variability to predict the probability of dying within 30, 60, and 120 minutes after WLSM.

## MATERIALS AND METHODS

### Inclusion/Exclusion

This study was an analysis of data collected prospectively from patients enrolled in the DePPaRT study (36), from 20 adult ICUs in Canada (16 sites), the Czech Republic (3 sites), and the Netherlands (1 site) between May 1, 2014, and May 1, 2018. Patients were eligible if there was a planned WLSM with the expectation of imminent death, had at least 15 minutes of high-quality electrocardiogram (ECG) and arterial blood pressure waveform data prior to WLSM, and had a confirmed time of death. Both DCD eligible and noneligible patients were included. Patients were excluded if they had an active pacemaker, a neurologic determination of death, or if WLSM was staggered (i.e., life-sustaining measures modified greater than 10 minutes after the first act of WLSM [37, 38]). The research protocol was approved by the relevant institutional review board or ethics committee at each site (coordinating site: Children's Hospital of Eastern Ontario Research Ethics Board, No. 14/08E, see **Supplemental Information**, http://links.lww.com/CCX/A964, for full list). All patients' surrogate decision makers provided written informed consent for participation in the study. Consent for organ donation was obtained independently from this study, according to local practice.

### Data Collection

The time of death was confirmed through waveform data adjudicated using custom software (36) or by chart data for successful DCD patients without waveform data after withdrawal. Basic demographic information, clinical characteristics, and details of WLSM were collected. Vital sign waveform data were captured from bedside monitors beginning from up to 1 hour prior to WLSM until up to 30 minutes after the declaration of death. Capture used a variety of methods that depended on the type of bedside monitor and central stations. Only waveform data collected prior to WLSM was used in this analysis.

### Model Features

We randomly assigned 2/3 of patients to the derivation cohort, and 1/3 of patients to the validation cohort, ensuring that the fraction of patients from each

country remained the same. To assess the performance of variability features relative to traditional features, we developed four different prediction models: 1) employing waveform variability features only, 2) clinical features only, 3) the physician's prediction of time to death alone, and 4) all features combined.

***Waveform Variability Features***. The ECG and arterial blood pressure waveform data from 60 to 5 minutes prior to WLSM were processed to obtain beat-to-beat event times series (R peak interval, systolic, diastolic, mean, and pulse blood pressures), using previously reported software (31, 33, 36, 39). A comprehensive suite of 14 variability measures for each blood pressure event time series and 15 for the R peak interval series was calculated for each patient using custom software (39), for a total of 71 variability measures (**Supplemental Table E1**, http://links.lww.com/CCX/A964). We used the average of each variability measure in the hour prior to WLSM as input features for our predictive model. In 15 DCD attempts, measurements were taken in the hour prior to transport out of the ICU. For more details, see Supplemental Information (http://links.lww.com/CCX/A964).

***Clinical Features***. Six clinical variables (Glasgow Coma Scale [GCS], positive end-expiratory pressure, pH, systolic blood pressure, spontaneous respiration rate, and use of analgesia) identified by previous studies (20, 22) were used as clinical features, taken from the patient's chart at the time of WLSM.

***Physician Prediction of Time to Death***. Prior to WLSM for each patient, the most responsible physician was asked if they expected the time to death to be within 1, 2, 3, or 6 hours after WLSM and to rate the confidence of their prediction (low, moderate, and high). A binary prediction of "timely death" was defined as a physician's prediction of death within 1 hour of WLSM with moderate or high confidence, and was assessed as a predictor for outcomes at 30, 60, and 120 minutes after WLSM.

## Predictive Model

***Model Development***. We used RSF (40) to develop our predictive models for all models except the model employing the physician's prediction alone, using the "ranger" (41) package in R Foundation for Statistical Computing (Vienna, Austria) (42) (see Supplemental Information, http://links.lww.com/CCX/A964). RSF employ an ensemble of decision trees, with patients and features randomly allocated per tree, providing a form of internal validation. RSFs are capable of handling a large number of features without overfitting (43, 44), fully nonparametric, and capable of detecting nonlinear effects and multiple interactions in predictors, while permitting correlation between predictors (40).

Each survival model yielded a probability of survival S(t) for each 15-minute interval up until 24 hours after WLSM, for each patient (**Supplemental Fig. E1**, http://links.lww.com/CCX/A964). We calculated the probability of dying at each time as $D(t) = 1 - S(t)$. The probability of dying at a specific time ($t$ = 30, 60, or 120 min) was then used as a score to predict if a patient would die within this time. Model performance at each time was assessed using the area under the receiver operating characteristic curve (ROC AUC, or AUC).

We optimized the set of available features used in each model by ranking features based on their importance values (43, 45), using a 10-fold cross validation of the derivation set, and employed a feature removal step to remove redundant features (see Supplemental Information, http://links.lww.com/CCX/A964). Finally, for each feature set, we combined all patients from the derivation cohort and the optimized features in a final model, which we tested on the validation set.

***Model Validation***. Each derived model (employing variability features, clinical features, physician prediction of timely death, or all features combined) was tested on the validation cohort. We used 5,000 bootstrap iterations to calculate CIs in the validation cohort.

***Sensitivity Analysis***. We performed a sensitivity analysis for the AUC using the probability scores for each model by combining all patients from both the derivation and the validation cohorts and separating them by DCD eligibility status or by country of origin.

***Model Calibration***. Model calibration was assessed at the population level by comparing the occurrence rate of rapid death with the mean predicted probabilities at each time for all patients. Calibration curves were generated by calculating the fraction of patients assigned a probability of dying that died within the given time frame, grouping the probabilities into deciles over 5,000 bootstrap iterations.

A recalibration step was included in the model pipeline to ensure probability estimates were accurate. For each cross-validation fold of the combined model for the derivation cohort, model probabilities were

recalibrated using equal mass bins (deciles) (46, 47). The calibration histogram learned from the derivation cohort was used to adjust the model probabilities of the validation cohort. See Supplemental Information (http://links.lww.com/CCX/A964) for more details.

## RESULTS

Out of 654 patients enrolled in the DePPaRT study (36), 429 were included in our secondary analysis (**Fig. 1**). We randomly assigned 288 patients (2/3) to the derivation cohort, and 141 patients (1/3) to the validation cohort, keeping the fraction of patients from each country the same (Canada [48%], the Czech Republic [44%], and the Netherlands [8%]). The median time to death was 56 and 54 minutes in the derivation and validation cohorts, respectively (**Table 1**). In the derivation and validation cohorts, respectively, 103 (36%) and 54 (38%) were considered DCD eligible by their organ donor organizations, 47 (16%) and 26 (18%) had DCD attempted, and 30 (10%) and 18 (13%) became successful organ donors.

### Model Derivation

We compare the performance of the derivation set (using 10-fold cross-validation) for the four different models in **Supplemental Table E2** (http://links.lww.com/CCX/A964). The waveform variability model performed similarly to the clinical model, although many variability features showed a higher Spearman correlation with time to death than the clinical features (**Supplemental Table E3**, http://links.lww.com/CCX/A964). The physician's prediction of timely death achieved slightly higher AUCs than the variability or clinical models. The combined model achieved the highest AUC scores at all three times, with AUC values of 0.79, 0.81, and 0.84 for death within 30, 60, and 120 minutes, respectively. During feature optimization, all clinical features were removed from the combined model, with only variability features and the physician's prediction of timely death remaining in the final model (Supplemental Information, http://links.lww.com/CCX/A964).

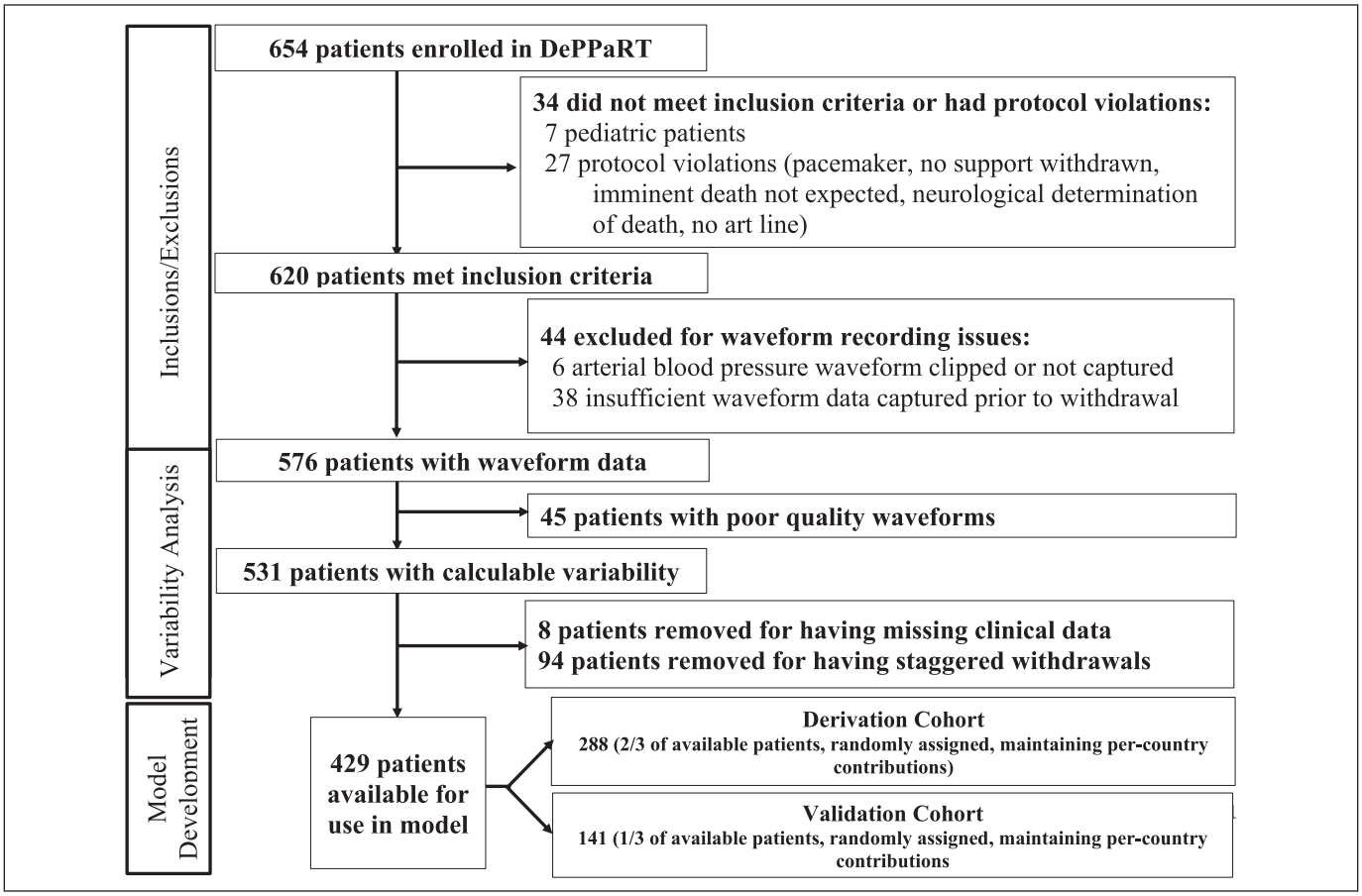


**Figure 1.** Flow diagram outlining inclusion/exclusion criteria for model development. DePPaRT = Death Prediction and Physiology after the Removal of Therapy.

## TABLE 1.

**Characteristics of Enrolled Patients in the Derivation and Validation Cohorts, As Well As Patients Excluded From Analysis.**

| Demographic Characteristics | Derivation Cohort ($n = 288$) | Validation Cohort ($n = 141$) | Excluded Patients ($n = 225$[a]) |
|---|---|---|---|
| Average age (SD, range) | 63 (16, 18–94) | 63 (15, 22–95) | 65 (15, 24–93) |
| Female gender | 114 (40%) | 55 (39%) | 75 (35%) |
| Chronic condition | 230 (80%) | 113 (80%) | 187 (87%) |
| Primary reason for ICU admission | | | |
| Neurologic | 152 (53%) | 75 (53%) | 83 (38%) |
| Cardiac | 9 (3%) | 7 (5%) | 11 (5%) |
| Respiratory | 37 (13%) | 20 (14%) | 43 (20%) |
| Sepsis/infection | 42 (15%) | 16 (11%) | 38 (18%) |
| Trauma | 13 (5%) | 6 (4%) | 8 (4%) |
| Other[b] | 35 (12%) | 17 (12%) | 33 (15%) |
| Median Glasgow Coma Scale at ICU admission (IQR) | 4 (3–9) | 3 (3–10) | 7 (3–14) |
| Avg. Acute Physiology and Chronic Health Evaluation II Score (IQR, range) | 28 (23–33, 7–55), $n = 286$ | 27 (21–34, 6–47) | 27 (21–32, 8–48) $n = 214$ |
| Reported traumatic brain injury | 31 (11%) | 10 (7%) | 21 (10%) |
| Median length of stay in ICU, d (range) | 4 (0–32) | 4 (0–34), $n = 140$ | 3 (0–61) |
| DCD organ donation organization eligible | 103 (36%) | 54 (38%) | 49 (23%) |
| DCD attempted | 47 (16%) | 26 (18%) | 14 (6%) |
| DCD successful donors | 30 (10%) | 18 (13%) | 12 (6%) |
| Median time to death, min (range) | 56 (4 min to 270 hr) | 54 (7 min to 110 hr) | 66 (0 min to 159 hr) $n = 214$[c] |
| Death within 30, 60, and 120 min | 39%, 51%, and 63% | 38%, 51%, and 65% | 29%, 47%, and 61% |
| Patients from Canada | 138 (48%) | 68 (48%) | 163 (75%) |
| Patients from the Czech Republic | 127 (44%) | 62 (44%) | 44 (20%) |
| Patients from the Netherlands | 23 (8%) | 11 (8%) | 9 (4%) |
| Life-Sustaining Measures | | | |
| Receiving invasive mechanical ventilation, with a mandatory ventilation mode | 253 (88%) | 126 (89%) | 184 (85%) |
| Extubated as part of withdrawal of life-sustaining measures | 184 (64%) | 91 (65%) | 119 (55%) |
| Vasopressors/inotropes: 0 | 130 (45%) | 76 (54%) | 68 (31%) |
| Vasopressors/inotropes: 1 | 111 (39%) | 49 (35%) | 76 (35%) |
| Vasopressors/inotropes: 2 | 32 (11%) | 12 (9%) | 38 (18%) |
| Vasopressors/inotropes: ≥ 3 | 15 (5%) | 4 (3%) | 34 (16%) |
| Receiving sedation | 194 (67%) | 107 (76%) | 172 (80%) |
| Receiving analgesia | 267 (93%) | 130 (92%) | 202 (94%) |

DCD = donation after circulatory death, IQR = interquartile range.

[b]Other reasons for admission include gastrointestinal bleeding, abdominal aortic aneurysm, multiple causes, hypovolemic shock, and multiorgan failure.

[a]Data from seven pediatric patients and two patients with missing data were not included, so values are reported for 216/225 unless indicated otherwise. See **Supplementary Table E6** (http://links.lww.com/CCX/A964) for demographics of the excluded population.

[c]Only 178 excluded patients had adjudicated times of death, so the clinically reported time of death was analyzed for this group instead.

Glasgow Coma Scale scores range from 3 to 15, with lower scores indicating a reduced level of consciousness. Scores on the Acute Physiology and Chronic Health Evaluation range from 0 to 71, with higher scores indicating more severe disease. The number of available data points (*n*) is indicated for each field with missing values.

### Model Validation

In the validation cohort, the variability model achieved higher AUCs than the clinical model at all times (AUCs of 0.7–0.72 for variability vs 0.63–0.68 for the clinical model) (**Table 2**), whereas the physician's prediction ranged from 0.72 to 0.76. The combined model achieved the highest AUC scores, with values of 0.78, 0.79, and 0.80 for the 30-, 60-, and 120-minute predictions, respectively.

### Sensitivity Analysis

As the model's primary use case is for DCD eligible patients, we performed a sensitivity analysis of DCD eligible versus DCD ineligible patients (**Supplemental Tables E4** and **E5** http://links.lww.com/CCX/A964). The DCD eligible population was younger, less likely to have preexisting conditions or require vasopressors/inotropes, and had lower GCS and Acute Physiology and Chronic Health Evaluation II scores, while more likely to

## TABLE 2.
**Model Performance in the Validation Cohort, for Predictions of Time to Death Within 30 min, 1 hr, or 2 hr From Withdrawal of Life-Sustaining Measures**

| | | Validation Cohort | | | | |
|---|---|---|---|---|---|---|
| Prediction | Predictors | Sensitivity (95% CI) | Specificity (95% CI) | Positive Predictive Value (95% CI) | Negative Predictive Value (95% CI) | Area Under the Receiver Operating Characteristic Curve (95% CI) |
| 30 min | Waveform variability alone | 0.66 (0.43–0.96) | 0.74 (0.39–0.9) | 0.6 (0.47–0.76) | 0.79 (0.72–0.94) | 0.72 (0.64–0.81) |
| | Brieva clinical features (20, 22) | 0.67 (0.42–0.96) | 0.67 (0.3–0.89) | 0.56 (0.46–0.72) | 0.77 (0.69–0.95) | 0.67 (0.58–0.76) |
| | Physician prediction of timely death | 0.73 (0.61–0.82) | 0.79 (0.69–0.87) | 0.71 (0.63–0.8) | 0.8 (0.73–0.87) | 0.76 (0.69–0.83) |
| | Variability, clinical features, and physician prediction | 0.86 (0.67–0.98) | 0.66 (0.45–0.8) | 0.58 (0.49–0.69) | 0.9 (0.8–0.98) | 0.78 (0.7–0.86) |
| 1 hr | Waveform variability alone | 0.67 (0.49–0.86) | 0.75 (0.51–0.9) | 0.73 (0.63–0.84) | 0.69 (0.6–0.81) | 0.72 (0.64–0.8) |
| | Brieva clinical features | 0.75 (0.52–0.96) | 0.61 (0.32–0.83) | 0.66 (0.59–0.77) | 0.7 (0.6–0.89) | 0.68 (0.59–0.77) |
| | Physician prediction of timely death | 0.62 (0.51–0.72) | 0.82 (0.72–0.91) | 0.81 (0.73–0.9) | 0.63 (0.57–0.71) | 0.72 (0.65–0.79) |
| | Variability, clinical features, and physician prediction | 0.87 (0.71–0.97) | 0.7 (0.53–0.85) | 0.75 (0.67–0.84) | 0.83 (0.71–0.95) | 0.79 (0.71–0.87) |
| 2 hr | Waveform variability alone | 0.68 (0.36–0.82) | 0.72 (0.52–0.96) | 0.82 (0.74–0.95) | 0.54 (0.44–0.67) | 0.7 (0.61–0.79) |
| | Brieva clinical features | 0.68 (0.53–0.89) | 0.65 (0.4–0.82) | 0.76 (0.68–0.85) | 0.57 (0.48–0.73) | 0.63 (0.54–0.73) |
| | Physician prediction of timely death | 0.55 (0.46–0.64) | 0.9 (0.8–0.97) | 0.94 (0.87–0.99) | 0.43 (0.38–0.49) | 0.73 (0.66–0.79) |
| | Variability, clinical features, and physician prediction | 0.78 (0.64–0.91) | 0.77 (0.58–0.9) | 0.86 (0.79–0.93) | 0.66 (0.55–0.81) | 0.8 (0.72–0.88) |

Death within the given time limit was defined as positive, and death outside the time interval was defined as negative. The median value over 5,000 bootstrap iterations is shown (95% confidence interval in brackets).

have traumatic brain injury, invasive mechanical ventilation, or be extubated as part of the WLSM. Despite these differences, DCD eligible and ineligible patients had similar times to death. We assessed the performance of our predictive models in DCD eligible patients only and found the combined model performed well at all three times (AUCs of 0.83, 0.80, and 0.81; see Supplemental Table E5, http://links.lww.com/CCX/A964).

We also assessed model performance by country and found that the combined model achieved higher AUC values in the Czech Republic compared with Canada at 60 and 120 minutes (0.85 and 0.84 vs 0.80 and 0.79). Of note, the physician's prediction of timely death performed poorly in Canada (16 sites, AUCs of 0.67–0.71) compared with the Czech Republic (three sites, AUCs of 0.81–0.84), with no overlap in the CIs.

### Model Calibration

The mean predicted probabilities at each time, calculated using the raw RSF probabilities, were underconfident compared with the empirical occurrence rate of rapid death, and a calibration step was included in the model pipeline. After recalibration, the mean predicted probability of the combined model was very close to the observed occurrence rate of death within 30, 60, and 120 minutes (**Table 3**), indicating that the model is well calibrated at the population level. Calibration curves calculated at these times, using deciles of the probability scores, demonstrate that the line of identity falls within the 95% CI for nearly all predicted probabilities (**Supplemental Fig. E2**, http://links.lww.com/CCX/A964).

Once probabilities are calibrated, the occurrence rate of predicted probabilities in the population will determine model utility. We show the cumulative distribution functions for the predicted probabilities and the occurrence rate of rapid death for each probability score tertile for the 30-, 60-, and 120-minute predictions in **Supplemental Figures E3** and **E4** (http://links.lww.com/CCX/A964). Selecting the proper probability cutoffs for a low versus a high risk of a rapid death will depend on the center, the organ time limit, and the current burden of DCD on the healthcare system. A lower cutoff of 38% at 30 minutes would classify 40% of the validation cohort as low risk, and only 11% of these low-risk patients would die within this time frame (4% of all patients and 12% of all patients that die in this time frame). At 2 hours, this cutoff would classify 16% of patients in the validation cohort as low-risk, with only 32% of that group dying within time limits (5% of patients and 8% of all patients with rapid death). An upper cutoff of 67% at 2 hours would classify 64% of patients in the validation cohort as high risk, of which 82% would die within time limits (52% of patients and 81% of all patients with rapid death) (**Fig. 2**; and **Supplemental Table E7**, http://links.lww.com/CCX/A964).

## DISCUSSION

This is the first study employing heart rate and blood pressure variability to predict the time to death after WLSM. A model combining vital sign variability and the physician's prediction of time to death achieved an ROC AUC of 0.80 for time to death within 2 hours.

Although we cannot claim superiority of the combined model due to the overlapping CIs for the

**TABLE 3.**

**Model Calibration at the Population Level at the Three Selected Time Points for the Combined Model, Comparing the Observed Proportion of Rapid Death With the Mean Predicted Probability at That Time Point**

| | Derivation (95% CI) | | | Validation (95% CI) | | |
|---|---|---|---|---|---|---|
| Combined Model | Observed Proportion of Rapid Death | Mean Predicted Probability (Raw RSF) | Mean Predicted Probability (Recalibrated) | Observed Proportion of Rapid Death | Mean Predicted Probability (Raw RSF) | Mean Predicted Probability (Recalibrated) |
| 30 min | 40 (35–46) | 34 (32–36) | 40 (37–42) | 36 (28–44) | 36 (33–39) | 43 (38–47) |
| 60 min | 51 (46–58) | 44 (41–46) | 51 (48–54) | 51 (43–60) | 46 (43–49) | 55 (51–60) |
| 120 min | 64 (59–70) | 54 (52–56) | 64 (60–67) | 65 (57–73) | 56 (53–59) | 67 (63–71) |

RSF = random survival forest.

95% confidence intervals were calculated over 5,000 bootstrap iterations.

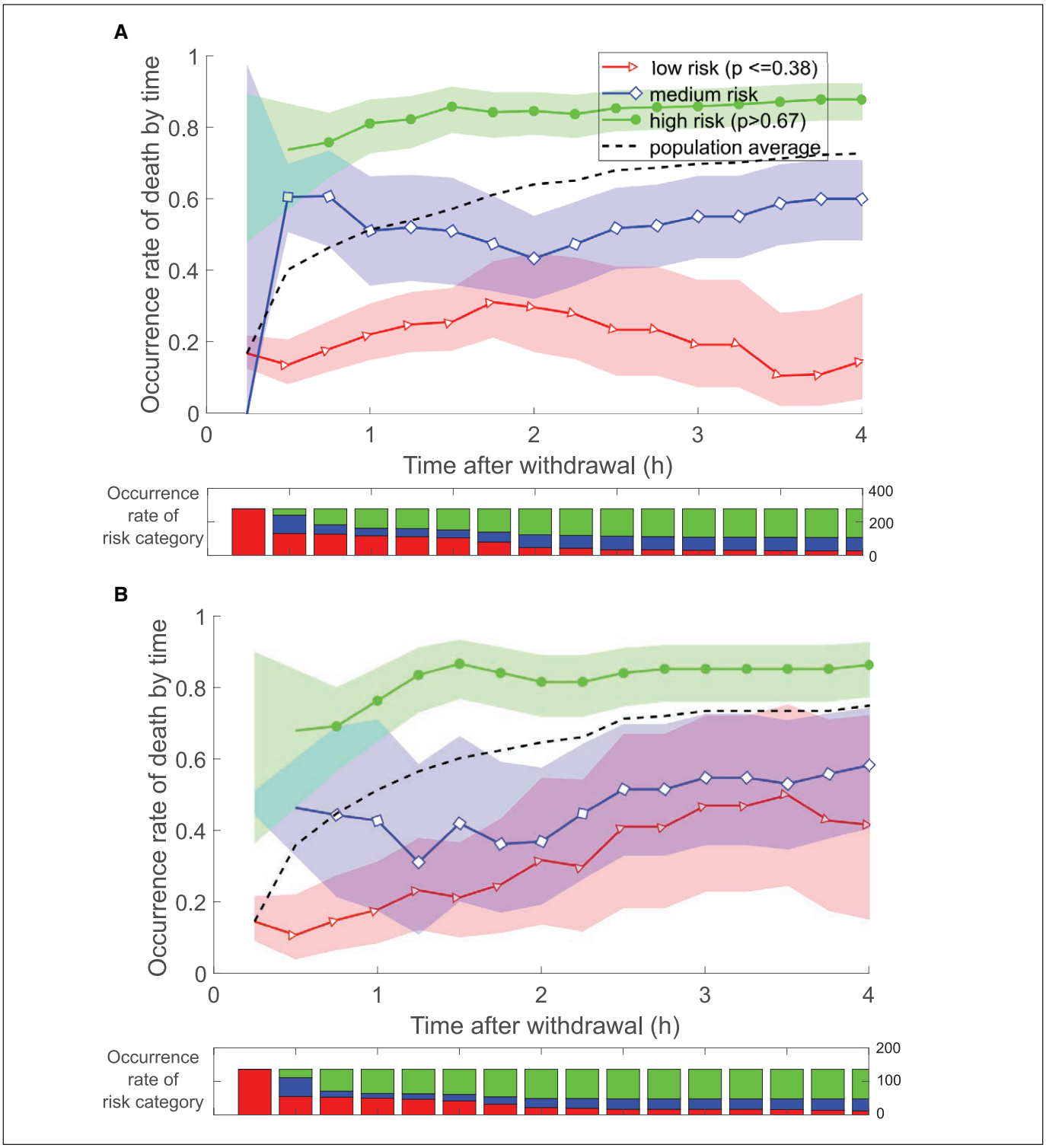


**Figure 2.** Occurrence rate of rapid death, for patients stratified into low, medium, or high-risk categories. Plots of the occurrence rate of rapid death as a function of time, for patients with probability scores in a low-risk (probability score < 0.38), medium-risk (038 < probability score < 0.67), or high-risk category (probability score > 0.67) at each time point, for the derivation (**A**) and validation (**B**) cohorts, using the recalibrated combined model. The *shaded areas* indicate the 95% CI of the occurrence rate for each risk group at each time point. The *histograms below each plot* show the proportion of patients that fall in each risk category at each time point.

different models, the most responsible physician's prediction performed significantly poorer in Canada, which enrolled patients from many sites ($n$ = 16), compared with the Czech Republic (three sites). The physician's prediction of time to death has had variable success in other work (21, 22), and our data suggest

that performance can deteriorate when a diverse range of sites are included. We hypothesize that physicians at lower volume sites that withdraw life-supporting measures or that perform DCD less frequently will be less experienced at predicting time to death. The combined model, however, performed well in all countries (Supplemental Table E5, http://links.lww.com/CCX/A964), suggesting that combining the physician's prediction with variability makes for a more reliable predictor. The clinical features used in this work added no additional benefit and were removed during feature optimization for the combined model. We chose to use clinical features from the models of Brieva et al (20, 22) because these features were easily available, were shown to have good performance, and did not require a pause in mechanical ventilation, which may be a barrier to implementation (27) in some jurisdictions and would not have been possible in this observational study. Future work may consider the inclusion of additional clinical features and investigate the use of variability analysis during a period of reduced mechanical ventilation support.

Altered heart rate variability is closely associated with altered autonomic function. Our findings complement those of Nijhoff et al (6), who recently validated the predictive score for cardiac death in patients in neurocritical state (DCD-N), employing three brainstem reflexes and an oxygenation index (24), in DCD eligible patients. The authors used the DCD-N score to predict death within 30, 60, and 120 minutes, with AUC values of 0.71 (0.66–0.77), 0.77 (0.71–0.83), and 0.80 (0.74–0.86), comparable with the performance of the combined model in our validation cohort (Table 2, and Supplemental Table E5, http://links.lww.com/CCX/A964). The authors noted that although the DCD-N model had good discrimination, it could not be calibrated as it did not provide a probability; a linear prediction model based on the DCD-N model achieved similar discrimination but was poorly calibrated. Our model provides a calibrated probability that death occurs within a given timeframe, rather than providing a Boolean prediction; we envision these probabilities better informing clinical decision-making within the context of each patient. In this use case, proper calibration of a model will be imperative for its successful uptake. The combination of vital sign variability and brainstem reflexes should be considered in future models.

Different organs are more sensitive to ischemia and have different time constraints for donation, which have gradually increased over the years in some locations (5) and may be further increased using reperfusion techniques (3). As RSFs provide a probability of survival versus time curve for each patient, it is not necessary to develop new models whenever these time constraints change. Our combined model shows high sensitivity and negative predictive values at early times after WLSM, whereas the specificity and positive predictive values improve as time increases. Ideally, all potential DCD candidates would have DCD attempted, and model probabilities would inform the surgeons and families of the likelihood of success. However, the severe acute respiratory syndrome coronavirus 2 pandemic has demonstrated that transplantation rates can be significantly reduced when resources are constrained, and marginal candidates are less likely to have DCD attempted (48); a model with a high positive predictive value would better identify cases that are likely to proceed when resources are limited.

As vital sign monitoring is continuous, prediction models employing waveform variability can be updated as the patient's condition changes prior to WLSM. It can take hours to complete assessment and tissue matching of DCD candidates (3), and repeatable prediction is likely to be useful. An initial prediction could be generated at the time of DCD consent to assess a patient's eligibility for DCD and an updated prediction closer to WLSM to provide an expectation for both the family and the DCD team, using the most recent data available. Ultimately, we envision waveform variability-based predictions being incorporated into bedside/central monitors and/or electronic health record systems, which would require further feasibility studies, randomized controlled trials, commercialization, and regulatory approval. We are currently assessing the feasibility of implementing real-time waveform capture for prediction of time to death in DCD patients in an observational study in multiple centers in Ontario, Canada.

This study has several limitations. This was a secondary analysis of the DePPaRT study (36), which did not prescribe a specific pattern of WLSM, and patients likely experienced a variety of withdrawal patterns that could vary by center or region. Patients with a staggered withdrawal process (vasopressors and ventilator settings gradually reduced after WLSM) were excluded. This staggered process likely results in a longer dying process (49) and would need to be considered explicitly in any prediction model (37, 38). Some prediction tools, such

as the United Network for Organ Sharing (37) and University of Wisconsin (26) prediction tools, employ a brief period with reduced ventilation support to assess respiratory drive. Due to the observational nature of this study, this was not possible, but future studies might investigate the use of vital sign waveform variability during such a period to improve prediction. Mechanical ventilation, sedatives, pressors, and inotropes have all been shown to impact variability and/or be useful in predicting time to death (7, 50) and should be considered in future work. Many clinical features that describe the degree of life support or assess neurologic function were not considered. Only ~1/3 of the patients in this study were eligible for DCD, and this model may perform differently when used on DCD patients.

The RSF method, while less prone to overfitting and designed to estimate survival probabilities over time, generates models that are more difficult to interpret and share. It is possible that other machine learning algorithms may show similar or improved performance. Although the data from three different countries were used to derive and validate our models, our model may perform differently in other regions. Further studies assessing the feasibility of implementation of a model employing waveform variability and the associated clinical decision support tool are required.

## CONCLUSIONS

For the first time, a predictive model incorporating vital sign variability, clinical features, and the physician's prediction of a rapid death was developed and validated in patients from three different countries. This combined model achieved ROC AUC values of 0.78, 0.79, and 0.80 for a prediction of time to death within 30 minutes, 1 hour, or 2 hours, respectively. The inclusion of a diverse range of sites from multiple countries improves the likelihood that this model will perform well at new sites. Future work will assess model performance with additional clinical features as well as the feasibility of real-time variability analysis within the context of DCD.

## ACKNOWLEDGMENTS

We are grateful for the input of our patient partners and the dedication and commitment of investigators and research coordinators across Canada, the Czech Republic, and the Netherlands. We are grateful to the patients and their families, without whom this study would not have been possible.

1 *Dynamical Analysis Laboratory, Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, ON, Canada.*

2 *Children's Hospital of Eastern Ontario Research Institute, Ottawa, ON, Canada.*

3 *Canadian Blood Services, Ottawa, ON, Canada.*

4 *Translational Research in Respiratory Diseases Program, McGill University Health Centre and Research Institute, Montreal, QC, Canada.*

5 *Department of Critical Care, Pulmonary Medicine, McGill University, Montreal, QC, Canada.*

6 *Division of Pediatric Critical Care, Children's Hospital of Eastern Ontario, University of Ottawa, Ottawa, ON, Canada.*

7 *Divisions of Thoracic Surgery & Critical Care Medicine, University of Ottawa, Ottawa, ON, Canada.*

*Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's website (http://journals.lww.com/ccejournal).*

*Dr. Scales, Dr. Herry, Dr. van Beinum, Ms. Hornby, Dr. Shahin, Dr. Dhanani, and Dr. Seely helped in design. Dr. Scales, Dr. Herry, Dr. van Beinum, and Ms. Hogue helped in acquisition. Drs. Scales, Scales, and Seely helped in analysis. All authors helped in interpretation and writing.*

*This work is part of the Canadian Donation and Transplantation Research Program and was supported by the Canadian Institutes of Health Research and partners (Grant Number 127880), The Research Institutes of the Children's Hospital of Eastern Ontario and Ottawa Hospital, Physician's Services Incorporated and the Karel Pavlík Foundation.*

*Dr. Seely is a patent holder, director, and shareholder of Therapeutic Monitoring Systems and focused on commercialization of variability-derived clinical decision support tools developed at the Ottawa Hospital Research Institute's Dynamical Analysis Laboratory. Dr. Scales, Dr. Herry, Dr. van Beinum, Ms. Hornby, Dr. Dhanani, and Dr. Seely are named on a patent related to estimating the time to death employing variability monitoring and physiological waveform analysis. The remaining authors have disclosed that they do not have any potential conflicts of interest.*

*For information regarding this article, E-mail: aseely@toh.ca*

*Design and implementation of the study, and article development were coordinated at The Ottawa Hospital and the Children's Hospital of Eastern Ontario.*

## REFERENCES

1. Kramer AH, Hornby K, Doig CJ, et al: Deceased organ donation potential in Canada: A review of consecutive deaths in Alberta. *Can J Anaesth* 2019; 66:1347–1355
2. Canadian Institute for Health Information: Annual Statistics on Organ Replacement in Canada: Dialysis, Transplantation and Donation, 2009 to 2018 [Internet]. 2019. Available at: https://www.cihi.ca/en/organ-replacement-in-canada-corr-annual-statistics-2019 Accessed November 6, 2020

3. Smith M, Dominguez-Gil B, Greer DM, et al: Organ donation after circulatory death: Current status and future potential. *Intensive Care Med* 2019; 45:310–321
4. NHS Blood and Transplant: Annual Report on the Potential Donor Audit [Internet]. 2019. Available at: https://www.odt.nhs.uk/statistics-and-reports/potential-donor-audit-report/. Accessed November 12, 2020
5. Bradley JA, Pettigrew GJ, Watson CJ: Time to death after withdrawal of treatment in donation after circulatory death (DCD) donors. *Curr Opin Organ Transplant* 2013; 18:133–139
6. Nijhoff MF, Pol RA, Volbeda M, et al: External validation of the DCD-N score and a linear prediction model to identify potential candidates for organ donation after circulatory death: A nationwide multicenter cohort study. *Transplantation* 2021; 105:1311–1316
7. Munshi L, Dhanani S, Shemie SD, et al: Predicting time to death after withdrawal of life-sustaining therapy. *Intensive Care Med* 2015; 41:1014–1028
8. Coleman NL, Bonner A: Exploring Australian intensive care physicians clinical judgement during donation after cardiac death: An exploratory qualitative study. *Aust Crit Care* 2014; 27:172–176
9. Mandell MS, Zamudio S, Seem D, et al: National evaluation of healthcare provider attitudes toward organ donation after cardiac death. *Crit Care Med* 2006; 34:2952–2958
10. McHaffie HE, Lyon AJ, Fowlie PW: Lingering death after treatment withdrawal in the neonatal intensive care unit. *Arch Dis Child Fetal Neonatal Ed* 2001; 85:F8–F12
11. Taylor LJ, Buffington A, Scalea JR, et al: Harms of unsuccessful donation after circulatory death: An exploratory study. *Am J Transplant* 2018; 18:402–409
12. Sarti AJ, Sutherland S, Healey A, et al: A multicenter qualitative investigation of the experiences and perspectives of substitute decision makers who underwent organ donation decisions. *Prog Transplant* 2018; 28:343–348
13. Walker W, Sque M: Family bereavement: A case study of controlled organ donation after circulatory death. *Nurs Crit Care* 2019; 24:229–234
14. Moran J: Ninety minutes to die. *J Palliat Med* 2014; 17:112–113
15. Kramer AH, Holliday K, Keenan S, et al: Donation after circulatory determination of death in western Canada: A multicentre study of donor characteristics and critical care practices. *Can J Anaesth* 2020; 67:521–531
16. Krmpotic K, Payne C, Isenor C, et al: Delayed referral results in missed opportunities for organ donation after circulatory death. *Crit Care Med* 2017; 45:989–992
17. Scalea JR, Redfield RR, Rizzari MD, et al: Outcomes and implications of DCD at a high-volume, single-center OPO in the United States. *Ann Surg* 2015; 263:6
18. Lindemann J, Dageforde LA, Vachharajani N, et al: Cost evaluation of a donation after cardiac death program: How cost per organ compares to other donor types. *J Am Coll Surg* 2018; 226:909–916
19. Montgomery JR, Highet A, Hobeika MJ, et al: Going the distance for procurement of donation after circulatory death livers for transplantation-does reimbursement reflect reality? *Clin Transplant* 2020; 34:e13780
20. Brieva J, Coleman N, Lacey J, et al: Prediction of death in less than 60 minutes after withdrawal of cardiorespiratory support in potential organ donors after circulatory death. *Transplantation* 2014; 98:1112–1118
21. Wind J, Snoeijs MG, Brugman CA, et al: Prediction of time of death after withdrawal of life-sustaining treatment in potential donors after cardiac death*. *Crit Care Med* 2012; 40:766–769
22. Brieva J, Coleman N, Lacey J, et al: Prediction of death in less than 60 minutes following withdrawal of cardiorespiratory support in ICUs. *Crit Care Med* 2013; 41:2677–2687
23. Long AC, Muni S, Treece PD, et al: Time to death after terminal withdrawal of mechanical ventilation: Specific respiratory and physiologic parameters may inform physician predictions. *J Palliat Med* 2015; 18:1040–1047
24. de Groot YJ, Lingsma HF, Bakker J, et al: External validation of a prognostic model predicting time of death after withdrawal of life support in neurocritical patients. *Crit Care Med* 2012; 40:233–238
25. Kotsopoulos AMM, Böing-Messing F, Jansen NE, et al: External validation of prediction models for time to death in potential donors after circulatory death. *Am J Transplant* 2018; 18:890–896
26. Lewis J, Peltier J, Nelson H, et al: Development of the University of Wisconsin donation after cardiac death evaluation tool. *Prog Transplant* 2003; 13:265–273
27. Coleman NL, Brieva JL, Crowfoot E: Prediction of death after withdrawal of life-sustaining treatments. *Crit Care Resusc* 2008; 10:278–284
28. Green GC, Bradley B, Bravi A, et al: Continuous multiorgan variability analysis to track severity of organ failure in critically ill patients. *J Crit Care* 2013; 28:879-e1
29. Schmidt H, Lotze U, Ghanem A, et al: Relation of impaired interorgan communication and parasympathetic activity in chronic heart failure and multiple-organ dysfunction syndrome. *J Crit Care* 2014; 29:367–373
30. Pontet J, Contreras P, Curbelo A, et al: Heart rate variability as early marker of multiple organ dysfunction syndrome in septic patients. *J Crit Care* 2003; 18:156–163
31. Barnaby DP, Fernando SM, Herry CL, et al: Heart rate variability, clinical and laboratory measures to predict future deterioration in patients presenting with sepsis. *Shock* 2019; 51:416–422
32. Chiew CJ, Liu N, Tagami T, et al: Heart rate variability based machine learning models for risk prediction of suspected sepsis patients in the emergency department. *Medicine (Baltimore)* 2019; 98:e14197
33. Seely AJ, Bravi A, Herry C, et al; Canadian Critical Care Trials Group (CCCTG): Do heart and respiratory rate variability improve prediction of extubation outcomes in critically ill patients? *Crit Care* 2014; 18:R65
34. Ryan ML, Ogilvie MP, Pereira BM, et al: Heart rate variability is an independent predictor of morbidity and mortality in hemodynamically stable trauma patients. *J Trauma* 2011; 70:1371–1380
35. Norris PR, Ozdas A, Cao H, et al: Cardiac uncoupling and heart rate variability stratify ICU patients by mortality. *Ann Surg* 2006; 243:11
36. Dhanani S, Hornby L, van Beinum A, et al; Canadian Critical Care Trials Group; Canadian Donation and Transplantation Research Program: Resumption of cardiac activity after withdrawal of life-sustaining measures. *N Engl J Med* 2021; 384:345–352

37. DeVita MA, Brooks MM, Zawistowski C, et al: Donors after cardiac death: Validation of identification criteria (DVIC) study for predictors of rapid death. *Am J Transplant* 2008; 8:432–441
38. Suntharalingam C, Sharples L, Dudley C, et al: Time to cardiac death after withdrawal of life-sustaining treatment in potential organ donors. *Am J Transplant* 2009; 9:2157–2165
39. Bravi A, Longtin A, Seely AJ: Review and classification of variability analysis techniques with clinical applications. *Biomed Eng Online* 2011; 10:90
40. Ishwaran H, Kogalur UB, Blackstone EH, et al: Random survival forests. *Ann Appl Stat* 2008; 2:841–860
41. Wright MN, Ziegler A: ranger: A fast implementation of random forests for high dimensional data in *C++* and *R*. *J Stat Softw* 2017; 77:1–17
42. R Core Team: R: A Language and Environment for Statistical Computing. Vienna, Austria, R Foundation for Statistical Computing, 2019
43. Ishwaran H, Kogalur UB, Chen X, et al: Random survival forests for high-dimensional data: Random survival forests for high-dimensional data. *Stat Anal Data Min ASA Data Sci J* 2011; 4:115–132
44. Breiman L: Random forests. *Mach Learn* 2001; 45:5–32
45. Nembrini S, König IR, Wright MN: The revival of the Gini importance? *Bioinformatics* 2018; 34:3711–3718
46. Zadrozny B, Elkan C: Obtaining calibrated probability estimates from decision trees and naive Bayesian classifiers. *Icml* 2001; 1:609–616
47. Gupta C, Podkopaev A, Ramdas A: Distribution-free binary classification: Prediction sets, confidence intervals and calibration. *Adv Neural Inf Proces Syst* 2020; 33:3711–3723
48. Stock PG, Wall A, Gardner J, et al; TTS Ethics Committee: Ethical issues in the COVID era: Doing the right thing depends on location, resources, and disease burden. *Transplantation* 2020; 104:1316–1320
49. Delaney JW, Downar J: How is life support withdrawn in intensive care units: A narrative review. *J Crit Care* 2016; 35:12–18
50. Karmali SN, Sciusco A, May SM, et al: Heart rate variability in critical care medicine: A systematic review. *Intensive Care Med Exp* 2017; 5:33