OPEN

Machine Learning Algorithms in Controlled Donation After Circulatory Death Under Normothermic Regional Perfusion: A Graft Survival Prediction Model

Rafael Calleja, MD,[1] Marcos Rivera, BSc,[2] David Guijo-Rubio, PhD,[2] Amelia J. Hessheimer, MD, PhD,[3] Gloria de la Rosa, MD, PhD,[4] Mikel Gastaca, MD, PhD,[5] Alejandra Otero, MD, PhD,[6] Pablo Ramírez, MD, PhD,[7] Andrea Boscà-Robledo, MD, PhD,[8] Julio Santoyo, MD, PhD,[9] Luis Miguel Marín Gómez, MD, PhD,[10] Jesús Villar del Moral, MD, PhD,[11] Yiliam Fundora, MD, PhD,[12] Laura Lladó, MD, PhD,[13] Carmelo Loinaz, MD, PhD,[14] Manuel C. Jiménez-Garrido, MD, PhD,[15] Gonzalo Rodríguez-Laíz, MD, PhD,[16] José Á. López-Baena, MD, PhD,[17] Ramón Charco, MD, PhD,[18] Evaristo Varo, MD, PhD,[19] Fernando Rotellar, MD, PhD,[20] Ayaya Alonso, MD,[21] Juan C. Rodríguez-Sanjuan, MD, PhD,[22] Gerardo Blanco, MD, PhD,[23] Javier Nuño, MD, PhD,[24] David Pacheco, MD, PhD,[25] Elisabeth Coll, MD, PhD,[4] Beatriz Domínguez-Gil, MD, PhD,[4] Constantino Fondevila, MD, PhD,[3] María Dolores Ayllón, MD, PhD,[1] Manuel Durán, MD, PhD,[1] Ruben Ciria, MD, PhD,[1] Pedro A. Gutiérrez, PhD,[2] Antonio Gómez-Orellana, MSc,[2] César Hervás-Martínez, PhD,[2] and Javier Briceño, MD, PhD[1]

Background. Several scores have been developed to stratify the risk of graft loss in controlled donation after circulatory death (cDCD). However, their performance is unsatisfactory in the Spanish population, where most cDCD livers are recovered using normothermic regional perfusion (NRP). Consequently, we explored the role of different machine learning-based classifiers as predictive models for graft survival. A risk stratification score integrated with the model of end-stage liver disease score in a donor-recipient (D-R) matching system was developed. **Methods.** This retrospective multicenter cohort study used 539 D-R pairs of cDCD livers recovered with NRP, including 20 donor, recipient, and NRP variables. The following machine learning-based classifiers were evaluated: logistic regression, ridge classifier, support vector classifier, multilayer perceptron, and random forest. The endpoints were the 3- and 12-mo graft survival rates. A 3- and 12-mo risk score was developed using the best model obtained. **Results.** Logistic regression yielded the best performance at 3 mo (area under the receiver operating characteristic curve = 0.82) and 12 mo (area under the receiver operating characteristic curve = 0.83). A D-R matching system was proposed on the basis of the current model of end-stage liver disease score and cDCD-NRP risk score. **Conclusions.** The satisfactory performance of the proposed score within the study population suggests a significant potential to support liver allocation in cDCD-NRP grafts. External validation is challenging, but this methodology may be explored in other regions.

(*Transplantation* 2025;109: e362–e370).

Received 27 February 2024. Revision received 21 September 2024.

Accepted 17 October 2024.

[1] *Hepatobiliary Surgery and Liver Transplantation Unit, Maimonides Biomedical Research Institute of Cordoba (IMIBIC), Hospital Universitario Reina Sofía, University of Córdoba, Córdoba, Spain.*

[2] *Department of Computational Sciences and Numerical Analysis, University of Córdoba, Córdoba, Spain.*

[3] *General and Digestive Surgery Department, Hospital Universitario La Paz, Madrid, Spain.*

[4] *Organización Nacional de Trasplantes, Madrid, Spain.*

[5] *Hepatobiliary Surgery and Liver Transplantation Unit, Biocruces Bizkaia Health Research Institute, Cruces University Hospital, University of the Basque Country, Bilbao, Spain.*

[6] *General and Digestive Surgery Department, Complejo Hospitalario Universitario de A Coruña, A Coruña, Spain.*

[7] *General and Digestive Surgery Department, Hospital Clínico Universitario Virgen de la Arrixaca, IMIB, El Palmar, Spain.*

[8] *General and Digestive Surgery Department, Hospital Universitari i Politècnic La Fe, Valencia, Spain.*

[9] *General and Digestive Surgery Department, Hospital Regional Universitario de Málaga, Spain.*

[10] *General and Digestive Surgery Department, Hospital Universitario Virgen del Rocío, Sevilla, Spain.*

[11] *General and Digestive Surgery Department, Hospital Universitario Virgen de las Nieves, Instituto de Investigación Biosanitaria ibs.GRANADA, Granada, Spain.*

[12] *Division of Hepatobiliary and General Surgery, Department of Surgery, Institut de Malalties Digestives I Metabòliques (IMDiM), Hospital Clínic, University of Barcelona, Barcelona, Spain.*

[13] *General and Digestive Surgery Department, Hospital Universitario de Bellvitge, Hospitalet de Llobregat, Spain.*

[14] *General and Digestive Surgery Department, Hospital Universitario 12 de Octubre, Madrid, Spain.*

[15] *General and Digestive Surgery Department, Hospital Universitario Puerta de Hierro, Majadahonda, Spain.*

INTRODUCTION

Liver transplantation (LT) is the best treatment option for several end-stage liver diseases. However, the shortage of donors and the emergence of new indications perpetuates an imbalance between donors and candidates, resulting in deaths and exclusions from the waiting list.[1,2] Strategies including donations after circulatory determination of death (DCD) and extended criteria donors aim to mitigate this situation. DCD donations have grown exponentially in recent years, increasing from 1207 in 2023 to >4700 in 2022 in the United States. In the United Kingdom and Spain, DCD represents 46% of donors and 42% of transplants, respectively.[3] However, grafts derived from DCD and extended criteria donors carry a higher risk than grafts obtained from donors declared dead based on the neurological criteria.[4] In the United States, the 5-y graft survival rates exceeded 75% across most LT categories, except for donors aged 65 y or older, recipients with a model for end-stage liver disease (MELD) score ≥40, and DCD liver recipients.[5]

In addition to the increased risk of these donors, there is a lack of consensus on the allocations of grafts generated using these strategies, leading to the development of scores such as the donor risk index (DRI),[6] survival outcomes after LT (SOFT),[7] and balance of risk (BAR),[8] which aim to assess the risk of donor-recipient (D-R) assignment to achieve better posttransplant results. The BAR score exhibits an area under the curve (AUC) of >0.70 for 90-d mortality and outperforms other predictive systems, including the DRI, MELD, and SOFT scores.[9] However, none of these systems align with the current organ allocation policies. Furthermore, they are not specific to the assignment of DCD livers.

An ideal allocation system should optimize D-R matching, prioritize candidates with the highest risk of death on the waiting list, and enhance the likelihood of graft survival success.

Specific scores, such as the University of California and Los Angeles (UCLA)-DCD,[10] King College Hospital (KCH)-DCD risk index,[11] or UK-DCD score, have been used for assessing the potential risk associated with LT in the context of DCD.[12] Although the UK-DCD score is a relevant contribution, it does not accurately predict outcomes in Spain owing to unique DCD characteristics[13]: it yields high graft survival rates even in high-risk cases (1-y graft survival rates of 100% and 91% in the "futile" and "high-risk" groups, respectively[14]) and it lacks a matching system.

Normothermic regional perfusion (NRP) is an in situ preservation technique implemented in several European countries to reduce the effects of ischemia/reperfusion injury in controlled DCD (cDCD) livers.[15] The cDCD program was regulated a decade ago in Spain, and its results and the use of in situ NRP have been highly satisfactory.[16,17] In a previous study, 506 of 803 cDCD livers were recovered with NRP (63%), whereas 258 were recovered with the standard rapid recovery technique (32%).[17] This study positioned NRP as the preferred technique by Spanish transplant groups for cDCD procedures.[17] Notably, most cDCD-NRP procedures in Spain are performed with antemortem cannulation of the vessels following specific authorization from the donor surrogate decision-makers or the donor themselves in case of first-person consent.

Different machine learning (ML) models have been used to improve organ allocation.[18-21] In the M.A.D.R.E. study, artificial neural networks (ANNs) predicted the probability of graft survival (AUC: 0.80) and graft loss (AUC: 0.82) with 90.79% and 71.42% accuracy, respectively.[20] These models surpassed traditional scores and generated

[16] General and Digestive Surgery Department, Hospital General Universitario de Alicante, Alicante, Spain.

[17] General and Digestive Surgery Department, Hospital General Universitario Gregorio Marañón General University Hospital, Madrid, Spain.

[18] General and Digestive Surgery Department, Hospital Universitario Vall d'Hebron, Barcelona, Spain.

[19] General and Digestive Surgery Department, Complejo Hospitalario Universitario de Santiago, Santiago de Compostela, Spain.

[20] General and Digestive Surgery Department, Hospital Universitario de Navarra, Pamplona, Spain.

[21] General and Digestive Surgery Department, Hospital Universitario Nuestra Señora de Candelaria, Santa Cruz de Tenerife, Spain.

[22] General and Digestive Surgery Department, Hospital Universitario Marqués de Valdecilla, Santander, Spain.

[23] General and Digestive Surgery Department, Hospital Universitario Infanta Cristina, Badajoz, Spain.

[24] General and Digestive Surgery Department, Hospital Universitario Ramón y Cajal, Madrid, Spain.

[25] General and Digestive Surgery Department, Hospital Universitario Río Hortega, Valladolid, Spain.

This research has received financial support from Mutua Madrileña XVIII Convocatoria de Ayudas a la Investigación (AP177512021). D.G.-R. has been supported by the "Agencia Estatal de Investigación (España)." MCIU/AEI/10.13039/501100011033 and European Union NextGenerationEU/PRTR (grant JDC2022-048378-I) and by the University of Córdoba through competitive grants for Andalusian society challenges (grant PP2F_L1_15). Antonio Manuel Gómez-Orellana has been supported by "Consejería de Transformación Económica, Industria, Conocimiento y Universidades de la Junta de Andalucía" (grant PREDOC-00489).

The authors declare no conflicts of interest.

R.C., M.R., D.G.-R., P.A.G., A.G.-O., C.H.-M., and J.B. contributed to study design and methodology and formal analysis. R.C., A.J.H., G.d.l.R., M.G., A.O., P.R., A.B.-R., Y.F., J.S., M.A.G.-B., J.V.d.M., L.L., C.L., M.C.J.-G., G.R.-L., J.A.L.-B., R.Ch., E.V., F.R., A.A., J.C.R.-S., G.B., J.N., D.P., E.C., B.D.-G., C.F., and J.B. contributed to data acquisition. R.C., M.R., D.G.-R., .H.-M., and J.B. contributed to writing the original draft. R.C., M.R., D.G.-R., A.J.H., G.d.l.R., M.G., A.O., P.R., A.B.-R., Y.F., J.S., M.A.G.-B., J.V.d.M., L.L., C.L., M.C.J.-G., G.R.-L., J.A.L.-B., R.Ch., E.V., F.R., A.A., J.C.R.-S., G.B., J.N., D.P., E.C., B.D.-G., C.F., M.D.A., M.D., R.Ci., P.A.G., A.G.-O., C.H.-M., and J.B. participated in writing—review and editing.

The research data belong to the different Spanish transplantation units, with the collaboration of the Spanish Liver Transplant Registry and the Spanish National Transplant Organization. The participants of this study did not give written consent for their data to be shared publicly; therefore, due to the sensitive nature of the research, supporting data are not available.

The institutions mentioned above played no part in the design or conduct of the study; collection, management, analysis, or interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

Visual abstract is available online at doi.org/10.1097/TP.0000000000005312.

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: Rafael Calleja, MD, Unit of Hepatobiliary Surgery and Liver Transplantation, University Hospital Reina Sofia, Avenida Menendez Pidal s/n, 14004 Cordoba, Spain. (rafael.calleja.sspa@juntandeandalucia.es).

ISSN: 0041-1337/20/1097-e362

DOI: 10.1097/TP.0000000000005312

a rule-based matching system. This methodology has been externally validated.[18] Although ML models are widely used in LT, they are not autonomous and should be considered decision-support tools.[19] Furthermore, it is unclear whether ML algorithms can improve allocation accuracy by estimating the risk of LTs from cDCD.

Therefore, we explored the performance of different ML techniques, including logistic regression (LR), to assess the risk of D-R matching in cDCD performed with NRP. We proposed an automatic D-R matching system by integrating the best model obtained with the currently used MELD score to assist liver allocation.

PATIENTS AND METHODS

Study Design

This retrospective observational cohort study was conducted on a cDCD-NRP LT cohort (n = 545) in Spain, as previously reported by Hessheimer et al.[17] The data set was obtained from the Spanish Liver Transplant Registry and the Spanish Organización Nacional de Trasplantes. A detailed description of cDCD donor and recipient selection and the NRP protocol has been provided by Hessheimer et al.[17] Immediately after initiating cannulation, sodium heparin will be administered. The heparin dose is specified in the cDCD national protocol.[22] However, there may be variations between centers. After excluding pediatric transplants (n = 2), acute liver failure (n = 1), and graft survival time not specified cases (n = 3), 539 D-R pairs were selected. This article adheres to the TRIPOD checklist (**Appendix 1, SDC,** http://links.lww.com/TP/D218).[23]

Variables and Outcomes

All cohort variables related to donor, recipient, and graft characteristics, including graft-related complications, are shown in Table 1. Different ML models were built considering the following 20 variables for each D-R pair: donor age, donor body mass index, A-NRP duration, location of withdrawal of life-sustaining therapy (WLST), antemortem cannulation, intubation period before WLST, asystolic warm ischemia time, total warm ischemia time, functional warm ischemia time (FWIT), cold ischemia time (CIT), aspartate aminotransferase (AST) slope, alanine aminotransferase (ALT) slope, lactate slope, recipient age, recipient body mass index, sex compatibility, recipient MELD (laboratory), retransplantation, and graft survival status. These variables were used as inputs to the ML models to estimate the probability of graft survival for D-R matching. At the start of A-NRP, an initial blood sample is collected, followed by additional samples every 30 min to assess parameters such as lactate and transaminases, which assist in graft viability evaluation. Lactate levels were measured in millimoles per liter and transaminases in units per liter. Detailed methodology for calculating lactate and transaminase slopes is provided in **Appendix 2** (**SDC,** http://links.lww.com/TP/D218).

Ischemia time was reported in minutes, except for CIT, which was expressed in hours. CIT starts with donor cold preservation until graft reperfusion. The total warm ischemia time is defined as the period between the WLST and the start of NRP. FWIT starts when the donor systolic blood pressure drops to <60 mm Hg and ends with the initiation of NRP. AWIT is the period from cardiac arrest to preservation, including the "no-touch" period. Graft failure is defined as death or retransplantation.

TABLE 1.

Donor-, recipient-, and transplant-related baseline characteristics

Characteristics	A-NRP (N = 539)
Donor	
Age, y	57 (49–67)
Sex male, n (%)	346 (64)
BMI, kg/m^2	26.604 (24.16–28.89)
Intubation before WLST, d	10 (4–11)
Cause of death, n (%)	
CVA	246 (45)
Cerebral anoxia	173 (32)
TBI	52 (10)
Other	68 (13)
A-NRP	
Antemortem cannulation, n (%)	500 (93)
WLST location OR, n (%)	394 (73)
ALT slope	0.178 (–0.04 to 0.257)
AST slope	0.121 (–0.075 to 0.267)
Lactate slope	–0.029 (–0.049 to 0.004)
TWIT, min	19 (13–23)
FWIT, min	13 (9–16)
AWIT, min	6 (5–6)
A-NRP duration, min	108 (81–126)
CIT, min	330 (270–375)
Preservation solution, n (%)	
UW or IGL-1	121 (22)
Celsior	327 (61)
HTK	73 (14)
Other	18 (3)
Recipient	
Age, y	57 (53–63)
Sex male, n (%)	425 (79)
Sex compatibility, n (%)	348 (64)
BMI, kg/m^2	27.08 (23.73–30.16)
Laboratory MELD score	13 (9–16)
Transplant indication, n (%)	
Cirrhosis	346 (65)
HCC	136 (25)
PSC	10 (2)
Retransplantation	17 (3)
Other	29 (5)
Graft-related complications, n (%)	
Early allograft dysfunction	79 (14)
Primary nonfunction	14 (2.6)
Hepatic artery thrombosis	21 (3.8)
Biliary complications, n (%)	58 (10.7)
Biliary stricture	37 (6.8)
Biliary leakage	13 (2.5)
ITBL	7 (1.3)
Others	1 (0.1)

A-NRP, abdominal normothermic regional perfusion; ALT, Alanine aminotransferase; AST, aspartate aminotransferase; AWIT, asystolic warm ischemia time; BMI, body mass index; CIT, cold ischemia; CVA, cerebrovascular accident; FWIT, functional warm ischemia; HCC, hepatocellular carcinoma; HTK, histidine-tryptophan-ketoglutarate; IGL-1, Institut Georges Lopez-1; ITBL, ischemic type biliary lesion; MELD, model of end-stage liver disease; OR, operating room; PSC, primary sclerosing cholangitis; TBI, trauma brain injury; TWIT, total warm ischemia time; UK DCD risk score, United Kingdom Donation after Circulatory Death risk score; UW, University of Wisconsin; WLST, withdrawal of life-sustaining therapy.

Graft survival is reported in months with a minimum follow-up of 1 y.

ML models were developed using a data set from a Spanish multicenter, representing the most significant cDCD LT experience under NRP. The 3- and 12-mo graft survival (3 M and 12 M models, respectively) periods were selected as the endpoints to explore the performance of different models (including ML models) in predicting liver graft survival. The score was integrated with the MELD score using a rule-based matching system.

ML Techniques

Five ML techniques were considered: LR, ridge classifier (RC), support vector classifier (SVC), multilayer perceptron (MLP), and random forest (RF).[21] Model performance was evaluated using the accuracy or correct classification rate and AUC. The ML models were trained using a hold-out methodology wherein 75% of the patterns (D-R matches) were randomly selected from different years to obtain an accurate model. The remaining 25% of the patterns were used for testing and validating the trained models. This last set of patterns was hidden and excluded from the training phase of the model. Moreover, a 10-fold cross-validation procedure was performed during the training stage (**Table S1, SDC,** http://links.lww.com/TP/D219). Both cohorts (training and validation) were analyzed to identify statistically significant differences that could bias the results of the best model obtained (**Table S2, SDC,** http://links.lww.com/TP/D219). MLP and RF were executed 10 times owing to their stochastic nature resulting from initialization. Hence, their results are expressed as mean±SD. These metrics served as performance evaluators and facilitated a more analytical and clinical evaluation of the input variables selected by each model and their weights.

Spanish Score System

After identifying the best ML model, the SPanish NRP cDCD risk score was constructed using the Framingham risk scheme.[24] This score was used to estimate graft survival associated with D-R matching at 3 and 12 mo (SP3M and SP12M). A lower score indicated a higher probability of graft survival. The scoring system was discretized into several categories to stratify the risk into low, medium, and high. The coefficients for each input variable ($\beta_1, \beta_2, ..., \beta_n$, where n is the number of input variables considered by the model) were identified. The corresponding reference values for each category were determined as the midpoint ($W_1REF, W_2REF, ..., W_nREF$). The base category, considered a reference risk factor, was determined to be the healthier state, and zero points were assigned. However, the worst states of the risk factors were assigned positive points. These points were assigned on the basis of the distance from each category to the base category and B, the number of regression units reflecting one point in the scoring system. The points were calculated using the following equation: $Points_{ij} = \beta_i \frac{(W_{ij} - W_iREF)}{B}$, where i represents the input variable considered by the model and j represents the categories.

Rule-based Matching System

An automatic D-R matching system was designed on the basis of the aforementioned score. It calculated the previously suggested score for each potential D-R match on the waiting list. The D-R pair with the lowest score, indicating better graft survival, was matched. The graft was allocated to the recipient with the highest MELD score in the case of a tie. The recipient's waiting list time was considered if a second tie occurred.

Statistical Analysis

Quantitative variables are reported as means and ranges, and qualitative variables are expressed as numbers and percentages. All values were logarithmically transformed to consider a similar magnitude. Missing values for each input variable have been reported (**Table S3, SDC,** http://links.lww.com/TP/D219) and were recovered using the mean values of the variables computed from the training split. The specific methodologies of the models, scores, and matching systems are explained in the following sections. All statistical analyses were performed using R version 4.2.0 (RStudio, PBC).

Ethics

All procedures, including obtaining informed consent, were conducted in accordance with the ethical standards of the Helsinki Declaration of 1975. The Spanish cDCD LT procedure is regulated by Royal Decree 1723/2012[25] and official national protocols.[22,26] This study was approved by the Institutional Review Board of Reina Sofía University Hospital.

RESULTS

Table 1 summarizes the cohort characteristics. The graft survival rates at 3 and 12 mo were 92.9% and 91%, respectively. In the cohort, of the 17 retransplanted patients, 5 cases occurred within the very early period (<2 wk), 6 cases from 2 wk to 3 mo, and 6 cases were performed after 3 mo. No liver grafts in this cohort underwent ex situ machine perfusion.

ML Techniques

RF and MLP achieved higher accuracy levels 92% at 3 M achieved by RF and MLP and 90% at 12 M with RF among all techniques. However, the results in terms of AUC were limited, ranging from 57% to 73%. The LR method achieved the best balance between accuracy and AUC (AUC: 82% and 83% for the 3 M and 12 M endpoints, respectively; Figure 1). The accuracy slightly decreased to 71% and 77%, respectively. RC and SVC presented limited results in terms of AUC. Table 2 presents the performances of the different ML models over the testing set for both endpoints. There were no statistically significant differences between the training and validation cohorts in our best model (LR method; **Table S2, SDC,** http://links.lww.com/TP/D219). The LR model assigned various coefficients to each input variable, indicating their weights. A negative coefficient indicated graft survival. Variables with a coefficient of <0.1 were excluded owing to their low significance in the model. Therefore, among the initial 20 variables, only 12 and 10 variables were included in the 3 M and 12 M models, respectively.

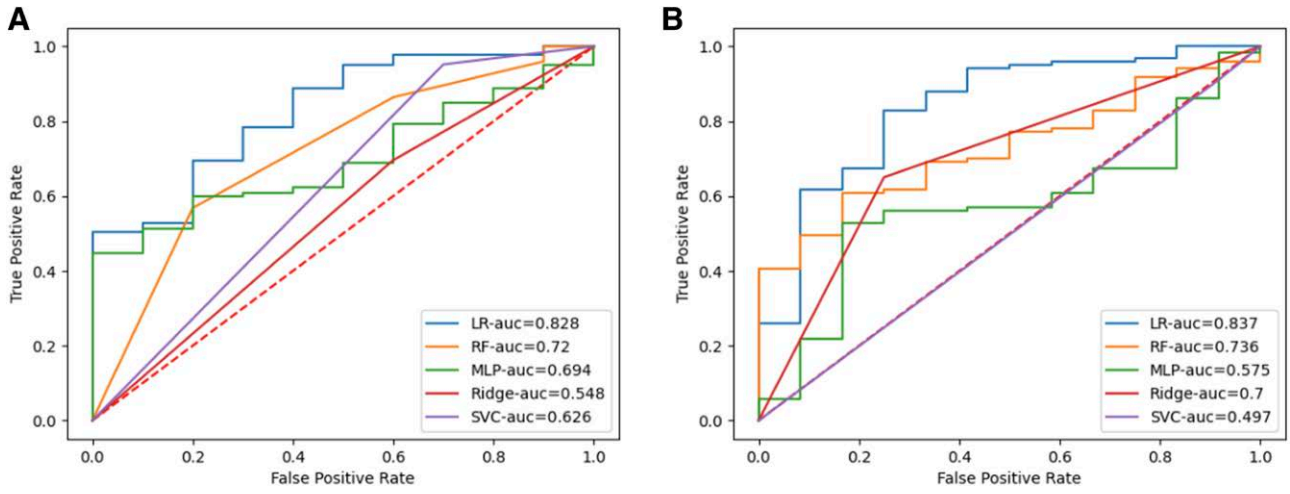

FIGURE 1. Performance of different models in 3-mo and 12-mo graft survival prediction based on ROC curve comparison. The LR (AUC = 0.837) shows superiority in both endpoints of graft survival prediction capability against other prediction models. The ROC curves showing the 3-mo endpoint model performance (A) and the 12-mo endpoint (B). AUC, area under the curve; CCR, correct classification rate; LR, logistic regression; MLP, multiLayer perceptron; RF, random forest; ROC, receiver operating characteristic; SVC, support vector classifier.

TABLE 2.

Results achieved by the ML techniques in terms of CCR and AUC evaluation metrics (values between 0 and 1) for the 3 M and 12 M endpoints

	3 M		12 M	
ML technique	**Accuracy**	**AUC**	**Accuracy**	**AUC**
MLP	0.926 ± 0	0.694 ± 0.097	0.864 ± 0.026	0.575 ± 0.094
RC	0.674	0.548	0.659	0.700
RF	0.921 ± 0.009	0.720 ± 0.074	0.903 ± 0.026	0.736 ± 0.032
SVC	0.904	0.626	0.837	0.497
LR	0.719	0.828	0.778	0.837

The values for the MLP and RF techniques are expressed as mean ± SD of the evaluation metrics, as they are stochastic techniques.
3 M, 3 mo; 12 M, 12 mo; AUC, area under the curve; CCR, correct classification rate; LR, logistic regression; ML, machine learning; MLP, multilayer perceptron; RC, ridge classifier; RF, random forest; SVC, support vector classifier.

Spanish NRP cDCD Score

Table 3 summarizes the coefficient values for each risk factor. **Tables S4 and S5** (**SDC,** http://links.lww.com/TP/D219) show the risk factors, categories obtained, and points associated with each one for the 3 M and 12 M endpoints, respectively. Hence, the values for the category ranges are on a logarithmic scale. The 3 M and 12 M scores ranged from 0 to 26 and 0 to 14, respectively. The 2 most influential variables, regardless of the endpoint, were "retransplantation" and "CIT." Figure 2 shows the differences in graft survival corresponding to each developed score (at 3 M and 12 M) according to the risk level. In our cohort, 53% and 61% of the patients were considered medium risk for transplantation at 3 and 12 mo, respectively. The number of high-risk D-R pairings did not exceed 15% for both endpoints.

The negative coefficients for lactate and AST/ALT slopes suggest that an increase in these slopes is associated with a higher risk of graft failure. This means that if lactate or AST/ALT levels rise over time (positive slope), it is associated with a worse graft prognosis. (**Appendix 2, SDC,** http://links.lww.com/TP/D218).

D-R Matching System

To better understand the score, 5 recipients and 10 donors were randomly selected from the database. Table 4 details the scores associated with each potential D-R and the MELD for each recipient. D-R matching was performed by selecting the pair with the lowest value with a difference of at least one point to the other potential pairs. For example, consider donors 1, 3, or 4. Donors 1, 3, and 4 were matched with recipients 1, 4, and 2, respectively. Matching was performed according to the MELD guidelines in the case of a tie. For example, consider donor 2, for which recipients 3 and 5 had a SP12M score of 4 points. The recipient with a higher MELD score, that is, recipient 3 with a MELD score 17, was selected. A similar scenario was observed for donor 8, with a tie between recipients 2 and 3. The recipient with the highest MELD score was selected (recipient 4).

DISCUSSION

cDCD has been widely used to address donor shortages. The extended criteria grafts were assumed to be at a higher risk; thus, scores such as MELD, DRI, SOFT, and BAR were

TABLE 3.

Coefficients obtained for the LR model at the 3 M and 12 M endpoints

3 M endpoint		12 M endpoint	
Input variable name	**Coefficient**	**Input variable name**	**Coefficient**
Retransplantation	–1.467	Retransplantation	–0.579
Cold ischemia time	–0.589	Cold ischemia time	–0.368
Antemortem cannulation	0.575	Recipient age	–0.187
Lactate slope	–0.305	ALT slope	–0.176
ALT slope	–0.220	Antemortem cannulation	0.159
AST slope	–0.180	MELD	–0.141
MELD	–0.164	Lactate slope	–0.135
WLST location OR	0.155	Sex compatibility	0.130
Intubation period before WLST, d	–0.141	WLST location OR	0.122
FWIT	–0.134	AST slope	–0.114
Sex compatibility	0.131		
Recipient age	–0.125		

The input variables were ordered according to the absolute value of the coefficient.
3 M, 3 mo; 12 M, 12 mo; ALT, alanine aminotransferase; AST, aspartate aminotransferase; FWIT, functional warm ischemia; LR, logistic regression; MELD, model of end-stage liver disease; OR, operating room; WLST, withdrawal of life-sustaining therapy.

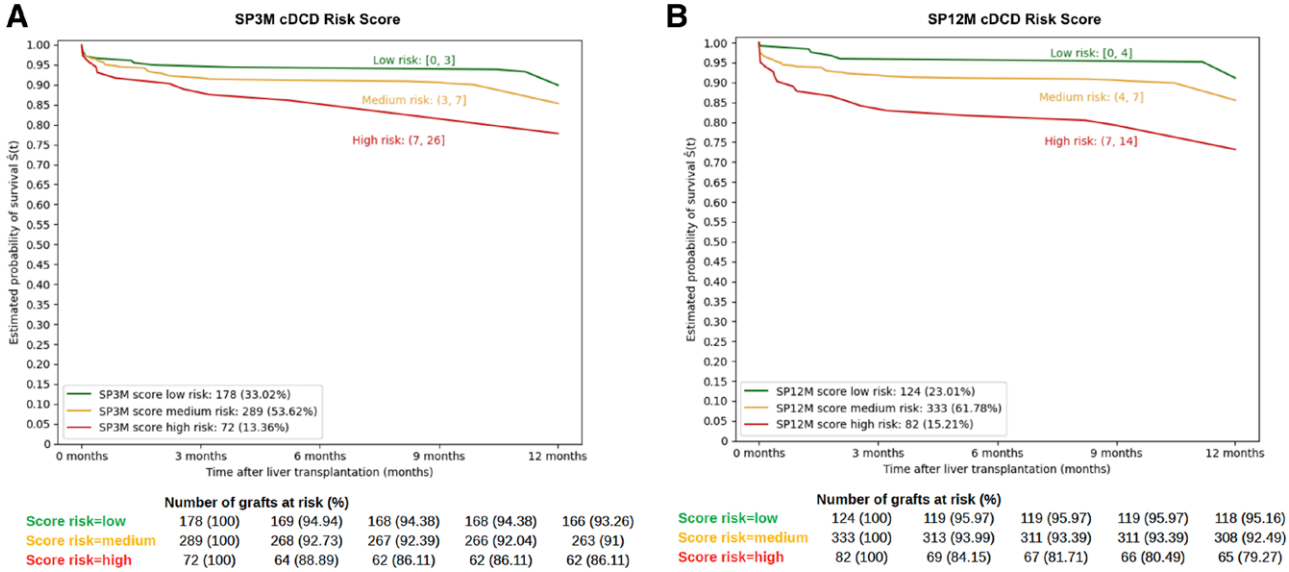

FIGURE 2. Five-year graft survival based on Spanish cDCD risk score stratification for 3- and 12-mo graft survival endpoints (SP3M and SP12M) represented by a Kaplan-Meier curve. The Kaplan-Meier curves representing the 3- and 12-mo endpoint score risk stratification are shown in A and B, respectively. The numeric range in parentheses represents the score points assigned to a level of risk. cDCD, controlled donation after circulatory death.

proposed to stratify the LT risk in this scenario.[27] The use of ML algorithms to improve D-R matching is not novel. Although ML algorithms have offered successful models in LT,[28] evidence on how ML can assist in allocating liver grafts obtained from cDCD donors with organ recovery performed by the use of NRP are lacking. This study is the first to explore performance of different ML techniques in this context.

This study initially assessed the potential role of ML algorithms in predicting liver graft survival using cDCD-NRP. The largest cDCD-NRP cohort[17] was analyzed to explore these algorithms. LR (AUC 0.82 and AUC 0.83, respectively) overpassed both the 3- and 12-mo graft survival prediction compared with other ML algorithms (SVC, RF, MLP, or RC). ML algorithms are typically associated with techniques such as RFs or ANNs. LR is sometimes seen as a traditional statistical method and less associated with the field of ML. However, LR is considered the simplest form of ML and remains a widely used technique. Although algorithms such as ANNs or RF have been used to estimate posttransplant graft survival, only 2 previous studies have integrated ANNs into a D-R matching system.[18,20] However, the heavy dependency of these models on the databases used for training limits their exportability and raises concerns about their applicability in diverse countries, highlighting the need for standardizing registry-based variables or developing region-specific models[29] to address this issue. ANNs and RF did not achieve satisfactory performance for 2 reasons: the small number of patients (a larger data set is required for training) and imbalanced classification of some variables. Previous studies used up to 57 variables

TABLE 4.

Demonstration of the D-R matching system

	D-R matching system					
	Recipient 1	Recipient 2	Recipient 3	Recipient 4	Recipient 5	
MELD	24	18	17	20	16	Allocation
Donor 1	**4**	6	5	8	7	Recipient 1
Donor 2	5	6	**4**	6	4	Recipient 3
Donor 3	5	9	4	**2**	9	Recipient 4
Donor 4	6	**3**	7	5	8	Recipient 2
Donor 5	**6**	8	8	9	7	Recipient 1
Donor 6	5	9	6	9	**4**	Recipient 5
Donor 7	9	**5**	8	9	6	Recipient 2
Donor 8	6	4	7	**4**	5	Recipient 4
Donor 9	7	8	4	8	**3**	Recipient 5
Donor 10	3	7	**2**	6	5	Recipient 3

A recipient is chosen for a specific donor. Green shading indicates the matching that is allocated by this prioritization system. In the case of ties, there are several potential recipients; hence, the one in bold is the matching to be performed (induced by the recipient MELD).
D-R, donor-recipient; MELD, model of end-stage liver disease.

from 1003 LTs[20] or external center data sets,[18] whereas this study used only 20 variables from 539 LTs. Thus, prediction models with many variables can be clinically tedious. Furthermore, unlike LR, where variable-outcome relationships are discernible, ANNs are often viewed as "black boxes." They do not provide a clear understanding of the clinical variables that influence their predictive ability. This is a crucial limitation because the model may serve as a decision-support system for D-R allocation, where transparency is mandatory. Hence, data protection policies and the complexities of ANNs and RF models hinder their clinical applicability.[30,31]

Among the 12 and 10 variables considered for the 3- and 12-mo risk score, respectively, CIT and retransplantation were the most relevant variables, supporting the findings reported by Hessheimer et al.[17] These variables were also important risk factors in the UCLA-DCD,[10] KCH-DCD risk index,[11] and UK-DCD scores.[12] MELD did not have a significant effect in our model, probably owing to the exclusion of certain factors (acute liver failures, hepatocellular carcinoma rate: 26%) and a low MELD score in our study. FWIT strongly predicts post-cDCD graft complications, including ischemic type biliary lesion incidence.[32,33] However, FWIT has no role in the 12 M model, and its role is marginal in the 3 M model, as NRP with antemortem cannulation shortens the duration of FWIT. Moreover, the distribution of FWIT was homogeneous in our database (mean: 13 min). Consequently, FWIT did not significantly affect the endpoint. However, its effect on other endpoints, such as ischemic type biliary lesion incidence, could be more significant, although the incidence in our cDCD-NRP population was low (1%). Some variables have not been included in other scores, such as antemortem cannulation (exclusive to our model owing to Spanish law), location of withdrawal of life support, or sex mismatch. Among these variables, antemortem cannulation achieved more relevance in both models. Specific variables related to NRP, such as lactate, AST, and ALT levels, were included. These variables were introduced as slopes rather than absolute values in an effort to obtain a dynamic variable that reflects the evolution of these values during NRP. Certain limitations of these variables as markers of viability have been discussed previously.[15] One pitfall of LR models with continuous variables, such as AST or ALT slopes, is the potential to over-penalize transitory physiological increases or secondary to procurement, which does not necessarily indicate graft injury. To address this issue, there are 2 possible alternatives: (1) reduce the number of missing values and increase the sample size, or (2) exclude transaminases from a future model, considering only the lactate slope. At 3 mo, the lactate slope has the greatest influence compared with transaminases. However, at 12 mo, while the ALT slope shows a higher risk coefficient, the values of all 3 variables are very similar. Overall, the 3M score presents a more asymmetric distribution of coefficient values compared with the 12 M score, which has a more homogeneous distribution. Liver retransplantation, especially in the first weeks, probably significantly impacts short-term graft survival compared with long-term outcomes.[34] Therefore, it seems reasonable to consider excluded retransplants from the model; however, it would lead to a more restrictive allocation system. Furthermore, this variable has been considered in other existing scores.[12] A larger data set would likely reduce the impact of this variable in a future model.

The second part of our study involved designing a cDCD-NRP score based on the best model obtained. Figure 2 shows that the 3 M and 12 M cDCD-NRP scores share similar ROC curves. Similar to the UK-DCD score,[12] the highest-risk combination in cDCD-NRP patients at 3 and 12 mo was CIT >6 h and retransplantation. This combination was identified in 5 D-R pairs (<1%) in our data set. From this, we can infer that it is critical to minimize the association with other risk factors whenever possible in patients requiring retransplantation to improve graft survival outcomes. Despite previous reports on these findings,[17] what our model provides the probabilistic stratification of graft survival risk that was not previously defined. Identifying different risk groups within our cohort may be a starting point for exploring the benefits of ex situ machine perfusion, which may play a significant role in exporting the applicability of our model. In regions with variations in cDCD protocols, demographics, or ischemia

times, the inclusion of ex situ machine perfusion as a protective factor may help decrease graft risk, compensating for other risk factors in the score.

A comparison against other scores (KCH, UCLA, or UK-DCD) could not be performed in this study as some input variables were not collected by Spanish centers. Although it has been previously stated that the UK-DCD score is not applicable in our region,[13] a hypothetical comparison would be biased. First, both countries differ in their donation protocols, retrieval techniques, and logistics, affecting, among others, ischemia times. Second, NRP-specific variables such as NRP duration, lactate, or transaminases are not present in UK-DCD.

Our score is a continuum of survival probability, not an all-or-nothing score, such as the SOFT or BAR scores. This score assigns a survival probability to each D-R pair, facilitating its integration into a D-R matching system, as we previously performed.[18,20] Through a system of rules, our score can be integrated with the MELD system. Hence, D-R matching would be based on graft survival probability (considering recipient, donor, and logistic factors) and the current allocation system (MELD-based) that prioritizes mortality on the waiting list. An ideal D-R matching system requires predefined variables, which our score satisfies, with the exception of cold ischemia, which is often estimated by the transplant team in our country.[20] In this study, the D-R allocation model is based on graft survival. Most studies focused on D-R matching have considered this endpoint. When analyzing the cohort, we observe survival rates >90% with the current allocation system. Therefore, it seems reasonable to consider alternative endpoints, such as posttransplant complications, immunosuppression-related complications, or the incidence of neoplasms. However, to develop an organ allocation support system that does not primarily focus on graft survival may be inadequate, as a D-R pairing that reduces the risk of acute kidney injury may not necessarily correlate with the best graft survival. In our opinion, graft survival should be the basis and all other endpoints should be complementary. In addition, certain improvements may be desirable: (1) to include the risk of waiting list dropout from the list of available recipients and (2) to assess the degree of penalization for high-risk recipients and establish compensation strategies. The solution to these issues probably involves advanced ML algorithms, such as multiobjective ANNs; however, this is unattainable within the current database. Hence, although our D-R matching proposal cannot replace the current allocation model, it provides valuable real-time decision support.

The study has 2 main limitations: the data set and its validation. The data set was obtained retrospectively. A widely criticized aspect of ML models is that they are not based on prospective data. Given the amount of data required to generate these models, this limitation is difficult to address. Moreover, the number of variables used was not sufficient to obtain ML models other than LR with an adequate AUC. The small number of variables, relatively small sample size, and low graft failure rate indicated that the weight distribution of the variables in our model is asymmetric. Despite the promising and applicable results, assessing whether increasing the number of variables is pertinent is advisable. In addition, we should be cautious about the role of lactate and transaminase levels. The heterogeneous collection of these variables, especially at the beginning of the cDCD program, has led to a missing values rate of 8% for lactate and >15% for transaminases. These percentages should be decreased to increase the confidence of the model. Finally, other endpoints should be considered in future studies. Currently, this data set does not facilitate establishing other endpoints of interest, such as the risk of waitlist dropout (requires a new data set). Nonetheless, the database is being updated to improve the model's robustness, evaluate whether other ML models would perform better, and integrate other endpoints.

This study aimed to develop a risk prediction model and validate it in our population. External validation was not considered initially as it is challenging and may be controversial. In addition, before performing this validation, an update of the data set would be necessary. The scenario in Spain is unique, and the differences in demographics or donation protocols in other countries[15] limit their applicability to other populations. Consequently, other scores do not perform adequately in our population. Since the advent of artificial intelligence and ML models, the future rests in region-specific models rather than global models.

In conclusion, the application of ML models has experienced significant growth in LT. ML has the potential to improve LT outcomes, especially in the context of organ allocation. Although these models cannot replace human decisions based on clinical experience, they can be improved as decision-support tools. They are database-dependent models, requiring constant training to improve their robustness. In this study, we developed an LR model for predicting graft survival in cDCD-NRP, integrated into a rule-based system. Despite its limitations, our model offers an alternative to current scoring systems in our population to assist organ allocation. Moreover, this methodology may contribute as a starting point for developing future models in other populations that can outperform current scores limitations and improve D-R matching. The role of ex situ machine perfusion is yet to be explored in this score but may promote its exportability.

ACKNOWLEDGMENTS

The authors acknowledge Fundación Mutua Madrileña, Instituto Maimónides de Investigación Biomédica, the Spanish Liver Transplantation Society, and the Spanish National Transplant Organization. Finally, the authors thank the Liver Transplantation Unit of Reina Sofía University Hospital and all the liver donors in Spain.

REFERENCES

1. Kim WR, Lake JR, Smith JM, et al. OPTN/SRTR 2013 annual data report: liver. *Am J Transplant*. 2015;15(Suppl 2):1–28.
2. Neuberger J. Liver transplantation in the United Kingdom. *Liver Transpl*. 2016;22:1129–1135.
3. WHO-ONT. Global Observatory on Donation and Transplantation. https://www.transplant-observatory.org/data-charts-and-tables/. Accessed March 11, 2023.
4. Ziogas IA, Kakos CD, Esagian SM, et al. Liver transplant after donation from controlled circulatory death versus brain death: a UNOS database analysis and publication bias adjusted meta-analysis. *Clin Transplant*. 2022;36:e14521.
5. Kwong AJ, Kim WR, Lake JR, et al. OPTN/SRTR 2019 annual data report: liver. *Am J Transplant*. 2021;21(Suppl 2):208–315.
6. Feng S, Goodrich NP, Bragg-Gresham JL, et al. Characteristics associated with liver graft failure: the concept of a donor risk index. *Am J Transplant*. 2006;6:783–790.

7. Rana A, Jie T, Porubsky M, et al. The survival outcomes following liver transplantation (SOFT) score: validation with contemporaneous data and stratification of high-risk cohorts. *Clin Transplant*. 2013;27:627–632.
8. Dutkowski P, Oberkofler CE, Slankamenac K, et al. Are there better guidelines for allocation in liver transplantation? A novel score targeting justice and utility in the model for end-stage liver disease era. *Ann Surg*. 2011;254:745–753.
9. Boecker J, Czigany Z, Bednarsch J, et al. Potential value and limitations of different clinical scoring systems in the assessment of short- and long-term outcome following orthotopic liver transplantation. *PLoS One*. 2019;14:e0214221.
10. Hong JC, Yersiz H, Kositamongkol P, et al. Liver transplantation using organ donation after cardiac death: a clinical predictive index for graft failure-free survival. *Arch Surg*. 2011;146:1017–1023.
11. Khorsandi SE, Giorgakis E, Vilca-Melendez H, et al. Developing a donation after cardiac death risk index for adult and pediatric liver transplantation. *World J Transplant*. 2017;7:203–212.
12. Schlegel A, Kalisvaart M, Scalera I, et al. The UK DCD Risk Score: a new proposal to define futility in donation-after-circulatory-death liver transplantation. *J Hepatol*. 2018;68:456–464.
13. Jochmans I, Hessheimer AJ, Neyrinck AP, et al; ESOT Workstream 04 of the TLJ (Transplant Learning Journey) Project. Consensus statement on normothermic regional perfusion in donation after circulatory death: report from the European Society for Organ Transplantation's Transplant Learning Journey. *Transpl Int*. 2021;34:2019–2030.
14. Hessheimer AJ, Coll E, Ruiz P, et al. The UK DCD Risk Score: still no consensus on futility in DCD liver transplantation. *J Hepatol*. 2019;70:1034–1035.
15. Melandro F, Basta G, Torri F, et al. Normothermic regional perfusion in liver transplantation from donation after cardiocirculatory death: technical, biochemical, and regulatory aspects and review of literature. *Artif Organs*. 2022;46:1727–1740.
16. Hessheimer AJ, Coll E, Torres F, et al. Normothermic regional perfusion vs. super-rapid recovery in controlled donation after circulatory death liver transplantation. *J Hepatol*. 2019;70:658–665.
17. Hessheimer AJ, de la Rosa G, Gastaca M, et al. Abdominal normothermic regional perfusion in controlled donation after circulatory determination of death liver transplantation: outcomes and risk factors for graft loss. *Am J Transplant*. 2022;22:1169–1181.
18. Ayllon MD, Ciria R, Cruz-Ramirez M, et al. Validation of artificial neural networks as a methodology for donor-recipient matching for liver transplantation. *Liver Transpl*. 2018;24:192–203.
19. Briceno J, Calleja R, Hervas C. Artificial intelligence and liver transplantation: looking for the best donor-recipient pairing. *Hepatobiliary Pancreat Dis Int*. 2022;21:347–353.
20. Briceno J, Cruz-Ramirez M, Prieto M, et al. Use of artificial intelligence as an innovative donor-recipient matching model for liver transplantation: results from a multicenter Spanish study. *J Hepatol*. 2014;61:1020–1028.
21. Guijo-Rubio D, Briceno J, Gutierrez PA, et al. Statistical methods versus machine learning techniques for donor-recipient matching in liver transplantation. *PLoS One*. 2021;16:e0252068.
22. Protocolo nacional de donación y trasplante hepático en donación en asistolia controlada. 2015. Available at https://www.ont.es/wp-content/uploads/2023/06/PROTOC1-1.pdf. Accessed November 13, 2023.
23. Collins GS, Reitsma JB, Altman DG, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. *BMC Med*. 2015;13:1.
24. Sullivan LM, Massaro JM, D'Agostino RB Sr. Presentation of multivariate data for clinical use: the Framingham Study risk score functions. *Stat Med*. 2004;23:1631–1660.
25. State Agency Official Gazette of the State. Royal Decree 1723/2012. Diagnosis of death based on circulatory and respiratory criteria. Available at https://www.boe.es/eli/es/rd/. Accessed November 13, 2023.
26. DONACIÓN EN ASISTOLIA EN ESPAÑA: SITUACIÓN ACTUAL Y RECOMENDACIONES, Documento de Consenso Nacional de la Organización Nacional de Trasplantes. (2012). Available at https://www.ont.es/wp-content/uploads/2023/06/Doc-de-Consenso-Nacional-sobre-Donacion-en-Asistolia.-Ano-2012.pdf. Accessed November 10, 2023.
27. Briceno J, Ciria R, de la Mata M. Donor-recipient matching: myths and realities. *J Hepatol*. 2013;58:811–820.
28. Bhat M, Rabindranath M, Chara BS, et al. Artificial intelligence, machine learning, and deep learning in liver transplantation. *J Hepatol*. 2023;78:1216–1233.
29. Ivanics T, So D, Claasen M, et al. Machine learning-based mortality prediction models using national liver transplantation registries are feasible but have limited utility across countries. *Am J Transplant*. 2023;23:64–71.
30. Kelly CJ, Karthikesalingam A, Suleyman M, et al. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med*. 2019;17:195.
31. Ruffle JK, Farmer AD, Aziz Q. Artificial intelligence-assisted gastroenterology- promises and pitfalls. *Am J Gastroenterol*. 2019;114:422–428.
32. Ho KJ, Owens CD, Johnson SR, et al. Donor postextubation hypotension and age correlate with outcome after donation after cardiac death transplantation. *Transplantation*. 2008;85:1588–1594.
33. Kalisvaart M, de Haan JE, Polak WG, et al. Onset of donor warm ischemia time in donation after circulatory death liver transplantation: hypotension or hypoxia? *Liver Transpl*. 2018;24:1001–1010.
34. Li L, Zhang Y, Xiao F, et al. Liver retransplantation: timing is equally important. *Medicine (Baltimore)*. 2023;102:e35165.