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

Authors: Nathan B Scales, Christophe L Herry, Amanda van Beinum, Melanie L Hogue, Laura Hornby, Jason Shahin, Sonny Dhanani, Andrew JE Seely

Online Data Supplement
Research Ethics Board Approvals

Children’s Hospital of Eastern Ontario Research Ethics Board: No. 14/08E (coordinating site)

Ottawa Health Science Network Research Ethics Board: Protocol # 20140337-01H (2 sites)

St. Michael’s Hospital Research Ethics Board: REB# 14-335

Western University Health Science Research Ethics Board: HSREB File Number 105752 (2 sites)

University of Alberta Health Research Ethics Board: Study ID Pro00063243

Nova Scotia Health Authority Research Ethics Board: NSHA REB ROMEO File # 1020827

Sunnybrook Health Sciences Centre Research Ethics Board: Protocol ID # 042-2015

University of British Columbia Clinical Research Ethics Board: UBC CREB # H14-02114

Hamilton Integrated Research Ethics Board: REB Project # 14-405

McGill University Health Centre Genetics/Population Research Ethics Board: #14-204 GEN

University of Calgary Conjoint Health Research Ethics Board: Ethics ID REB14-1337 (3 sites)

Mount Sinai Hospital Research Ethics Board: #14-0340-E

Queen’s University Health Sciences & Affiliated Teaching Hospitals Research Ethics Board (HSREB): ROMEO/TRAQ# 6018723

Ethics Committee of the General University Hospital, Prague: #503/15 S-IV (2 sites)

Ethics Committee of the Medical Faculty of Charles University: EK-VP/68/0/2014

Maastricht UMC REB: METC 16-4-174
Time of death determination

For the majority of patients, time of death was determined through beat-by-beat adjudication of the waveform data, as determined by at least two independent adjudicators, as described for the main DePPaRT study (1), and time of death was defined as 5 minutes after the last arterial blood pressure pulse with a pulse pressure above 5 mmHg. If the last pulse was not adjudicable, other waveform data were used to confirm that the time of death determination from the chart was sufficiently accurate for the time-to-death analyses performed in this paper (i.e., electrical activity stopped within minutes of the charted time, or waveform activity confirmed to continue for longer than 2 hours). For DCD patients without waveform data available after withdrawal of life supporting measures, the chart data were assumed to be correct as these patients were closely monitored at the time of death. Patients that could not have their time to death confirmed as described above were excluded. As this was part of an observational study, we did not specify death determination methods to be used by the bedside physicians. The time of the last pulse determined by adjudication included any resumptions that may have occurred after a cessation in cardiac activity.

Heart Rate and Blood Pressure Variability

Vital sign waveform data were captured from bedside monitors beginning from up to one 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 (GE, Philips, Spacelabs, or Draeger). Waveform capture relied on tools or APIs developed by the vendors themselves, or using software developed by third party companies. For Philips monitors/central stations, we used Philips Holter/Research Data Export, the Philips Data Export Interface, or BedMasterEx software; for GE monitors, we used GE’s iCollect or BedMasterEx software; for Draeger, we used Draeger’s WinAccess API; and for Spacelabs, we used the Intersys Clinical Suite (ICS) to extract the data. For organ donation patients, we collected data while the patient was in the ICU, and in a few cases during/after transport to the PACU if this was done prior to WLSM. We did not collect data from the operating room.
Deidentified waveform data, collected using the applicable method at each site, were uploaded to the secure study server. These data were then downloaded over a secure connection to the Ottawa Hospital Research Institute for analysis.

The sampling frequency for the electrocardiogram (ECG) ranged from 125Hz to 500Hz, with a median value of 224Hz. While sampling rates between 250 and 500 Hz may be considered optimal (2), other work has suggested that sampling rates of at least 125 Hz are sufficient to capture clinically relevant differences in heart rate variability (3). The lowest sampling rate of 125Hz was only used for 25 Philips recordings which did not include R-peak detections from Philips’ Arrhythmia Monitoring Algorithm, which uses a higher sampling rate for R-peak detection than available in the downsampled exported waveforms. The sampling frequency for the arterial blood pressure (ABP) ranged from 100Hz to 300Hz. Blood pressure variability is less studied than heart rate variability, but as the arterial blood pressure waveform peaks are not as narrow as the R-peaks in the ECG waveform, sampling rates of 100Hz should be sufficient. The sampling rates used in this study were predetermined by the monitoring companies and the relevant export method.

The ECG and ABP waveform data from 60 minutes to 5 minutes prior to WLSM were processed to obtain beat-to-beat event times series (R-peak-to-R-peak interval (RRI), systolic (SBP), diastolic (DBP), mean, and pulse blood pressures), using previously reported software (1, 4–6).

In most studies employing heart rate variability analysis, ectopic beats and arrhythmias are filtered from the R-R interval time series prior to analysis (2, 5). However, at the end of life, we found that arrhythmias were so common that such filtering would result in the removal of approximately 25% of enrolled patients, as they would have insufficient data for analysis. Consequentially, we did not filter all ectopic beats and arrhythmias, but instead performed a coarse filtering on each RRI series to remove missed beats only, using as a threshold a relative change in adjacent RRIs of at least 85%, as well as any intervals outside the normal physiologic range. This coarse filtering may have contributed to the relatively lower p-values of RRI-based measures in Supplementary Table 3. Future work will examine the possibility of performing separate analysis pipelines based on the degree of arrhythmia present.

A low pass, elliptic infinite impulse response filter was applied to the raw arterial blood pressure waveform to facilitate peak detection. The systolic and diastolic events were identified on a beat-
by-beat basis. Dicrotic notches were removed by discarding any peaks <0.25 seconds from an identified peak, or any peaks with less than half the amplitude of neighboring peaks occurring within 0.5 seconds. Regions with disconnections or clipping of either systolic or diastolic pressures were removed. Following Sun et al (7), we defined poor quality beats as beats with SBP > 300, DBP < 20, mean ABP < 30, mean ABP > 200, ΔSBP > 20, ΔDBP > 20 mmHg, pulses with a slope of >−40 mmHg/100 ms, or with an instantaneous heart rate <20 or >200 beats/minute. We removed beats with a pulse pressure <10 mmHg rather than the suggested 20 mmHg (7), as several patients had very low pulse pressures prior to withdrawal.

A comprehensive suite of 14 variability measures for each of the 4 blood pressure event time series, and 15 for the RRI time series (see Supplemental Table E1), was calculated for each patient using custom software, including statistical, nonlinear, entropic, information theoretic measures, and measures using symbolic dynamics, for a total of 71 variability features. Variability was averaged across all windows of 750 beats, which corresponds to approximately 8 minutes (the mean HR in our population was 94 beats/min), so that most patients would have at least 2 windows for analysis (the median number of windows per patient was 6, with an IQR of 4-8 and a range of 1-12). The RRI time series was normalized by the mean RRI interval in each window prior to calculation of variability (8).

**Model Development**

The derivation cohort was created by randomly selecting 2/3 of the patients with usable waveform data, with the remaining 1/3 of the patients set aside for the validation cohort. Patients were selected so that each cohort contained an equivalent proportion of patients from each country.

We used random survival forests to develop our predictive models, using the **ranger** (9) package in R (10). Missing values were imputed using the MissForest algorithm (11) through the **missRanger** package, an iterative imputation method that also employs random forests. The time to death for each patient was divided into 15-minute intervals (rounded up), up to 24 hours after WLSM. Each patient was assigned an integer value for their time to death (1 to 96) along with a Boolean value to indicate if their time to death was censored. All patients with a time to death greater than 24 hours were censored.
The survival model yielded a probability of survival $S(t)$ for each 15-minute interval up until 24 hours after WLSM, for each patient. We calculated the probability of dying at a specific time as $D(t) = 1 - S(t)$. The probability of dying at a specific time (for example, $t = 30, 60, \text{or } 120$ minutes) was then used as a score to predict if a given patient would die within this time. Note that each model made predictions using features collected prior to WLSM only.

We optimized the set of variability features used in the model by ranking features based on their impurity-corrected importance values (12) over a 10-fold cross validation of the derivation set, with each variable $j$ given a cumulative score of $CS(j) = \sum m - k_{ij} + 1$, where $m$ is the number of features in the dataset, and $k_{ij}$ is the $i$th ranking (over the 10 folds) of the $j$th variable. The time-to-death model was retrained on the training set using only the 30 highest ranked variables, adding one variable at a time, starting with the highest ranked variable, with a 10-fold cross-validation per added variable. In general, each added feature would increase the mean AUC ROC, until the performance “plateaued”. Using all $S$ features up to the feature with the highest observed AUC, we then removed features one at a time, in the original order that they were added, and tested the performance of the model with the removed variable, using 10-fold cross-validation. If the reduced model had equal or better performance than the highest observed AUC, the feature was removed from the model. This process continued for all $S$ features.

Using only the reduced set of the top ranked variables, we then performed another 10-fold cross validation of the derivation cohort to determine the performance of the reduced model in the derivation cohort. Finally, we combined all patients from the derivation cohort and the top ranked features in a final model, which we tested on the validation cohort. We used 5000 bootstrap iterations to calculate confidence intervals in the validation cohort.

The top variables selected for the variability-based model were, starting from the highest ranked variable: dbp_DFA_AUC, mabp_mupm, dbp_mupm, mabp_Poincare_SD2, sbp_mupm, dbp_Poincare_SD2, rri_ShannEn, rri_gcount, rri_mupm, rri_DFA_AUC, sbp_DFA_Alpha_1, dbp_gcount, rri_Poincare_SD2, mabp_QSE

The top variables selected for the clinical model were, starting from the highest ranked variable: systolic blood pressure (chart), analgesia use, Glasgow Coma Scale score prior to WLSM, spontaneous respiration rate, and positive end-expiratory pressure (PEEP).
The top variables selected for the combined model were, again starting from the highest ranked variable: Physician Prediction of Timely Death, mabp_DFA_AUC, dbp_Poincare_SD2, rri_DFA_AUC, mabp_mupm, rri_Poincare_SD2, mabp_Poincare_SD2, dbp_gecount, rri_mupm, dbp_mupm, rri_QSE, sbp_mupm, dbp_ShannEn, dbp_QSE, rri_CVI. Note that no clinical features were retained in the reduced model. Please see the Supplemental Table E1 for the definitions of the variability features.
## Variability Features

| Abbreviation | Longer Name               | Definition                                                                                                                                                                                                 |
|--------------|---------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| sbp          | systolic blood pressure   | Beat-by-beat measurement of the systolic blood pressure from the arterial blood pressure waveform                                                                                                      |
| mabp         | mean blood pressure       | Beat-by-beat measurement of the mean blood pressure from the arterial blood pressure waveform                                                                                                           |
| dbp          | diastolic blood pressure  | Beat-by-beat measurement of the diastolic blood pressure from the arterial blood pressure waveform                                                                                                       |
| abp          | pulse pressure            | Beat-by-beat measurement of the pulse pressure from the arterial blood pressure waveform                                                                                                                  |
| rri          | R-peak to R-peak interval | Beat-by-beat measurement of the R-peak to R-peak interval of the ECG waveform                                                                                                                             |
| mupm         | Mean Rate Per Minute      | $60/(\text{mean of beat-to-beat measurement or R-R interval over 750 beat window})$. For RRI, this would give the mean heart rate. For blood pressure, a mean SBP of 120 would give a mupm of 0.5.                     |
| DFA_Alpha_1  | Detrended Fluctuation Analysis, Scaling Exponent Alpha 1 | Detrended Fluctuation Analysis quantifies how the fluctuations in a signal scale with the number of samples of that signal. The trend is removed using a linear regression of the cumulative time series. The fluctuations are measured as the standard deviation in each window of the detrended series, over a variety of window sizes. DFA Alpha 1 is the slope on the log-log plot of the fluctuations over increasing time scales for the detrended integrated series. |
| DFA_AUC      | Detrended Fluctuation Analysis, Area under the Curve | The area under the curve of the log-log plot of the fluctuations over increasing time scales for the detrended integrated series.                                                                 |
| Poincare_SD1 | Poincaré standard deviation 1 | The standard deviation of the Poincaré plot perpendicular to the line of identity (equivalent to the square root of the mean squared difference of successive intervals, or RMSSD). The Poincaré plot plots the value of $f(x+1)$ vs $f(x)$ (i.e. each subsequent point is plotted on the y-axis, and the current point is plotted on the x-axis). SD1 was not used as a feature on its own, but was used in the calculation of CVI. |
| Poincare_SD2 | Poincaré standard deviation 2 | The standard deviation of the Poincaré plot along the line of identity (related to the standard deviation).                                                                                         |
| ShannEn | Shannon Entropy | The entropy of the time series in the window. ShannEn = −\sum p(i)ln(p(i)), where p(i) is the probability of the i\text{th} histogram element for the histogram of all observed values in the window. The sum is over all bins used in the histogram. |
|---|---|---|
| QSE | Quadratic Sample Entropy | The sample entropy (SE) is the negative natural logarithm of the conditional probability that a dataset of length N, having repeated itself for m samples within a tolerance \( \varepsilon \), will repeat itself for \( m + 1 \) samples, without allowing self matches. The QSE adjusts the sample entropy so that different tolerance windows have comparable entropy. |
| gcount | Grid Count | The time series is transformed into a grid, similar to a discretized Poincaré plot, with each pixel assigned a 1 if it was visited during the time series, and 0 otherwise. The Grid Count sums the total over all pixels, divided by the number of pixels in the grid. Also known as Box Count. |
| CVI | Cardiac Vagal Index | Defined as \( \log_{10}(16 \text{ SD1 \ SD2}) \), where SD1 and SD2 are the standard deviations of the Poincaré plot. Conceptually, it is the log of the area of an ellipse defined by axes of SD1, SD2. |
| LF_HF_ratio_Lombscargle | Ratio of the low and high frequency spectral components | The ratio of the low and high frequency components, calculated using the LombScargle method. Used only for the RRI time series. |
| Multifractal_c2 | Multifractal spectrum cumulant of the second order | A parameter describing the width of the multifractal spectrum, which is a measure of the amount of variability at different analysis scales, measured using a wavelet transform. |
| Multiscale_Entropy | Multiscale entropy | The sample mean is calculated for progressively larger window sizes of 1:N, yielding N time series of sample means. The entropy is calculated for each time series; the multiscale entropy is given by the average entropy over all time series. |
| SymDp1_2 | Symbolic Dynamics: Percentage of one variation sequences | The amplitude of the time series is divided into 4 unequal (nonuniform) regions, and each assigned a symbol in a 4-character alphabet. Words are developed using 3 successive symbols. A one variation sequence is defined as a word with one successive difference (and two successive identical symbols, i.e. b,b,a). The percentage of words with one variation is calculated. |
**SymDp2_2**  
**Symbolic Dynamics:**  
Percentage of two variation sequences  
Similar to above, but quantifying the percentage of words with two-variation sequences, which are defined as words with two successive differences (i.e. a,b,a).

**eScale**  
**Embedding Scaling Exponent**  
An estimate of how the variance of the time series, reconstructed in a phase space with a lag of 1, changes with embedding dimension m (i.e. x(t), x(t-1)… x(t-m)). The slope of a least-squares fit of the log-log plot of the variance vs m yields the embedding scaling exponent.

**pDpR**  
**Recurrence quantification analysis: determinism/recurrences**  
The time series is plotted in a phase space, and the distance between points is calculated. Distances below a given threshold are considered “recurrences”. The % recurrences is calculated as the number of recurrences over the total possible number of recurrences. The % determinism is calculated by the number of recurrences that are part of a diagonal, divided by the number of occurrences (excluding the line of identity). The ratio of these two percentages yields pDpR.

---

**Model Calibration**

**Assessing calibration**

Model calibration was assessed at the population level by comparing the occurrence rate of rapid death at 30 minutes, 1 hr, and 2 hrs with the mean predicted probabilities at those times for all patients. Calibration curves were generated by calculating the fraction of patients with an estimated probability of dying that died within the given time frame, using equal mass probability bins (deciles). Confidence intervals were determined over 5000 bootstrap iterations. Histograms of the fraction of cases with probabilities falling in each bin, as well as the fraction of cases that died within the time period, were also calculated and displayed under each calibration curve (Supplemental Figure E2).

**Recalibration**

The raw random forest survival probabilities tended to be underconfident. A recalibration step was included in the model derivation pipeline, employing a uniform-mass binning method (13,
14). For each fold of the cross-validation step, the training set was divided into two datasets, one to learn the deciles of the probability scores (40%), and the second to assess the empirical probability of death within each decile (60%). This was done at each time point (every 15 minutes), using 100 bootstrap iterations. Tests using larger bootstrap runs confirmed 100 bootstrap runs provided sufficient precision. The probability scores in the test set for each fold were then adjusted according to the observed probability of dying in the corresponding decile. The mean, median, standard deviation, interquartile range, and 95% confidence of the bootstrapped recalibrated scores were recorded at each time point for each patient. Isotonic regression was then used to ensure that the mean probabilities over time were monotonically increasing, as would be expected. This typically resulted in only minor corrections to the mean probability scores, which were always within the specified 95% confidence interval.

To recalibrate the probability scores from the validation cohort, the derivation cohort was split into two test sets to learn the deciles of the probability scores (40%) and the empirical probabilities of dying within each decile (60%), again over 100 bootstrap iterations. The validation scores were recalibrated using the observed probabilities in the corresponding deciles, again recording the mean, standard deviation, and confidence intervals at each time point for each patient. Isotonic regression was used to ensure that the mean values were monotonically increasing. Although recalibration using binned probabilities has the potential to negatively impact a model’s discrimination, the AUC values of the recalibrated model were not appreciably different than those of the uncalibrated model (not shown).
Supplemental Figure E1) Exemplary probability of dying curves generated by the random survival forest models for three different patients, with times to death provided in the legend. The black dashed lines show the times at which the scores are generated for the prediction models (t=0.5, 1, and 2h).
Supplemental Table E2: Model performance in the derivation cohort, for predictions of time to death within 30 minutes, 1 hour, or 2 hours from WLSM. Death within the given time limit was defined as positive, and death outside the time interval was defined as negative. Sens: sensitivity. Spec: specificity. PPV: positive predictive value. NPV: negative predictive value. ROC AUC: area under the receiver operating characteristic curve. The median value over 10-fold cross-validation runs is shown (95% confidence interval in brackets).
| Variable                                         | Spearman correlation | Uncorrected p value |
|-------------------------------------------------|----------------------|----------------------|
| Physician Prediction of Timely Death            | -0.54                | 1E-33                |
| mabp (DFA_AUC)                                  | 0.40                 | 6E-18                |
| dbp (DFA_AUC)                                   | 0.38                 | 8E-17                |
| mabp (Poincare_SD2)                             | 0.33                 | 7E-13                |
| pH*                                             | 0.33                 | 2E-12                |
| dbp (Poincare_SD2)                              | 0.33                 | 2E-12                |
| sbp (DFA_AUC)                                   | 0.32                 | 1E-11                |
| mabp (QSE)                                      | 0.31                 | 7E-11                |
| Systolic Blood Pressure (Chart)*                | 0.29                 | 6E-10                |
| sbp (mupm)                                      | 0.29                 | 1E-09                |
| mabp (mupm)                                     | 0.28                 | 2E-09                |
| dbp (QSE)                                       | 0.28                 | 4E-09                |
| rri (ShannEn)                                   | 0.28                 | 5E-09                |
| GCS prior to WLSM*                              | 0.28                 | 5E-09                |
| rri (DFA_AUC)                                   | 0.28                 | 5E-09                |
| rri (QSE)                                       | 0.27                 | 6E-09                |
| rri (Poincare_SD2)                              | 0.26                 | 7E-08                |
| rri (gcount)                                    | 0.25                 | 8E-08                |
| sbp (DFA_Alpha_1)                               | 0.25                 | 2E-07                |
| rri (CVI)                                       | 0.25                 | 2E-07                |
| abp (mupm)                                      | 0.24                 | 3E-07                |
| dbp (DFA_Alpha_1)                               | 0.21                 | 7E-06                |
| dbp (mupm)                                      | 0.21                 | 1E-05                |
| PEEP prior to WLSM*                             | -0.21                | 1E-05                |
| dbp (gcount)                                    | 0.20                 | 4E-05                |
| dbp (ShannEn)                                   | 0.19                 | 6E-05                |
| rri (mupm)                                      | -0.18                | 0.0002               |
| Spontaneous Respiration Rate*                   | 0.17                 | 0.0004               |
| Analgesia used*                                 | -0.09                | 0.06                 |

Supplemental Table E3: Spearman correlation of features with time to death after WLSM. Out of the 60 available variability features, only those identified as most important by the random survival forest algorithm are shown, as well as the 6 clinical features used by Brieva et al. (15, 16) (indicated with an *) and the physician prediction of timely death. Correlations were calculated using the combination of the derivation and validation cohorts. Correlation values were not used in the feature selection process. mabp: mean arterial blood pressure (beat-by-beat). sbp: systolic arterial blood pressure (beat-by-beat). dbp: diastolic arterial blood pressure (beat-by-beat); abp: pulse (arterial) pressure (beat-by-beat); PEEP: positive end-expiratory pressure; GCS: Glasgow Coma Scale score.
Supplemental Figure E2: Assessment of calibration for the recalibrated combined model in the derivation (A) and validation (B) cohorts. The solid lines show the observed proportion of patients assigned a given probability from the recalibrated model that experienced a rapid death by each time point (t=0.5, 1, or 2 hours). The dashed black line of identity is the expected observed proportion for a given probability. The shaded regions indicate the 95% confidence interval (over 5000 bootstrap iterations). The histograms below each figure show the overall fraction of patients that had a probability score within the window of interest. The yellow histograms show the fraction of patients with these probability scores that died within the stated timeframe. Probability scores were divided into 10 bins with equal numbers of patients (deciles).
Supplemental Figure E3: Occurrence rate of rapid death (death < 30 minutes (left), < 60 minutes (middle), < 120 minutes (right), for patients with calibrated probability scores in the lowest, middle, or highest tertile of the derivation cohort (left to right within each panel). The validation cohort used the same probability cutoffs as the derivation cohort (shown on the x-axis). The probability scores were calculated from a calibrated random survival forest model employing the highest ranked vital sign variability features, clinical features from Brieva et al. (15, 16), as well as the physician’s prediction of timely death. The dashed purple and the dash-dotted blue horizontal lines show the occurrence rate of rapid death in the derivation and validation cohorts respectively as a whole. The numbers below the “f” and “s” indicate the number of deaths in the “fast” or “slow” category for each tertile.
### Demographic Characteristics

| Demographic Characteristics | DCD Eligible (N=157) | DCD Ineligible (N = 272) |
|-----------------------------|----------------------|-------------------------|
| Average age (STD, range)    | 57 (13, 22-79)       | 67 (16, 18-95)          |
| Female gender               | 65 (41%)             | 104 (38%)               |
| Chronic condition           | 97 (62%)             | 246 (90%)               |
| Primary reason for ICU Admission |                   |                         |
| Neurologic                  | 113 (72%)            | 114 (42%)               |
| Cardiac                     | 5 (3%)               | 11 (4%)                 |
| Respiratory                 | 14 (9%)              | 43 (16%)                |
| Sepsis/infection            | 2 (1%)               | 56 (21%)                |
| Trauma                      | 12 (8%)              | 7 (3%)                  |
| Other\(^a\)                 | 11 (7%)              | 41 (15%)                |
| CPR in previous 24hr        | 22 (14%)             | 39 (14%)                |
| Median GCS at ICU admission (IQR) | 3 (3-6) | 5 (3-11) |
| Avg. APACHEII Score (IQR, range) | 26 (21-30, 11-46) | 29 (23-35, 6-55) N=270 |
| Reported TBI                | 20 (13%)             | 21 (8%)                 |
| Median LOS in ICU, days (range) | 4 (0-23) N=156 | 3 (0-34) |
| DCD ODO eligible            | 157 (100%)           | 0                       |
| DCD attempted               | 73 (46%)             | 0                       |
| DCD successful donors       | 48 (31%)             | 0                       |
| Median Time to Death, hours (IQR, range) | 0.89 (0.31-6.82, 0.07-269.9) | 0.92 (0.31-4.18, 0.11-220.4) |
| Death within 0.5, 1, 2h (%) | 43%, 51%, 61%        | 36%, 51%, 65%           |

### Life Sustaining Measures

| Life Sustaining Measures          | DCD Eligible (N=157) | DCD Ineligible (N = 272) |
|-----------------------------------|----------------------|-------------------------|
| Receiving IMV                     | 151 (96%)            | 228 (84%)               |
| Extubated as part of WLSM         | 136 (87%)            | 139 (51%)               |
| On Vasopressors/Inotropes         | 55 (35%)             | 168 (62%)               |
| Receiving Sedation                | 113 (72%)            | 188 (69%)               |
| Receiving Analgesia               | 144 (92%)            | 253 (93%)               |

Supplemental Table E4: Demographics data by DCD eligibility status. \(^a\): other reasons for admission include gastrointestinal bleeding, abdominal aortic aneurysm, multiple causes, hypovolemic shock, and multiorgan failure. GCS: scores on the Glasgow Coma Scale (range from 3 to 15, with lower scores indicating a reduced level of consciousness). Scores on the Acute Physiology and Chronic Health Evaluation (APACHE) II range from 0 to 71, with higher scores indicating more severe disease. TBI: traumatic brain injury. LOS: length of stay. DCD: donation after circulatory death. ODO: organ donation organization. IMV: invasive mechanical ventilation. WLSM: withdrawal of life sustaining measures. STD: standard deviation. IQR: interquartile range.
| Prediction | Model                  | ROC AUC (95%CI)                                                                 |
|-----------|------------------------|-------------------------------------------------------------------------------|
|           | Derivation (N=288)     | Validation (N=141) | Canada (N=206) | Czech Republic (N=189) | Netherlands (N = 34) | DCD Eligible (N=157) |
| 30 minutes| Variability alone      | 0.69 (0.52-0.8)      | 0.72 (0.64-0.81) | 0.74 (0.67-0.8) | 0.63 (0.54-0.71) | 0.7 (0.52-0.87) | 0.73 (0.65-0.8) |
|           | Brieva features        | 0.72 (0.51-0.91)     | 0.67 (0.58-0.76) | 0.73 (0.67-0.8) | 0.66 (0.57-0.74) | 0.72 (0.53-0.88) | 0.69 (0.61-0.77) |
|           | Physician prediction   | 0.73 (0.68-0.79)     | 0.76 (0.69-0.83) | 0.71 (0.64-0.77) | 0.81 (0.76-0.86) | 0.85 (0.72-0.95) | 0.75 (0.67-0.82) |
|           | Combined               | 0.79 (0.64-0.98)     | 0.78 (0.64-0.86) | 0.79 (0.64-0.85) | 0.79 (0.72-0.85) | 0.85 (0.68-0.97) | 0.83 (0.76-0.89) |
| 1 hour    | Variability alone      | 0.76 (0.56-0.87)     | 0.72 (0.64-0.8)  | 0.78 (0.64-0.78) | 0.72 (0.64-0.79) | 0.65 (0.51-0.84) | 0.73 (0.64-0.8)  |
|           | Brieva features        | 0.73 (0.55-0.92)     | 0.68 (0.59-0.77) | 0.77 (0.70-0.83) | 0.72 (0.64-0.79) | 0.59 (0.50-0.79) | 0.7 (0.62-0.78)  |
|           | Physician prediction   | 0.77 (0.72-0.82)     | 0.72 (0.65-0.79) | 0.69 (0.62-0.75) | 0.84 (0.79-0.89) | 0.79 (0.68-0.9)  | 0.72 (0.65-0.79) |
|           | Combined               | 0.81 (0.71-1)        | 0.79 (0.71-0.87) | 0.79 (0.63-0.86) | 0.85 (0.79-0.9)  | 0.74 (0.55-0.91) | 0.8 (0.73-0.87)  |
| 2 hours   | Variability alone      | 0.74 (0.66-0.84)     | 0.7 (0.61-0.79)  | 0.76 (0.69-0.83) | 0.72 (0.64-0.79) | 0.77 (0.57-0.92) | 0.71 (0.62-0.79) |
|           | Brieva features        | 0.73 (0.52-0.9)      | 0.63 (0.54-0.73) | 0.75 (0.68-0.82) | 0.67 (0.59-0.74) | 0.62 (0.51-0.86) | 0.7 (0.61-0.78)  |
|           | Physician prediction   | 0.77 (0.72-0.82)     | 0.73 (0.66-0.79) | 0.67 (0.61-0.74) | 0.84 (0.79-0.89) | 0.74 (0.64-0.84) | 0.73 (0.66-0.79) |
|           | Combined               | 0.84 (0.6-0.97)      | 0.8 (0.72-0.88)  | 0.79 (0.71-0.85) | 0.84 (0.77-0.89) | 0.81 (0.61-0.95) | 0.81 (0.74-0.88) |

Supplemental Table E5: Sensitivity analysis for model performance (ROC AUC) in different subgroups, including patients from Canada, the Czech Republic, the Netherlands, or DCD eligible patients. Prediction performance was assessed using the prediction results from derivation and validation cohorts combined. The results from the derivation and validation cohorts are included for comparison. Physician prediction is the physician prediction of timely death within 1 hour. The “Combined” model combined the variability features, clinical features, and physician prediction.
| Demographic Characteristics For Excluded Patients<sup>a</sup> | Staggered withdrawals<sup>b</sup> (N=94) | Recording problems (N=44) | Poor quality waveforms (N = 45) | Other exclusions<sup>c</sup> (N=33) |
|--------------------------------------------------|--------------------------------|--------------------------|-------------------------------|--------------------------------|
| Average age (STD, range)                         | 65 (15, 24-92)               | 65 (13, 39-93)           | 61 (15, 28-83)                | 67 (16, 24-90)                |
| Female gender                                    | 29 (31%)                     | 14 (32%)                 | 19 (42%)                      | 13 (39%)                      |
| Chronic condition                                | 86 (91%)                     | 38 (86%)                 | 38 (84%)                      | 25 (76%)                      |
| Primary reason for ICU Admission                 |                               |                          |                               |                               |
| Neurologic                                       | 27 (29%)                     | 23 (52%)                 | 19 (42%)                      | 14 (42%)                      |
| Cardiac                                          | 4 (4%)                       | 2 (5%)                   | 1 (2%)                        | 4 (12%)                       |
| Respiratory                                      | 25 (27%)                     | 5 (11%)                  | 7 (16%)                       | 6 (18%)                       |
| Sepsis/infection                                 | 20 (21%)                     | 5 (11%)                  | 11 (24%)                      | 2 (6%)                        |
| Trauma                                           | 2 (2%)                       | 2 (5%)                   | 2 (4%)                        | 2 (6%)                        |
| Other<sup>d</sup>                                | 16 (17%)                     | 7 (16%)                  | 5 (11%)                       | 5 (15%)                       |
| CPR in previous 24hr                              | 10 (11%)                     | 2 (5%)                   | 9 (20%)                       | 5 (15%)                       |
| Median GCS at ICU adm (IQR)                      | 11 (3-15)                    | 3.5 (3-10)               | 6 (3-11)                      | 6 (3-10)                      |
| Avg. APACHEII Score (IQR, range)                 | 26 (20-31, 8-46)             | 29 (25-33, 11-48)        | 27 (22-33, 13-48)             | 29 (22-33, 9-46) N=31         |
| Reported TBI                                     | 5 (5%)                       | 9 (20%)                  | 6 (13%)                       | 1 (3%)                        |
| Median LOS in ICU, days (range)                  | 5 (0-61)                     | 3 (0-42)                 | 2 (0-35)                      | 3 (0-54)                      |
| DCD ODO eligible                                 | 17 (18%)                     | 16 (36%)                 | 9 (20%)                       | 7 (21%)                       |
| DCD attempted                                    | 2 (2%)                       | 4 (9%)                   | 5 (11%)                       | 3 (9%)                        |
| DCD successful donors                            | 1 (1%)                       | 4 (9%)                   | 4 (9%)                        | 3 (9%)                        |
| Median Time to Death, min (IQR, range)           | 114<sup>e</sup> (58-340, 17-1873) | 34<sup>e</sup> (20-405, 6-3849) | 32<sup>e</sup> (13-78, 1-2473) | 125<sup>e</sup> (24-388, 1-9566) N=31 |
| Death within 0.5, 1, 2h (%)                       | 9%, 28%, 53%                 | 45%, 61%, 68%            | 49%, 67%, 80%                 | 32%, 42%, 48%                 |
| Patients from Canada                             | 72 (77%)                     | 35 (80%)                 | 35 (78%)                      | 22 (67%)                      |
| Patients from the Czech Rep.                     | 22 (23%)                     | 4 (9%)                   | 9 (20%)                       | 4 (12%)                       |
| Patients from the Neth.                          | 0 (0%)                       | 5 (11%)                  | 1 (2%)                        | 7 (21%)                       |
| Life Sustaining Measures                         |                               |                          |                               |                               |
| Receiving IMV                                    | 75 (80%)                     | 43 (98%)                 | 41 (91%)                      | 25 (76%)                      |
| Extubated as part of WLSM                        | 43 (46%)                     | 33 (75%)                 | 26 (58%)                      | 17 (51%)                      |
| On Vasopressors/Inotropes:                       | 76 (91%)                     | 23 (52%)                 | 29 (64%)                      | 20 (61%)                      |
| Receiving Sedation                               | 83 (88%)                     | 31 (70%)                 | 33 (73%)                      | 25 (76%)                      |
| Receiving Analgesia                              | 90 (96%)                     | 42 (95%)                 | 42 (93%)                      | 28 (85%)                      |

Supplemental Table E6: Demographics data for excluded patients, by reason of exclusion. a Data from 7 pediatric patients and 2 patients with missing data were not included in this analysis of excluded patients b Staggered withdrawals were defined as cases where additional measures of life sustaining measures were removed more than 10 minutes after the first act of withdrawal. c Other reasons for exclusion include protocol violations (N = 27, active pacemaker, no support withdrawn, imminent death not expected, neurological determination of death, no art line) and missing clinical data (N = 6). d Other reasons for admission include gastrointestinal bleeding, abdominal aortic aneurysm, multiple causes, hypovolemic shock, and multiorgan failure. e Double-sided Mann–Whitney U test comparing time to death distributions of derivation cohort and staggered patients gives p=0.0004. f Only 26 patients with recording problems and 13 patients with other reasons for exclusion had adjudicated times of death, so the clinically reported time of death was analyzed for these groups instead. Mann–Whitney U tests suggest time to death distributions were not different from the derivation cohort for these two groups (p>0.24). g Many patients flagged with poor quality waveforms had very low systolic and pulse pressures, and beats were filtered using the quality algorithms. Time to death was likely faster due to low blood pressures (Mann–Whitney U test gives p=0.008). GCS: scores on the Glasgow Coma Scale (range from 3 to 15, with lower scores indicating a reduced level of consciousness). Scores on the Acute Physiology and Chronic Health Evaluation (APACHE) II range from 0 to 71, with higher scores indicating more severe disease. TBI: traumatic brain injury. LOS: length of stay. DOD: donation after circulatory death. ODO: organ donation
organization. **IMV**: invasive mechanical ventilation, with a mandatory ventilation mode. **WLSM**: withdrawal of life sustaining measures. **STD**: standard deviation. **IQR**: interquartile range.
Supplemental Figure E4: Cumulative distribution functions for the prediction probabilities from the combined, calibrated model for the derivation (der) and validation (val) cohorts. These curves can be used to estimate the impact of a proposed probability cut-off. For example, approximately 50% of patients had probability scores >70% at 2 hours, and approximately 25% of patients had probabilities between 40% and 70% at 2 hours. Approximately 40% of patients had probability scores <20% at 30 minutes.
| Risk Category | Parameter or Metric                                      | Derivation | Validation |
|---------------|---------------------------------------------------------|------------|------------|
|               | Time (min)                                              | 30         | 60         | 120        | 30         | 60         | 120        |
| Low Risk      | number at low risk                                      | 131        | 117        | 47         | 55         | 50         | 22         |
|               | fraction at low risk                                    | 0.47       | 0.42       | 0.17       | 0.40       | 0.37       | 0.16       |
|               | number that died rapidly at low risk                    | 18         | 26         | 14         | 6          | 9          | 7          |
|               | fraction that died rapidly at low risk                  | 0.14       | 0.22       | 0.30       | 0.11       | 0.18       | 0.32       |
|               | low risk patients that died rapidly/all patients that died rapidly | 0.16 | 0.18 | 0.08 | 0.12 | 0.13 | 0.08 |
|               | low risk patients that died rapidly/low patients        | 0.06       | 0.09       | 0.05       | 0.04       | 0.07       | 0.05       |
| Medium Risk   | number at medium risk                                   | 109        | 45         | 76         | 56         | 14         | 27         |
|               | fraction at medium risk                                 | 0.39       | 0.16       | 0.27       | 0.41       | 0.10       | 0.20       |
|               | number that died rapidly at medium risk                 | 66         | 23         | 33         | 26         | 6          | 10         |
|               | fraction that died rapidly at medium risk               | 0.61       | 0.51       | 0.43       | 0.46       | 0.43       | 0.37       |
|               | medium risk patients that died rapidly/all patients that died rapidly | 0.59 | 0.16 | 0.19 | 0.53 | 0.09 | 0.11 |
|               | medium risk patients that died rapidly/medium patients  | 0.24       | 0.08       | 0.12       | 0.19       | 0.04       | 0.07       |
| High Risk     | number at high risk                                     | 38         | 116        | 155        | 25         | 72         | 87         |
|               | fraction at high risk                                   | 0.14       | 0.42       | 0.56       | 0.18       | 0.53       | 0.64       |
|               | number that died rapidly at high risk                   | 28         | 94         | 131        | 17         | 55         | 71         |
|               | fraction that died rapidly at high risk                 | 0.74       | 0.81       | 0.85       | 0.68       | 0.76       | 0.82       |
|               | high risk patients that died rapidly/all patients that died rapidly | 0.25 | 0.66 | 0.74 | 0.35 | 0.79 | 0.81 |
|               | high risk patients that died rapidly/medium patients    | 0.10       | 0.34       | 0.47       | 0.13       | 0.40       | 0.52       |

**Supplemental Table E7:** Statistics of patients falling into risk categories defined as low risk ($p \leq 0.38$), medium risk ($0.38 < p \leq 0.67$), or high risk ($p > 0.67$), as shown in Figure 2 of the main text and Supplementary Figure E4. Prediction probabilities were calculated using the combined calibrated model.
Supplemental Material References

1. Dhanani S, Hornby L, van Beinum A, et al.: Resumption of Cardiac Activity after Withdrawal of Life-Sustaining Measures. *N Engl J Med* 2021; 384:345–352

2. Task Force of The European Society of Cardiology and The North American Society of Pacing and Electrophysiology: Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. *European Heart Journal* 1996; 17:354–381

3. Ellis RJ, Zhu B, Koenig J, et al.: A careful look at ECG sampling frequency and R-peak interpolation on short-term measures of heart rate variability. *Physiol Meas* 2015; 36:1827–1852

4. 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

5. Bravi A, Longtin A, Seely AJ: Review and classification of variability analysis techniques with clinical applications. *Biomed Eng Online* 2011; 10:27

6. Seely AJ, Bravi A, Herry C, et al.: Do heart and respiratory rate variability improve prediction of extubation outcomes in critically ill patients? *Crit Care* 2014; 18:R65

7. Sun JX, Reisner AT, Mark RG: A signal abnormality index for arterial blood pressure waveforms. In: 2006 Computers in Cardiology. IEEE; 2006. p. 13–16.

8. Sacha J: Why should one normalize heart rate variability with respect to average heart rate. *Front Physiol* 2013; 4

9. Wright MN, Ziegler A: *ranger*: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R [Internet]. *J Stat Soft* 2017; 77[cited 2020 Dec 11] Available from: http://www.jstatsoft.org/v77/i01/

10. R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria; 2019.

11. Stekhoven DJ, Buhlmann P: MissForest--non-parametric missing value imputation for mixed-type data. *Bioinformatics* 2012; 28:112–118

12. Nembrini S, König IR, Wright MN: The revival of the Gini importance? *Bioinformatics* 2018; 34:3711–3718

13. Zadrozny B, Elkan C: Obtaining calibrated probability estimates from decision trees and naive Bayesian classifiers. *Icml* 2001; 1:609–616

14. Gupta C, Podkopaev A, Ramdas A: Distribution-free binary classification: prediction sets, confidence intervals and calibration [Internet]. *arXiv:2006.10564 [cs, math, stat]* 2021; [cited 2021 Sep 29] Available from: http://arxiv.org/abs/2006.10564
15. Brieva J, Coleman N, Lacey J, et al.: Prediction of Death in Less Than 60 Minutes Following Withdrawal of Cardiorespiratory Support in ICUs*: Critical Care Medicine 2013; 41:2677–2687

16. 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