Prospective Multicenter Observational Cohort Study on Time to Death in Potential Controlled Donation after Circulatory Death Donors - Development and External Validation of Prediction Models

Kotsopoulos, Angela; Vos, Piet; Witjes, Marloes; Volbeda, Meint; Franke, Hildegard; Epker, Jelle; Sonneveld, Hans; Simons, Koen; Bronkhorst, Ewald; Mullers, Ruud

Published in:
Transplantation

DOI:
10.1097/TP.0000000000004106

Prospective Multicenter Observational Cohort Study on Time to Death in Potential Controlled Donation After Circulatory Death Donors—Development and External Validation of Prediction Models: The DCD III Study

Angela Kotsopoulos, MD,[1] Piet Vos, CRN,[1] Marloes Witjes, PhD,[2] Meint Volbeda, MD,[3] Hildegard Franke, CRN,[3] Jelle Epker, MD, PhD,[4] Hans Sonneveld, MD,[5] Koen Simons, MD, PhD,[6] Ewald Bronkhorst, MStat,[7] Ruud Mullers, MStat, PhD,[8] Nichon Jansen, PhD,[9] Hans van der Hoeven, MD, PhD,[2] and Wilson F. Abdo, MD, PhD[2]

Background. Acceptance of organs from controlled donation after circulatory death (cDCD) donors depends on the time to circulatory death. Here we aimed to develop and externally validate prediction models for circulatory death within 1 or 2 h after withdrawal of life-sustaining treatment. **Methods.** In a multicenter, observational, prospective cohort study, we enrolled 409 potential cDCD donors. For model development, we applied the least absolute shrinkage and selection operator (LASSO) regression and machine learning–artificial intelligence analyses. Our LASSO models were validated using a previously published cDCD cohort. Additionally, we validated 3 existing prediction models using our data set. **Results.** For death within 1 and 2 h, the area under the curves (AUCs) of the LASSO models were 0.77 and 0.79, respectively, whereas for the artificial intelligence models, these were 0.79 and 0.81, respectively. We were able to identify 4% to 16% of the patients who would not die within these time frames with 100% accuracy. External validation showed that the discrimination of our models was good (AUCs 0.80 and 0.82, respectively), but they were not able to identify a subgroup with certain death after 1 to 2 h. Using our cohort to validate 3 previously published models showed AUCs ranging between 0.63 and 0.74. Calibration demonstrated that the models over- and underestimated the predicted probability of death. **Conclusions.** Our models showed a reasonable ability to predict circulatory death. External validation of our and 3 existing models illustrated that their predictive ability remained relatively stable. We accurately predicted a subset of patients who died after 1 to 2 h, preventing starting unnecessary donation preparations, which, however, need external validation in a prospective cohort.

(*Transplantation* 2022;106: 1844–1851).

Received 23 August 2021. Revision received 24 January 2022.

Accepted 26 January 2022.

[1] *Department of Intensive Care, Elisabeth TweeSteden Hospital, Tilburg, The Netherlands.*

[2] *Department of Intensive Care Medicine, Radboud University Medical Center, Nijmegen, The Netherlands.*

[3] *Department of Critical Care, University of Groningen, University Medical Center, Groningen, The Netherlands.*

[4] *Department of Intensive Care Medicine, Erasmus University Medical Center, Rotterdam, The Netherlands.*

[5] *Department of Intensive Care Medicine, Isala Hospital, Zwolle, The Netherlands.*

[6] *Department of Intensive Care Medicine, Jeroen Bosch Hospital, Hertogenbosch, The Netherlands.*

[7] *Department of Health Evidence, Radboud University Medical Center, Nijmegen, The Netherlands.*

[8] *Pipple, Eindhoven, The Netherlands.*

[9] *The Dutch Transplant Foundation, Leiden, The Netherlands.*

The authors declare no funding or conflicts of interest.

A.K., W.F.A., N.J., and P.V. participated in conceptualization of the study. P.V., M.W., M.V., H.F., J.E., H.S.., and K.S. participated in inclusion of patients and data entry in the Case Report Form in their hospital. A.K. was responsible for the curation of the data. A.K. did the first analysis of the data. E.B. evaluated the statistical analysis of the data and developed the least absolute shrinkage and selection operator prediction model. R.M. developed the artificial intelligence model. A.K., N.J., P.V., E.B., H.v.d.H., and W.F.A. participated in the study design, data interpretation, and the writing of the article. A.K., P.V., N.J., and W.F.A. had access to the raw data. A.K. and P.V. have verified the data used in the study. A.K., N.J., and W.F.A. had the final responsibility for the decision to submit for publication. All authors read, edited, and approved the final version of the article.

ClinicalTrials.gov; NCT04123275.

Supplemental Visual Abstract; http://links.lww.com/TP/C382.

Supplemental digital content (SDC) is available for this article. Direct URL citations appear in the printed text, and links to the digital files are provided in the HTML text of this article on the journal's Web site (www.transplantjournal.com).

Correspondence: Angela Kotsopoulos, MD, Department of Intensive Care, Elisabeth TweeSteden Hospital, Hilvarenbeekseweg 60, 5000LC, Tilburg, The Netherlands. (a.kotsopoulos@etz.nl).

ISSN: 0041-1337/20/1069-1844

DOI: 10.1097/TP.0000000000004106

INTRODUCTION

The number of patients donating their organs after controlled donation after circulatory death (cDCD) is increasingly leading to a large pool of potential organ donors apart from brain dead donors.[1] A limiting factor in cDCD is that circulatory death must occur within a defined time frame after withdrawal of life-sustaining treatment (WLST) for organs to be medically suitable for transplantation. Up to 1 out of 4 potential cDCD donors will not die within this defined time frame and thus will not be able to donate their organs.[2] This leads to disappointment within the families of these donors and the patients on the waiting list and to inefficient utilization of hospital facilities and wasted healthcare costs.[3] Therefore, in several countries with DCD programs, for example, United Kingdom, the Netherlands, Australia, China, and the United States, efforts have been made to develop models predicting time to death.[2,4-13] As clinical practices and donation policies differ considerably within countries, external validation of a prediction model is a necessary step before implementing a prediction model in daily practice. External validation provides insight into the extent of the reproducibility in a new cohort, as prediction models usually perform less in another setting[14]; however, this step is often omitted.[2,9,15]

The predictive performance of previous models showed that 10% to 18% of the patients who would die within the necessary time frame would be misclassified, which limits implementation with the current large shortage of transplantable organs.[4]

However, if a model is highly accurate in detecting which patients would certainly not die within the necessary time frame, predicting even a small percentage of such patients could be beneficial, as it could be used to prevent starting an organ donation procedure. The approach to predict which patients will die beyond the time frame with 100% accuracy for such a purpose has not been done before.

In this multicenter prospective observational study, including consecutive potential cDCD donors, our primary aim was to develop a prediction model for death in 1 or 2 h. We also aimed to externally validate our model and previously published models using our data sets. Prediction model development studies rarely assessed the impact of WLST practices and palliative medication use and dosages on time to death.[5] Therefore, we also aimed to assess the impact of end-of-life care on time to death.

MATERIALS AND METHODS

Study Population

We prospectively enrolled consecutive patients, meeting the cDCD criteria used by the Dutch Transplant Foundation (**Table S1, SDC,** http://links.lww.com/TP/C381), from 3 university and 3 teaching hospitals in the Netherlands.[16] Patients were excluded if they were <18 or >75 y of age, were not mechanically ventilated, or were clinically brain dead at the time of assessment.

Study Design and Data Source

We performed a multicenter, observational, prospective cohort study of potential cDCD patients and collected data between June 2015 and July 2018. The study design has been previously published and registered at clinicaltrials.gov (NCT04123275).[17] In short, hemodynamic, pulmonary, and neurological features were assessed no later than 30 min before WLST. Dosages of analgesics and opioids were recorded from 30 min before WLST until circulatory death. Different types of opioids were converted into morphine equivalent doses (15 μg of intravenous fentanyl or remifentanil or 2 μg of intravenous sufentanil were equivalent to 1 mg of intravenous morphine).[18,19]

In each hospital, local researchers prospectively collected all data. Data were stored in an internet-based electronic Case Report Form supported by Research Manager (ISO/IEC 27001 certified). Before patient enrollment, all local researchers received on-site personal training regarding the completion of the electronic Case Report Form. The lead research team performed regular site visits and randomly assessed data entry for accuracy and completeness.

Results of this study are reported according to the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis Or Diagnosis and Strengthening the Reporting of Observational Studies in Epidemiology statements.[20,21]

The Medical Ethics Committee Brabant, the Netherlands, has approved the study protocol (NW2014-36). The Medical Ethics Committee boards of all participating hospitals assessed and consented with the study protocol.

External Validation or Our Model and Previously Published Models

In addition to internal validation, we also sought to externally validate our models. External validation is the process of evaluating a prediction model in a different cohort of patients. The prediction model tested can be newly developed or already existing. The same applies to the validation cohort, as long as it is independent of the cohort in which the prediction model was developed. As such, first, we contacted and asked the authors of previously published studies on prediction models, including a potential cDCD population, if we could obtain their data sets to externally validate the DCD III study models. This resulted in 3 data sets, 2 from the Netherlands and 1 from the United Kingdom (**Table S2, SDC,** http://links.lww.com/TP/C381).[5,11,15] Variables included in the data sets of the 3 obtained cohorts had only a limited overlap with our developed prediction models ("least absolute shrinkage and selection operator [LASSO] models") and could therefore not be used to externally validate our models. In addition, we previously collected data on a cDCD cohort of 92 patients ("Elisabeth-TweeSteden Hospital [ETZ] cohort").[2] The data of the ETZ cohort were collected before the start of this DCD III study, and thus, there were no overlapping patients between the ETZ and DCD III cohorts. As such, we used the data set of the ETZ cohort to externally validate our DCD III models.

As a second step, we identified 8 previously published prediction models that we selected to validate externally using either the DCD III or ETZ cohorts (**Table S3, SDC,** http://links.lww.com/TP/C381). We contacted the corresponding authors of these previous studies to obtain their data for external validation. Five models were excluded because not all the variables used in their prediction models were registered in our validation cohort (DCD III cohort).[4,5,7,12,13] The remaining models, including patients from the Netherlands (de Groot) and the United Kingdom (Davila and Suntharalingam), were validated using the data sets of the DCD III and ETZ cohorts.[6,9,11] **Table S4**

SDC, http://links.lww.com/TP/C381) shows the characteristics of the prediction models.

Statistical Analysis

Categorical variables were expressed as a percentage and continuous variables as mean ± standard deviation or median and interquartile range, depending on homogeneity. Univariable logistic regression analysis was performed to explore the relationship between a set of variables on the binary outcome (death within 1 h, yes or no). Our database had <3.4% missing data encountered in only 2 variables.

We analyzed the probability to predict death within and beyond 1 or 2 h, as these are the thresholds that are mostly used internationally for cDCD donation. For the development of the prediction models, we applied 2 different data analytic methods. First, the LASSO method was used for optimal variable selection and development of the multivariable logistic regression model on 80% of the patient population (development sample). The predictive performance of this model was internally validated on the remaining 20% of the population (validation sample). We aimed at an SE of 0.04 for the area under the curve (AUC), with an expected mortality of approximately 50% within 1 h and 65% within 2 h after WLST, and an anticipated AUC of 0.84.[2,5,7,11] The discriminatory ability of the prediction model was assessed using the AUC. The AUC as found in the 20% testing part of the sample is presented. The sample size for this study was calculated (based on the LASSO method as the primary analysis) at 400 patients.

To identify the patients dying beyond 1 or 2 h, we used the same models as for death within 1 or 2 h. This approach was analyzed because if we were able to predict, 100% accurately, death after 1 or 2 h, this could be applicable in the daily clinical practice even if the model only predicted a small percentage of patients accurately. We determined the optimal threshold as the value (computed by the linear predictors of these models when applied on the validation data set) with the highest negative predictive value. A more detailed description of the statistical analysis can be found in the previously published study protocol.[17]

Second, for the development of the machine learning models, a light gradient boosting machine (LightGBM) algorithm was used for both variable selection and model development. To overcome the limited amount of data available, 10-fold cross-validation was performed on the complete data set and repeated 50 times to assure stability of the model.[22] The performance of the models was assessed by looking at the distribution of the AUC for the different folds. Within the model, we used a threshold on the likelihood of dying within or beyond 1 or 2 h. Recall is the number of donors identified for dying beyond 1 or 2 h. By using a precision–recall curve and threshold curve, we can adjust the model in a way that precision is higher, but recall is lower, enabling us to select a threshold at which the precision is high, but the recall is still relevant.

External validation of the LASSO models was performed by calculating the probability of death and thus the AUC, computed after application of the linear predictor on the data set of the ETZ cohort taking the outcome status (death) into account. The AUC determines discrimination of the models and refers to the ability to distinguish between circulatory death within 1 h (or 2 h) and longer. Calibration refers to the concordance between the absolute predicted probability and observed death and is graphically displayed in calibration plots.[23] Over- or underestimation of predicted probabilities is often encountered in the new cohort. Recalibration of the models can overcome this problem. For that purpose, we used the offset procedure to update the intercept of the linear predictors and adjust the regression coefficients for optimism (shrinkage).[23,24] We used the same approach for external validation of the existing models on the data set of our cohort.

Statistical analyses were performed using IBM SPSS, version 24, the R software (R Project for Statistical Computing, version Microsoft R Open 3.6.1), and Python version 3.8.5, 64-bit LightGBM package, version 3.1.1.

RESULTS

Patient Characteristics

In a period of 40 mo, 425 consecutive potential cDCD donors were evaluated. Sixteen patients were suspected to be brain dead and were excluded, leading to inclusion of 409 patients, who formed the DCD III cohort.

Circulatory death within 1 or 2 h after WLST occurred in 55% and 63% of patients, respectively. The median time to death was 43 min (range, 17–432). In 9% of cases, death occurred after >24 h after WLST.

Table 1 shows the clinical characteristics of the patients. There were no differences in baseline characteristics between the patients dying within and after 1 h. Table 2 shows main collected characteristics and measures during the intensive care unit (ICU) and WLST process. More than 70% of patients in both groups experienced secondary neurological injury during their ICU stay. Patients dying within 1 h had significantly lower Glasgow Coma Scale (GCS)/motor scores, more often absent brain stem reflexes, higher ventilatory settings and oxygenation index, and spent less time spent on the ventilator than patients dying after 1 h.

Predicting Death Within 1 or 2 h

Table 3 shows the linear predictors (lp) and AUCs using LASSO for death within 1 or 2 h. The resulting models included easy-to-assess bedside clinical features incorporating several brain stem reflexes, ventilator triggering, (items of) the GCS, administration of vasopressors, the oxygenation index, and subarachnoid hemorrhage, combined in a mathematical equation known as the lp, whose value can predict the probability of death with the following equation: $\text{lp}=\frac{e^{lp}}{1+e^{lp}}$. In the validation sample, 3 patients (4%) had the highest predicted probability of death (77%). The group with the lowest observed death included 12 patients (16%) and had a predicted probability of death of 32% within 1 h. The model had a positive predictive value of 66%. The prediction model was overoptimistic in the lower and intermediate observed death groups and underoptimistic in the highest observed death group, meaning that potential donors are missed in the latter and unnecessary procedures will be started in the former (**Figure S1, SDC,** http://links.lww.com/TP/C381). Similar calculations were performed for death within 2 h with comparable results (positive predictive value of 69%).

The variables selected after applying LightGBM are also shown in Table 3. The variable importance determined by

TABLE 1.

Baseline characteristics of 409 potential cDCD patients dying within or after 1 h after WLST

Variable	Death within 1 h n = 224 (55%)	Death after 1 hZn = 185 (45%)	OR	95% CI	*P*
Time to death, min, (range)	18 (11–26)	476 (177–1202)			
Age, mean (±SD), y	56.7 (±14.9)	58.6 (±11.7)	0.99	0.97-1.00	0.170
Male,	145 (65%)	104 (56%)	1.43	0.95-2.13	0.079
BMI, mean (±SD)	26.1 (±4.7)	26.2 (±4.9)	1.01	0.96-1.04	0.796
APACHE IV, mean (±SD)	94.4 (±25.9)	91.9 (±26)	1.00	0.99-1.01	0.298
Hemicraniectomy	22 (10%)	26 (14%)	0.18	0.66-0.36	0.121
Admission diagnosis					
Anoxic encephalopathy	53 (24%)	50 (27%)	0.88	0.56-1.38	0.581
Traumatic brain injury	51 (23%)	43 (23%)	1.06	0.62-1.79	0.834
Subarachnoid hemorrhage	51 (23%)	33 (18%)	1.38	0.78-2.28	0.257
Intracranial hemorrhage	36 (16%)	36 (19%)	0.93	0.53-1.64	0.817
Cerebrovascular accident	18 (8%)	15 (8%)	1.49	0.65-3.42	0.342
Respiratory	6 (3%)	4 (2%)	0.72	0.19-2.71	0.638
Other[a]	9 (4%)	4 (2%)	2.12	0.61-7.33	0.234

Categorical variable are described as number (%) and continuous variables as mean (±SD).
P value was calculated using univariable logistic regression analysis.
[a]Other includes encephalitis, Huntington disease and trauma, meningitis, intracerebral abscess, aspiration pneumonia complicating minor trauma, complication after meningioma resection, methanol intoxication, and refractory epilepsy.
APACHE, acute physiology age chronic health evaluation; BMI, body mass index; cDCD, controlled donation after circulatory death; CI, confidence interval; OR, odds ratio; SD, standard deviation; WLST, withdrawal of life-sustaining treatment.

the permutation and gain method is shown in **Figure S2 (SDC,** http://links.lww.com/TP/C381). The AUCs of the artificial intelligence (AI) models were comparable with the LASSO models. The prediction variables included in the LASSO and AI models also overlapped. The AUCs (range, 0.73–0.88) and variables included in the 8 previously published prediction models are summarized in **Table S3 (SDC,** http://links.lww.com/TP/C381).

Predicting Death Beyond the 1 or 2 h Time Frame With 100% Accuracy

Similar calculations as mentioned for predicting death within 1 h were performed to predict death beyond 1 h with a 100% accuracy, meaning that if the model predicts that a patient does not die within 1 or 2 h, no false positives should be present.

Using a threshold that 100% correctly classified patients dying after 1 h, we found that the LASSO model classified 12 of 75 patients (16%) who died after 1 h with a 100% accuracy (**Table S5, SDC,** http://links.lww.com/TP/C381). Likewise, predicting death after 2 h led to a threshold that predicted 10 of 75 patients (13%) of the validation group without any false positives.

Similarly, we calculated the thresholds for 100% accuracy (no false positives) for the AI models, which led to the correct classification of 9 of 185 patients (5%) for the death after 1 h and 6 of 153 patients (4%) for death after 2 h (Table 4).

External Validation of 3 Previously Published Models and the Current LASSO Models From the DCD III Study Using the ETZ Cohorts

The predictive performances expressed as the AUC and calibration are shown in Table 5 and **Figures S3 and S4, (SDC,** http://links.lww.com/TP/C381). All models showed a modest discrimination (AUC range, 0.63–0.86) when validated in the data sets of the DCD III and ETZ cohorts. Calibration plots showed that most models slightly overestimated the probabilities of death in those patients who died within the necessary time frame, whereas our LASSO models overestimated the probability of death in those who did not die within the time frame (see **Figures S3 and S4, SDC,** http://links.lww.com/TP/C381). Recalibration of the models was therefore performed. In our ETZ validation cohort, we were not able to replicate the finding of identifying a subgroup with certain death after 1 or 2 h.

Because WLST practices could influence timing of death and are only rarely taken into account in previous studies, we also sought to assess the effect of WLST on timing of death.

Medication

Seventy-seven percent of patients used some type of sedation or analgesia, as a continuous infusion, during the course of WLST (Table 2). The mean midazolam dose was significantly higher before and after WLST in the <1 h group (**Figure S5, SDC,** http://links.lww.com/TP/C381). There were no significant differences in the doses of morphine equivalents administered before and after WLST in both groups (**Figure S5, SDC,** http://links.lww.com/TP/C381). Less than 10% of all patients received a bolus of midazolam (maximum dose 10 mg) or opioids before or after WLST.

WLST Practice

In all patients, supportive care was withdrawn simultaneously, meaning that together with the endotracheal tube, all medication (except medication for palliative care), tube feeding, and fluid administration were discontinued. In 21% of all patients, external ventricular drainage was present, which was removed or closed in almost all cases at the time of WLST.

All patients were either extubated (98.5%) or disconnected from the ventilator (remaining 1.5%). Nursing care, such as secretion clearance or lateral tilt position variation after WLST, was applied significantly more frequently in the >1 h group (both $P < 0.001$), however, in a small number of patients (9.5%).

TABLE 2.

Clinical parameters and medication use just before and throughout WLST of 409 potential cDCD patients dying within or after 1 h after WLST

Variable	Death within 1 h n = 224 (55%)	Death after 1 h n = 185 (45%)	OR	95% CI	P[a]
Neurological examination					
Absent pupillary reflex bilateral	148 (66%)	80 (43%)	2.56	1.71-3.82	<0.001
Absent corneal reflex bilateral	159 (71%)	76 (41%)	3.51	2.33-5.29	<0.001
Absent cough reflex	139 (62%)	58 (31%)	3.58	2.37-5.40	<0.001
GCS ≤4[b]	202 (92%)	120 (66%)	5.70	3.21-10.1	<0.001
Absent motor response or extensor (M1–M2)[b]	206 (94%)	134 (74%)	5.61	2.73-9.74	<0.001
Ventilation					
Controlled mode of ventilation	152 (68%)	73 (39%)	3.23	2.15-4.85	<0.001
Triggering mechanical ventilation	117 (52%)	153 (83%)	0.22	0.14-0.35	<0.001
Mechanical ventilation time, h, mean (±SD)	80 (±96)	128 (±151)	0.996	0.995-0.998	<0.001
Mean FiO_2, mean (±SD)	0.42 (±0.17)	0.36 (±0.13)	1.02	1.01-1.04	<0.001
PEEP, cmH_2O, mean (±SD)	7.4 (2.5)	6.7 (±2.2)	1.12	1.03-1.22	0.007
Peak inspiratory pressure, cmH_2O, mean (±SD)	19.6 (±6.2)	17.9 (±4.8)	1.01	1.06-1.14	<0.001
Pao_2, kPa, mean (±SD)	15.0 (±7.7)	13.6 (±5.7)	1.03	0.99-1.07	0.077
OI,[c] mean (±SD)	6.3 (±5.4)	4.8 (±3.4)	1.09	1.04-1.16	<0.001
Hemodynamics					
Mean arterial pressure, mm Hg, mean (±SD)	111 (±4)	90 (±3)	0.99	0.98-1.00	0.320
Heart rate per minute, mean (±SD)	89 (±26.6)	87 (±24.6)	1.01	0.99-1.01	0.214
Secondary neurological injury					
Total	175 (78%)	134 (72%)	1.35	0.86-2.13	0.183
Vasospasm	6 (2%)	7 (4%)	0.70	0.23-2.12	0.528
Recurrent hemorrhage	24 (11%)	18 (10%)	1.11	0.58-2.12	0.744
Cerebral infarction	33 (15%)	42 (23%)	0.58	0.35-0.97	0.039
Cerebral edema	48 (21%)	29 (16%)	1.46	0.88-2.44	0.140
Epilepsy	31 (14%)	28 (15%)	0.90	0.51-1.56	0.711
Hydrocephalus	56 (25%)	37 (20%)	1.33	0.83-2.13	0.231
Cerebral herniation	70 (31%)	39 (21%)	1.70	1.08-2.67	0.021
Medication use					
Norepinephrine use before WLST	64 (28%)	17 (9%)	3.95	2.22-7.03	<0.001
Norepinephrine dose before WLST, ug/kg/min, mean (±SD)	0.18 (±0.21)	0.11 (±0.12)	15.0	0.21-1030	0.209
Midazolam use before WLST	128 (57%)	120 (65%)	0.85	0.69-1.03	0.112
Midazolam use after WLST	129 (57%)	126 (68%)	0.63	0.42-0.95	0.029
Midazolam dose before WLST, mg/h, mean (±SD)	9.8 (±8.7)	7.4 (±6.3)	1.04	1.01-1.08	0.022
Midazolam dose after WLST, mg/h, mean (±SD)	9.8 (±8.7)	7.7 (±6.3)	1.03	1.01-1.07	0.040
Propofol use before WLST	46 (21%)	23 (12%)	1.16	1.01-1.33	0.031
Propofol use after WLST	47 (21%)	23 (12%)	1.87	1.08-3.21	0.024
Propofol dose before WLST, mg/h, mean (±SD)	188 (±109)	201 (±129)	0.99	0.99-1.00	0.645
Propofol dose after WLST, mg/h, mean (±SD)	186 (±109)	204 (±132)	0.99	0.99-1.00	0.560
Morphine equivalents use before WLST	157 (70%)	140 (75%)	0.75	0.48-1.17	0.208
Morphine equivalent use after WLST	157 (70%)	147 (79%)	0.60	0.38-0.95	0.032
Morphine equivalent dose before WLST, mg/h, mean (±SD)	10.8 (±23)	7.3 (±13.9)	1.01	0.99-1.02	0.148
Morphine equivalent dose after WLST, mg/h, mean (±SD)	10.9 (±24)	7.8 (±13.6)	1.01	0.99-1.02	0.194

[a] *P* value was calculated using univariable logistic regression analysis.
[b] Analysis as a dichotomous variable with GCS ≤4 vs GCS ≥ and M1/M2 vs ≥M3.
[c] Mechanical ventilator settings and arterial blood gas analysis were integrated into the OI. OI is a continuous variable and is calculated as mean airway pressure × FiO_2 × 100/Pao_2 in kPa; mean airway pressure = peak inspiratory pressure + PEEP/2). An OI >4.2 was defined as elevated.
Categorical variables are described as numbers (%) and continuous variables as mean (±SD).
cDCD, controlled donation after circulatory death; CI, confidence interval; GCS, Glasgow Coma Scale; FiO_2, fraction of inspired oxygen; M1, no response to stimulus; M2, extension response in response to pain; M3, Flexion in response to pain; N, number; OI, oxygenation index; OR, odds ratio; PEEP, positive end-expiratory pressure; Pao_2, Po_2 in arterial blood; SD, standard deviation; WLST, withdrawal of life-sustaining treatment.

DISCUSSION

This is the largest multicenter prospective cohort study of potential cDCD donors on time to circulatory death. Our study found that corneal and cough reflex, ventilator triggering, the motor score, administration of vasopressors, and the GCS are the main predictors for circulatory death using a regression analysis with an AUC of the model of 0.77. This improved only slightly using AI–machine learning modeling, which is most likely because of the limited number of patients for AI modeling. Using such a model

TABLE 3.

AUC of predicted probabilities for death for different time frames of circulatory death for the LASSO and AI models

	AUC	Features per model
LASSO for death within 1 h after WLST	0.77[a]	lp[b]: 1.05−0.32 × corneal reflex present (yes = 1, no = 0) −0.33 × cough reflex present (yes = 1, no = 0) + 0.01 × motor score (M1 or M2 = 1, M3 or more = 0) + 0.11 × norepinephrine given (yes = 1, no = 0) − 0.21 × triggering mechanical ventilator (yes = 1, no = 0) −0.09 × GCS
LASSO for death within 2 h after WLST	0.79[a]	lp[c]: 1.18−0.26 × pupillary reflex present (yes = 1, no = 0) −0.36 × corneal reflex present (yes = 1, no = 0) −0.54 × cough reflex present (yes = 1, no = 0) + 0.57 × motor score (M1 or M2 = 1, M3 or more = 0) + 0.02 × OI + 0.18 × norepinephrine given (yes = 1, no = 0) −0.48 × triggering mechanical ventilator (yes = 1, no = 0) + 0.11 × SAH (yes = 1, no = 0) −0.11 × GCS
AI for death within 1 h[d]	0.79	APACHE II × APACHE IV score, ventilatory hours, length of ICU stay (in hours), triggering the mechanical ventilator, FiO_2, the GCS before WLST, and presence of cough reflex
AI for death within 2 h[d]	0.81	Equal variables as for death within 1 h

OI is a continuous variable and is calculated as mean airway pressure × FiO_2 × 100/Pao_2 in kPa; mean airway pressure = peak inspiratory pressure + positive end-expiratory pressure/2.
[a]This is the AUC as found in the validation sample, and it represents the model fit of the development sample.
[b]This is the lp derived from the development sample.
[c]This is the lp derived from the development sample.
[d]All AI models included the same variables.
AI, artificial intelligence; APACHE, acute physiology age chronic health evaluation; AUC, area under the receiver operating characteristic curve; FiO_2, fraction of inspired oxygen; GCS, Glasgow Coma Scale; ICU, intensive care unit; LASSO, least absolute shrinkage and selection operator; lp, linear predictor; M1, no response to stimulus; M2, extension response in response to pain; M3, Flexion in response to pain; OI, oxygenation index; SAH, subarachnoid hemorrhage; WLST, withdrawal of life-sustaining treatment.

would mean that nearly 23% of the patients would be wrongly classified as not dying within 1 or 2 h. With the large shortage of transplantable organs, operationalizing such models in clinical practice would mean that an unacceptable high percentage of potential organ donors would be missed. Previous models had similar issues as shown in **Table S3 (SDC,** http://links.lww.com/TP/C381).

As a second step, we calculated whether we could predict death after the 1 to 2 h time frame with 100% accuracy in order not to miss any donors. We were able to correctly identify 4% to 16% of the patients who would die beyond the 1 to 2 h time frame. Although this seems small, the high accuracy means that it could be used in clinical practice. Predicting which patients would not die within the necessary time frame, cases that should not be prepared for a donor procedure can be identified. This would prevent using precious resources, such as bed occupancy on the ICU, booking an operating theater, having a procurement team and transplantation team standby, use of ancillary diagnostics to test suitability of the organs and matching with acceptors, selecting potential patients on the transplantation waiting list, etc. Costs of starting an unsuccessful organ donation procedure have never been fully calculated, which we estimate would be at approximately €20.000 per patient in the Netherlands. Importantly, being able to accurately select which patients will die beyond the necessary time frame also prevents disappointment within the family of these patients and the ICU team when an organ donation procedure fails because of timing of death.

Our final step was external validation. We encountered several obstacles in that process. Previous models used different variables, which prevented us from performing extensive external validation. For the same reason,

TABLE 4.

Precision–recall trade-off for different thresholds for death after 1 or 2 h of the AI prediction model

Threshold	Total deaths after 2 h	True positive	False positive	Precision	Recall
0.50	153	89	48	0.65	0.58
0.80	153	21	6	0.78	0.14
0.85	153	6	2	0.75	0.04
0.88[a]	153	6	0	1	0.04
0.9	153	5	0	1	0.03
0.95	153	1		1	0.01
Threshold	Total deaths after 1 h	True positive	False positive	Precision	Recall
0.50	185	122	49	0.71	0.66
0.80	185	47	6	0.89	0.25
0.85	185	30	4	0.88	0.16
0.88	185	17	2	0.89	0.09
0.9[b]	185	9	0	1	0.05
0.95	185	0	0	0	0

[a]For each patient for whom the probability of not deceasing within 2 h was >0.88, it was 100% correctly predicted that he did not decease within 2 h. Using this threshold, the model would have predicted 6 patients of 153 who did not die within 2 h.
[b]For each patient for whom the probability of not deceasing within 1 h was >0.9, it was 100% correctly predicted that he did not decease within 1 h. Using this threshold, the model would have predicted 9 patients of 185 who did not die within 1 h.
AI, artificial intelligence.

TABLE 5.

External validation of 3 existing models for death within 1 h of the DCD III cohort and ETZ cohort and LASSO models for death within 1 and 2 h of the ETZ cohort

	Published by the author			Cohorts for external validation					
				DCD III cohort (n = 409)			ETZ cohort (n = 92)		
		95% CI			95% CI			95% CI	
Models	AUC	Lower	Upper	AUC	Lower	Upper	AUC	Lower	Upper
de Groot et al[9,a] (n = 82)	0.77	0.69	0.90	0.74	0.70	0.79	0.86	0.77	0.95
Davila et al[6,b] (n = 178)	0.83	0.76	0.90	0.70	0.65	0.75	0.80	0.70	0.90
Suntharalingam et al[11,c] (n = 191)				0.63	0.58	0.68	0.63		
LASSO 1 h (n = 75)	0.77						0.80	0.69	0.91
LASSO 2 h (n = 75)	0.79						0.82	0.73	0.92

Analysis method: The AUC is based on the predicted probabilities of death within 1 h for each patient in the external validation data sets of the cohorts, calculated by the linear predictor per model as provided by the authors.

[a]Linear predictor = −2.52 + 1.54 × absent corneal reflex (yes = 1, no = 0) + 1.08 × absent cough reflex (yes = 1, no = 0) + 1.18 × extensor or absent motor response (yes = 1, no = 0) + 0.13 × oxygenation index × 0.13.

[b]Linear predictor = 2.4 + 1.6 × (0 = no inotropes;1 = yes inotropes) − 1.27 × (0 = ≤40 y, 1 = >41 y) − 1.92 × (0 = no gag/cough;1 = gag/cough present).

[c]Suntharalingam et al[11] used Cox regression analysis. The baseline cumulative hazard for survival at 1 h was 0.44. The PI is 0.44 + (age 30–40 y × LN [0.70])+ (age 41–50 y × LN [0.46]) + (age >50 y × LN [0.37]) + (pressure support ventilation × LN [1.67]) + (F_{IO_2} × LN [1.012]). The baseline cumulative hazard of the DCD III cohort was 0.45.

AUC, area under the curve; CI, confidence interval; DCD, donation after circulatory death; ETZ, Elisabeth-TweeSteden Hospital; LASSO, least absolute shrinkage and selection operator; LN, natural logarithm; N, number of patients included in each cohort; PI, prognostic index.

we were only able to validate our LASSO models in a small retrospective data set. The size of the data set likely also influenced the results of the external validation of the LASSO models. In the external cohort, an equivalent subset of patients who died after the 1 to 2 h time frame could not be addressed, as the LASSO models performed less accurately. We anticipated this outcome, as a less accurate performance is often encountered in a different patient population. Almost all prediction studies have focused trying to predict correctly the majority of patients, and unfortunately, this has not resulted in tools that can be implemented. Our data show that it could be worthwhile to develop tools to predict a minority of patients with high accuracy instead. With regard to our models, these need external validation, ideally using prospectively collected data, before their true potential can be valued.[23]

Before implementing prediction models, clinicians should be aware of the characteristics of their own patient population in relation to those in the models. Heterogeneity in clinical practices, for example, in timing of prognostication or end-of-life care, may influence patient selection and thus variables included in the models. A model that is not recalibrated in the new setting can result in over- or underestimation of predicted probabilities of death, hampering its use in clinical decision making.

With increasing share of cDCD donors, the effort of developing prediction models is clearly visible in the literature. Eighty percent of such studies have been published in the past decade; however, despite these efforts, prediction models have not been implemented in clinical practice, mainly because of their modest predictive ability. The accuracy data of our models to predict which patients would die after 1 to 2 h show that this approach has potential and needs confirmation by other cohorts.

As mentioned in earlier reports, we found that sedation and analgesia dosages did not influence timing of death.[2,18,19,25,26] We could not support the finding of previous studies in a general ICU population in which lower doses of morphine were associated with more severe neurological damage.[27] All hospitals adhered to the guidelines published by the Dutch Intensive Care Society on withholding and withdrawing of life-sustaining treatment and palliative care of ICU patients.[28] Therapeutic treatment provided (noncomfort medications such as vasopressors, inotropes, antibiotics, intravenous fluids, [par]enteral feeding and endotracheal tube) was withdrawn simultaneously in almost all cases. Synchronous withdrawal of all treatments was also the practice in previous studies on cDCD patients.[2,4-6,11]

Our study also has some limitations. Prediction models are statistical models built on variables collected from patients in a specific setting. Therefore, the discriminatory ability of some variables can vary in different cohorts; however, in contrast to several earlier reports, we only included patients fulfilling organ donation criteria, which makes our sample more generalizable than many of the earlier prediction studies. Additionally, our data set is the largest prospective cohort. We included patients from different types of hospitals (university medical centers as well as teaching hospitals with differences in their patient focus). We did not assess the clinical opinion of the medical team. Previous studies showed that the clinical opinion alone was not accurate enough.[4,5] We also did not assess laboratory results.[6] Although severe abnormalities in laboratory findings could be related to timing of death, they also mirror a poor clinical organ condition. Such patients would most likely not fulfill the cDCD criteria. As such, we do not expect that laboratory criteria would have changed the prediction model significantly.

In summary, this large, multicenter, prospective cohort of potential cDCD donors found it easy to assess clinical features and to estimate time to death after WLST. We were able to reliably predict a small percentage of patients who would die beyond 1 or 2 h. Selecting such patients not to enter the organ donation process could prevent unnecessary costs, use of precious resources, and disappointment within the donor and acceptor families; however, further evaluation in a large independent cohort is needed.

External validation of previously published models showed that all have modest accuracy, hampering their use in daily clinical practice. Based on our data, WLST practices are not associated with timing of death and as such have no influence on donor potential.

ACKNOWLEDGMENTS

The authors are grateful to all intensivists, trainees, and nursing staff who collected the data. They also acknowledge the support and advice of the Dutch Transplant Foundation staff. They also want to thank Dr T. Wind, M. Nijhoff, Dr C. Moers, and Dr C. Watson for providing their data sets for the realization of external validation. They gratefully thank Prof Dr F.M.P. van Haren (Director of Intensive Care at the St George Hospital in Sydney and Prof at the Australian National University in Canberra) for providing insight regarding the process, legal and ethical issues, best practices, and care of the patient and family in organ donation in Australian and New Zealand Intensive Cares and also Dr Dale Gardiner (Associate Medical Director, Deceased Organ Donation, NHSBT, and Intensive Care Consultant at Nottingham University Hospital) for giving information regarding time to death and the use of prediction models in the United Kingdom.

REFERENCES

1. Haase-Kromwijk B. *NTS Jaarverslag 2017- Nederlandse Transplantatie Stichting*. Dutch Transplant Foundation. 2018. Available at https://www.transplantatiestichting.nl/files/Jaarverslag/nts-jaarverslag-2018.pdf?02b0bab58b. Accessed October 26, 2021.
2. 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.
3. 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.
4. 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.
5. 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.
6. Davila D, Ciria R, Jassem W, et al. Prediction models of donor arrest and graft utilization in liver transplantation from Maastricht-3 donors after circulatory death. *Am J Transplant*. 2012;12:3414–3424.
7. 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.
8. Pine JK, Goldsmith PJ, Ridgway DM, et al. Predicting donor asystole following withdrawal of treatment in donation after cardiac death. *Transplant Proc*. 2010;42:3949–3950.
9. 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.
10. Rabinstein AA, Yee AH, Mandrekar J, et al. Prediction of potential for organ donation after cardiac death in patients in neurocritical state: a prospective observational study. *Lancet Neurol*. 2012;11:414–419.
11. 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.
12. Xu G, Guo Z, Liang W, et al. Prediction of potential for organ donation after circulatory death in neurocritical patients. *J Heart Lung Transplant*. 2018;37:358–364.
13. Okahara S, Snell GI, McDonald M, et al. Improving the predictability of time to death in controlled donation after circulatory death lung donors. *Transpl Int*. 2021;34:906–915.
14. Ramspek CL, Jager KJ, Dekker FW, et al. External validation of prognostic models: what, why, how, when and where? *Clin Kidney J*. 2021;14:49–58.
15. 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*. 2020;105:1311–1316.
16. Dutch Transplant Foundation. *Modelprotocol. Post-mortale orgaan en weefsel donatie*. 2019. Available at https://www.transplantatiestichting.nl/medisch-professionals/modelprotocoll. Accessed October 1, 2021.
17. Kotsopoulos AMM, Vos P, Jansen NE, et al. Prediction model for timing of death in potential donors after circulatory death (DCD III): protocol for a multicenter prospective observational cohort study. *JMIR Res Protoc*. 2020;9:e16733.
18. Chan JD, Treece PD, Engelberg RA, et al. Narcotic and benzodiazepine use after withdrawal of life support: association with time to death? *Chest*. 2004;126:286–293.
19. Mazer MA, Alligood CM, Wu Q. The infusion of opioids during terminal withdrawal of mechanical ventilation in the medical intensive care unit. *J Pain Symptom Manage*. 2011;42:44–51.
20. von Elm E, Altman DG, Egger M, et al; STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. *Int J Surg*. 2014;12:1495–1499.
21. Zamanipoor Najafabadi AH, Ramspek CL, Dekker FW, et al. TRIPOD statement: a preliminary pre-post analysis of reporting and methods of prediction models. *BMJ Open*. 2020;10:e041537.
22. How to calculate the fold number (k-fold) in cross validation? Data Science Stacke Exchange. 2021. Accessed September 26, 2021.
23. Van Calster B, McLernon DJ, van Smeden M, et al; Topic Group 'Evaluating diagnostic tests and prediction models' of the STRATOS initiative. Calibration: the Achilles heel of predictive analytics. *BMC Med.* 2019;17:230.
24. Royston P, Altman DG. External validation of a Cox prognostic model: principles and methods. *BMC Med Res Methodol*. 2013; 13:33.
25. Rocker GM, Heyland DK, Cook DJ, et al. Most critically ill patients are perceived to die in comfort during withdrawal of life support: a Canadian multicentre study. *Can J Anaesth*. 2004;51:623–630.
26. Epker JL, Bakker J, Kompanje EJ. The use of opioids and sedatives and time until death after withdrawing mechanical ventilation and vasoactive drugs in a dutch intensive care unit. *Anesth Analg*. 2011;112:628–634.
27. Mayer SA, Kossoff SB. Withdrawal of life support in the neurological intensive care unit. *Neurology*. 1999;52:1602–1609.
28. Gerritsen T. *Richtlijn nalaten en staken van behandeling en palliatieve zorg na het staken van behandeling bij volwassen IC patiënten. Dutch Intensive Care Society (NVIC guidelines)*. 2009. Available at https://www.nvic.nl/richtlijn/einde-leven-2021/. Accessed June 13, 2021.