



Article

Forecasting Postoperative Delirium in Older Adult Patients with Fast-and-Frugal Decision Trees

Maria Heinrich ^{1,2,†}, Jan K. Woike ^{3,4,†}, Claudia D. Spies ¹ and Odette Wegwarth ^{1,4,5,*}

¹ Department of Anesthesiology and Operative Intensive Care Medicine (CCM, CVK), Charité-Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, 13353 Berlin, Germany

² Berlin Institute of Health@Charité (BIH), Anna-Louisa-Karsch 2, 10178 Berlin, Germany

³ School of Psychology, University of Plymouth, Plymouth PL4 8AA, UK

⁴ Max Planck Institute for Human Development, Center for Adaptive Rationality, 14195 Berlin, Germany

⁵ Heisenberg Chair for Medical Risk Literacy and Evidence-Based Decisions, Charité-Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, 10117 Berlin, Germany

* Correspondence: odette.wegwarth@charite.de; Tel.: +49-30-450-531-056; Fax: +49-30-450-551-909

† These authors contributed equally to this work.

Abstract: Postoperative delirium (POD) is associated with increased complication and mortality rates, particularly among older adult patients. However, guideline recommendations for POD detection and management are poorly implemented. Fast-and-frugal trees (FFTrees), which are simple prediction algorithms, may be useful in this context. We compared the capacity of simple FFTrees with two more complex models—namely, unconstrained classification trees (UDTs) and logistic regression (LogReg)—for the prediction of POD among older surgical patients in the perioperative setting. Models were trained and tested on the European BioCog project clinical dataset. Based on the entire dataset, two different FFTrees were developed for the pre-operative and postoperative settings. Within the pre-operative setting, FFTrees outperformed the more complex UDT algorithm with respect to predictive balanced accuracy, nearing the prediction level of the logistic regression. Within the postoperative setting, FFTrees outperformed both complex models. Applying the best-performing algorithms to the full datasets, we proposed an FFTree using four cues (Charlson Comorbidity Index (CCI), site of surgery, physical status and frailty status) for the pre-operative setting and an FFTree containing only three cues (duration of anesthesia, age and CCI) for the postoperative setting. Given that both FFTrees contained considerably fewer criteria, which can be easily memorized and applied by health professionals in daily routine, FFTrees could help identify patients requiring intensified POD screening.

Keywords: fast-and-frugal decision trees; postoperative outcomes; postoperative delirium; clinical data prediction; medical decision making



Citation: Heinrich, M.; Woike, J.K.; Spies, C.D.; Wegwarth, O. Forecasting Postoperative Delirium in Older Adult Patients with Fast-and-Frugal Decision Trees. *J. Clin. Med.* **2022**, *11*, 5629. <https://doi.org/10.3390/jcm11195629>

Academic Editor: Patrice Forget

Received: 12 August 2022

Accepted: 21 September 2022

Published: 24 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Postoperative delirium (POD) is an acute and sudden change in the mental state, characterized by fluctuating levels of attention, consciousness and cognition [1,2]. The occurrence of POD is associated with increased complication and mortality rates [3,4] and may be related to the development of long-term cognitive disorders [5–7]. Incidence depends on predisposing and precipitating risk factors [8–10] and ranges from 10–50% [11,12], and older people are particularly susceptible to POD [13].

Given the risks associated with undetected postoperative delirium, it is important to have tools available to detect POD reliably and in a timely manner. According to the recommendations of the evidence-based and consensus-based guidelines on postoperative delirium [11], screening for delirium should be performed once per shift, at least twice

a day, for 5 days after surgery in all patients, and predisposing and precipitating risk factors should be attenuated whenever possible. However, the implementation of these measures requires a considerable allocation of personnel resources and time, and it may thus come as no surprise that these guideline recommendations are poorly implemented [14]. Furthermore, a number of predictive models have been developed in the past to guide the prediction of POD; however, these also often require extensive assessment, and their clinical implications remain unclear [15–20]. These examples suggest that a detection tool that would be supportive of use in clinical care needs to be simple to keep the number of personnel and the time costs of the assessment as low as possible. Fast-and-frugal trees (FFTrees)—binarizing prediction algorithms based on limited information search—can provide such a simple structure and have demonstrated the capacity to facilitate accurate decisions in a variety of medical domains [21–25]. For instance, when predicting whether a patient presenting with chest pain should be admitted to the coronary care unit or to a normal ward, an FFTree consisting of only three yes-or-no questions performed comparably with a dedicated decision support tool (heart disease predictive instrument (HDPI)) requiring 50 pieces of information. These findings likely go against the common assumption that “more information is always better”, particularly in the medical domain, where most professionals may feel that, to make a good prediction or diagnosis, gathering more rather than less information reduces the risk of error. However, the relation between the amount of information and the quality of prediction is often an inverse U-shaped curve [26,27], specifically when situational uncertainty is high, as is the case in most medical situations including the prediction of POD. When situational uncertainty is high, model robustness is key [28,29]. Complex models, by using as much information as possible, fit “noise” and idiosyncrasies in the presented dataset that do not generalize to a new sample of patients. The result is “overfitting”, which conflicts with the robustness of a model and, thus, with the accuracy of prediction. Furthermore, it is important to note that the POD risk detected at admission (predisposing factors) increases substantially during the operation, and the impact of anesthesia and surgery (precipitating factors, such as trauma, stress, medication, depth of anesthesia, blood pressure fluctuations, transfusions) warrant the reassessment of risk. This means that models for POD must be adaptable and must include the conditions associated with surgery.

The aim of this work was to examine if FFTrees are able to sufficiently predict POD. To address the requirements of perioperative medicine, we built a pre-operative FFTree based on pre-operative parameters and further built a postoperative FFTree with modeling that additionally considered intraoperative parameters. Moreover, we compared the ability to predict unseen cases in the two FFTree construction methods [22]—which are based on limited information search—with those of two compensatory models; namely, unconstrained classification trees (based on the classification and regression trees (CART) algorithm) and logistic regression.

2. Materials and Methods

2.1. Overview of the Present Study

The work reported herein was performed on data initially acquired via the BioCog project, a prospective multicenter observational study conducted at the Charité–Universitätsmedizin Berlin, Department of Anesthesiology and Operative Intensive Care Medicine, Berlin, Germany, and the University Medical Center Utrecht, Department of Intensive Care Medicine, Utrecht, the Netherlands. This work was a secondary analysis performed for the purpose of generating FFTrees considering various influencing variables from the BioCog database generated at the study site Charité–Virchow Klinikum ($n = 394$, see Figure 1) in relation to the development of POD. The secondary analysis was approved by the local Ethics Committee (ref: EA2_048_18, 16 July 2020) and conducted in accordance with the Declaration of Helsinki.

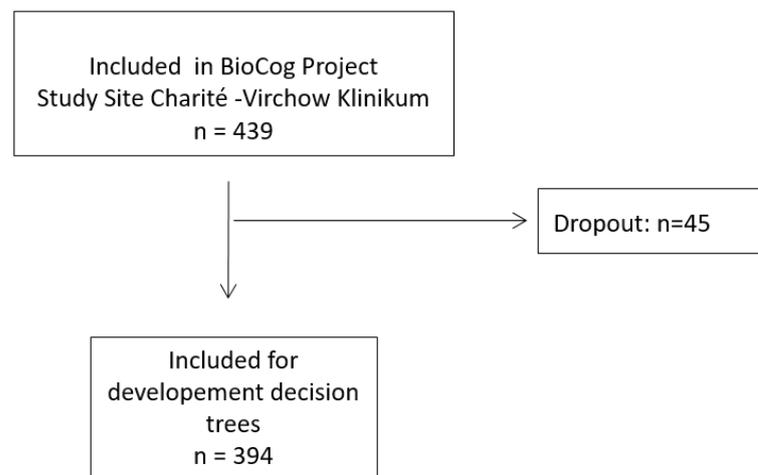


Figure 1. Flow chart.

The BioCog dataset was based on patients aged ≥ 65 years who were scheduled for elective surgery and presented with a Mini-Mental-State-Examination (MMSE) score of 23 points or higher (for detailed inclusion and exclusion criteria, see [30]).

2.2. Assessment of Postoperative Delirium

The models of this work inferred whether each respondent was at risk of POD as defined by the criteria of the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) [31]. Patients were considered delirious if they met any one of the following criteria:

- ≥ 2 cumulative points on the Nursing Delirium Screening Scale (Nu-DESC) and/or a positive Confusion Assessment Method (CAM) score;
- a positive CAM score for the Intensive Care Unit (CAM-ICU);
- a patient chart review that showed descriptions of delirium (e.g., confused, agitated, drowsy, disorientated, delirious, received antipsychotic therapy).

Delirium screening was started in the recovery room and repeated twice per day at 08:00 and 19:00 (± 1 h) for up to seven days after surgery. Delirium assessment was conducted independently of the routine hospital procedures by a research team that was trained and supervised by psychiatrists and delirium experts.

2.3. Cues

We aimed to develop two different models predicting a patient's POD status based on (i) pre-operative cues alone and (ii) both pre-operative and intraoperative cues.

For the pre-operative model, each model was meant to categorize a patient as being or not being at risk of POD based on the following cues: age; sex; body height; body mass index; physical status according to the American Society of Anesthesiologists (ASA PS); Charlson Comorbidity Index (CCI) [32]; comorbidities, such as arterial hypertension, coronary artery disease, diabetes mellitus, stroke or transient ischemic attack, in medical history; education according to International Standard Classification of Education (ISCED) [33]; MMSE; pre-operative cognitive impairment (for details, see Supplementary Information S1); impaired activities of daily living according to Barthel (ADL) [34], as well as Lawton and Brody (IADL) [35]; malnutrition according to the Mini-Nutritional Assessment (MNA) [36]; pre-operative frailty status (for details, see Supplementary Information S2); depression according to the geriatric depression scale (GDS) [37,38]; pre-operative long-term medication with benzodiazepines; hazardous alcohol consumption based on the AUDIT score [39]; current smoker status; pack years and site of surgery (intracranial vs. intrathoracic, intra-abdominal or pelvic vs. peripheral). The postoperative model used, in addition to these

pre-operative pieces of information, the duration of anaesthesia and the administration of premedication before surgery (benzodiazepines, clonidine, antihistaminergics, etc.).

2.4. Model Comparison

Two FFTree construction algorithms (the ifan algorithm (FFTi) and the dfan algorithm (FFTd)) [22] were compared with logistic regression and an unconstrained classification tree algorithm (UDT) based on CART [39] for the pre-operative and the complete dataset separately. We chose a maximum number of five cues for the FFTi algorithm and a maximum number of four cues for the FFTd algorithm. The criteria for ifan and dfan were set to balanced accuracy. For UDT, we weighted misclassifications of positive cases higher than the misclassification of negative cases (based on the ratio of negative to positive cases in the training set) to aim for good performance in terms of balanced accuracy. We used the rpart package in R for UDT [40], which implements most of the CART algorithms [39] (with the minimum splitting size set to 20 and the complexity parameter to 0.00001). For the binary logistic regression model (LogReg), cues that were provided to the corresponding tree models were included in the regression model. To target the criterion of balanced accuracy, we set the threshold to transform probability estimates into predictions to the base rate observed in the training set. We used the implementation of logistic regression in the glm command in R.

2.4.1. Training and Test Set

In order to estimate the predictive performance of each model, the dataset was repeatedly randomly split into training and test (prediction set) sets, with an equal number of cases in each. Trees were constructed and parameters estimated based on the training set, and performance was measured based on the test set alone. The performance measure used was balanced accuracy—the mean of sensitivity and specificity—and models were estimated with the aim of achieving high values on this measure. Based on the model comparison, a tree construction algorithm was chosen to build two final trees (pre- and postoperative) based on the full dataset ($n = 394$). All analyses were run in R (version 4.1.2) [41].

2.4.2. Model Comparison Procedure

In preparing the dataset, missing values were replaced before starting the model comparison. For the following variables, a missing value was replaced by the sample median: ISCED, GDS, pack years, duration of anaesthesia. For categorical variables, missing values were replaced by the mode, which was 0 (no impairment) in the case of arterial hypertension, coronary artery disease, diabetes mellitus, stroke or transient ischemic attack in medical history, pre-operative cognitive impairment, ADL, IADL and the administration of premedication before surgery and 3 (no impairment) in the case of MNA.

In each trial of the model comparison, the full dataset was randomly split into training and test sets with an equal number of cases ($n = 197$). Models were estimated using the training set and performance (sensitivity, specificity and balanced accuracy) was measured using the test set alone. This procedure was repeated 1000 times for the more time-intensive FFTree construction algorithm (FFTd) and 10,000 times for all others. In each trial, the same training-test split was applied for each model, with FFTd being restricted to the first 1000 splits.

3. Results

3.1. Patient Characteristics

Altogether, we used data for 394 older adult surgical patients; 99 patients (25.1%) fulfilled the criteria for POD (see Table 1 for patient characteristics).

Table 1. Patient characteristics (n = 394).

Characteristic	POD (n = 99) 25.10%	Non POD (n = 295) 74.90%	<i>p</i>
	n = 394		
Age (years)	74 [71;77]	72 [68;76]	0.004 ^a
Sex			
Female	51 (41.5%)	145 (49.2%)	0.684 ^b
ASA PS			
1–2	43 (43.4%)	203 (68.8%)	<0.001 ^b
3–4	56 (56.6%)	92 (31.2%)	
Charlson Comorbidity Index	2.14 ± 1.5 ^c	1.43 ± 1.6 ^c	<0.001 ^a
Frailty status			
Pre-frail	50 (51.0%)	143 (49.1%)	<0.001 ^b
frail	30 (30.6%)	30 (10.3%)	
Site of surgery			
intracranial	2 (2.0%)	7 (2.4%)	<0.001 ^b
intrathoracic, intra-abdominal or pelvic	67 (67.7%)	122 (41.4%)	
peripheral	30 (30.3%)	166 (56.3%)	
Duration of Anaesthesia (min)	360 [220;495]	157 [100;260]	<0.001 ^b

Data are expressed as medians [25th quartile; 75th quartile] except for categorical data, which are expressed as frequencies (percentages). *p*-values are with respect to Mann–Whitney U test (^a) or Chi-squared test (^b) between patients with or without POD. Additionally, data for Charlson Comorbidity Index are presented as means ± SD (^c). *p* ≤ 0.05 was considered as statistically significant. SD—standard deviation, POD—postoperative delirium, ASA PS—physical status according to the American Society of Anesthesiologists, min—minutes.

3.2. Model Comparison

For each model comparison, we report the mean performance of the four algorithms for the training and test sets and present visualizations of the distribution of prediction results across trials.

3.2.1. Performance of Pre-Operative Models

The model performance for all four models is summarized in Table 2. Performance for the training set represents the ability to predict criterion values based on already known cue values. The table reports sensitivity, specificity and balanced accuracy (the average of sensitivity and specificity).

Unconstrained decision trees exhibited the best average balanced accuracy (0.803), followed by logistic regression and the two FFTree models. The difference between the best and worst model was over 0.11. The standard error for FFTd models was higher due to the smaller number of trials (1000 vs. 10,000). It should be noted that the standard error for balanced accuracy was lower than that for sensitivity and specificity: Models tended to trade off sensitivity against specificity across trials, resulting in more stable values for the average. The performance of unconstrained decision trees suffered the most, changing the order of performance in the test set.

The distribution of results across trials (see Figure 2) demonstrated the variability across trials and put the average differences into context. Further analysis showed that the LogReg model outperformed the FFTi model in 76.0% of the trials, but it was outperformed by the FFTi model in 23.9% of the trials.

Table 2. Average performance for algorithms across trials with pre-operative information. The table presents means and standard errors separately for the four algorithms with the training set (fitting) and test set (prediction). For each combination, the table reports the mean and standard error of the mean for sensitivity, specificity and balanced accuracy across trials. FFTi, UDT and LogReg were tested in 10,000 trials and FFTd in the first 1000 of these only.

		Training			Prediction		
		Sensitivity	Specificity	Bal. Accuracy	Sensitivity	Specificity	Bal. Accuracy
FFTi	M	0.693	0.682	0.688	0.578	0.644	0.611
	SE	(0.0013)	(0.0013)	(0.0002)	(0.0015)	(0.0014)	(0.0003)
FFTd	M	0.751	0.689	0.720	0.562	0.625	0.593
	SE	(0.0037)	(0.0037)	(0.0007)	(0.0044)	(0.0040)	(0.0011)
UDT	M	0.868	0.738	0.803	0.52	0.626	0.573
	SE	(0.0006)	(0.0006)	(0.0002)	(0.0010)	(0.0007)	(0.0004)
LogReg	M	0.737	0.747	0.742	0.581	0.692	0.637
	SE	(0.0004)	(0.0003)	(0.0003)	(0.0007)	(0.0004)	(0.0003)

M—mean, SE—standard error of the mean, FFTi—fast-and-frugal tree construction using the ifan algorithm, FFTd—fast-and-frugal tree construction using the dfan algorithm, UDT—unconstrained decision trees based on the CART algorithm, LogReg—logistic regression.

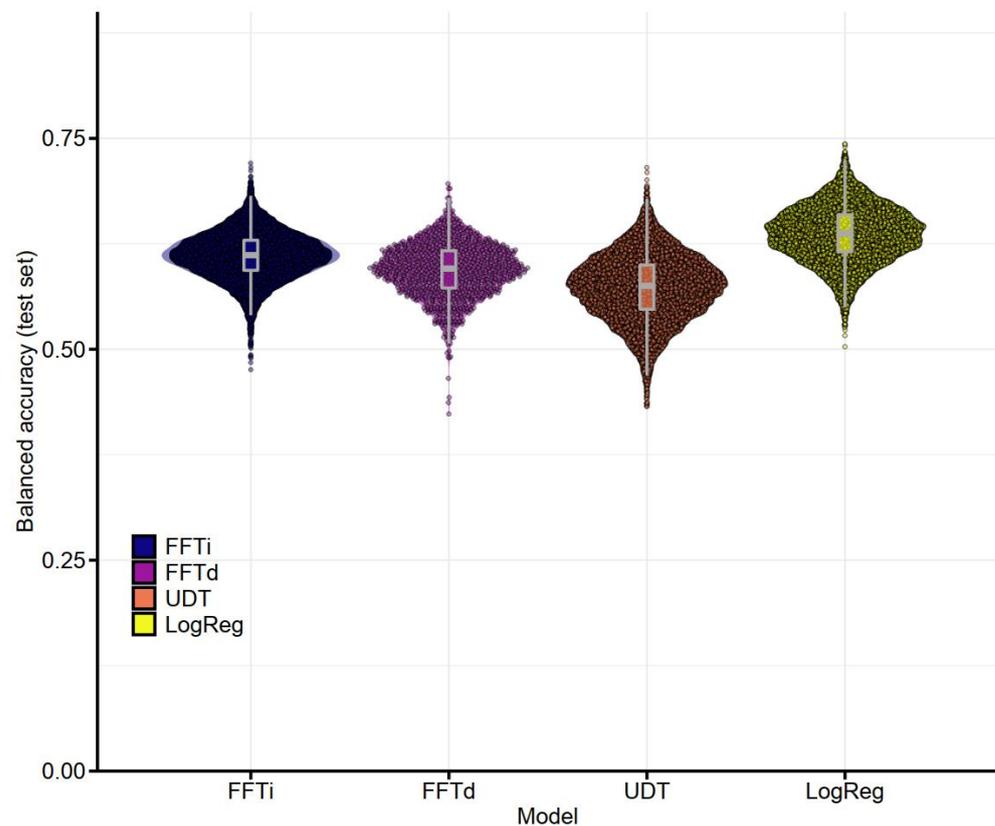


Figure 2. Balanced accuracy in the test set (prediction task) for the four algorithms across trials with pre-operative cues only. The graph shows boxplots and violin plots, with dots representing the results of individual trials (10,000 trials for FFTi, UDT and LogReg; 1000 trials for FFTd). FFTi—fast-and-frugal tree construction using the ifan algorithm, FFTd—fast-and-frugal tree construction using the dfan algorithm, UDT—unconstrained decision tree based on the CART algorithm, LogReg—logistic regression.

3.2.2. Performance of Postoperative Models

Adding the postoperative variables improved the performance of all models, both in fitting the training set and predicting the test set (see Table 3).

Table 3. Average performance of algorithms across trials with full information. The table presents means and standard errors separately for the four algorithms with the training set (fitting) and test set (prediction). For each combination, the table reports the mean and standard error of the mean for sensitivity, specificity and balanced accuracy across trials. FFTi, UDT and LogReg were tested in 10,000 trials and FFTd in the first 1000 of these only.

		Training			Prediction		
		Sensitivity	Specificity	Bal. Accuracy	Sensitivity	Specificity	Bal. Accuracy
FFTi	M	0.767	0.723	0.745	0.698	0.695	0.696
	SE	(0.0008)	(0.0008)	(0.0002)	(0.0011)	(0.0009)	(0.0003)
FFTd	M	0.815	0.756	0.786	0.698	0.71	0.704
	SE	(0.0025)	(0.0024)	(0.0007)	(0.0032)	(0.0028)	(0.0010)
UDT	M	0.896	0.784	0.840	0.625	0.694	0.660
	SE	(0.0005)	(0.0005)	(0.0002)	(0.0010)	(0.0007)	(0.0004)
LogReg	M	0.792	0.803	0.798	0.632	0.749	0.690
	SE	(0.0004)	(0.0003)	(0.0003)	(0.0007)	(0.0004)	(0.0003)

M—mean, SE—standard error of the mean, FFTi—fast-and-frugal tree construction using the ifan algorithm, FFTd—fast-and-frugal tree construction using the dfan algorithm, UDT—unconstrained decision tree based on the CART algorithm, LogReg—logistic regression.

Again, the order of performance changed between fitting and prediction, with unconstrained decision trees showing the best fitting performance (0.840) and the worst prediction performance (0.660). In contrast to the previous comparison, the FFTrees outperformed both alternative models in prediction. The FFTd algorithm showed a better average balanced accuracy in prediction (0.704) than the FFTi algorithm (0.696). FFTi performed better than LogReg in 56.1% of the trials, and FFTd performed better than LogReg in 61.9% of the trials and better than FFTi in 56.2% of the trials. The distribution of balanced accuracy in the prediction task is shown in Figure 3. Thus, these differences were not generated by outliers but by a general tendency to outperform the competitors.

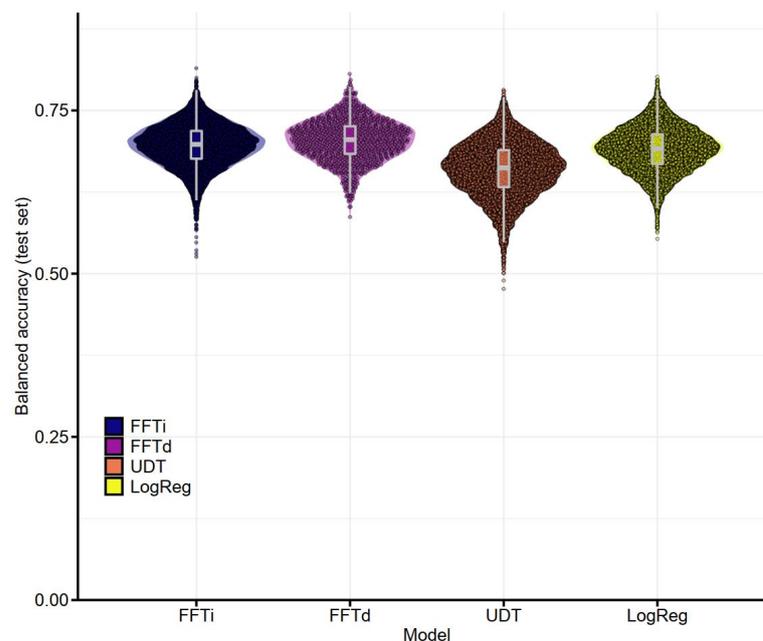


Figure 3. Balanced accuracy in the test set (prediction task) for the four algorithms across trials with all cues: The graph shows boxplots and violin plots, with dots representing the results of individual trials (10,000 trials for FFTi, UDT and LogReg; 1000 trials for FFTd). FFTi—fast-and-frugal tree construction using the ifan algorithm, FFTd—fast-and-frugal tree construction using the dfan algorithm, UDT—unconstrained decision tree based on the CART algorithm, LogReg—logistic regression.

3.3. Decision Trees Based on Full Dataset

Based on the results of the model comparison, we estimated two fast-and-frugal trees based on the full dataset, one for the pre-operative set of variables and one for the complete, postoperative set of variables. We present visualizations of the resulting trees and report performance statistics based on the full dataset (n = 394).

3.3.1. Pre-Operative FFTree

The first (pre-operative decision tree) was based on pre-operative data. Following the results of the model comparison, we chose the ifan algorithm to construct the tree. The resulting pre-operative decision tree contained four cues (CCI, site of surgery, ASA PS and frailty status) and indicated a sensitivity of 0.84 and a specificity of 0.46 with a balanced accuracy of 0.65 (see Figure 4). On average, 1.8 cue values had to be looked up to make a decision for the cases in the dataset, and 93% of the provided information was ignored on average. Over 52% of all cases were classified as positive after the first question (with a CCI larger than 1), and 71 of the 99 positive cases were in this group. The tree achieved a higher sensitivity at the cost of specificity (the unweighted accuracy was 0.56).

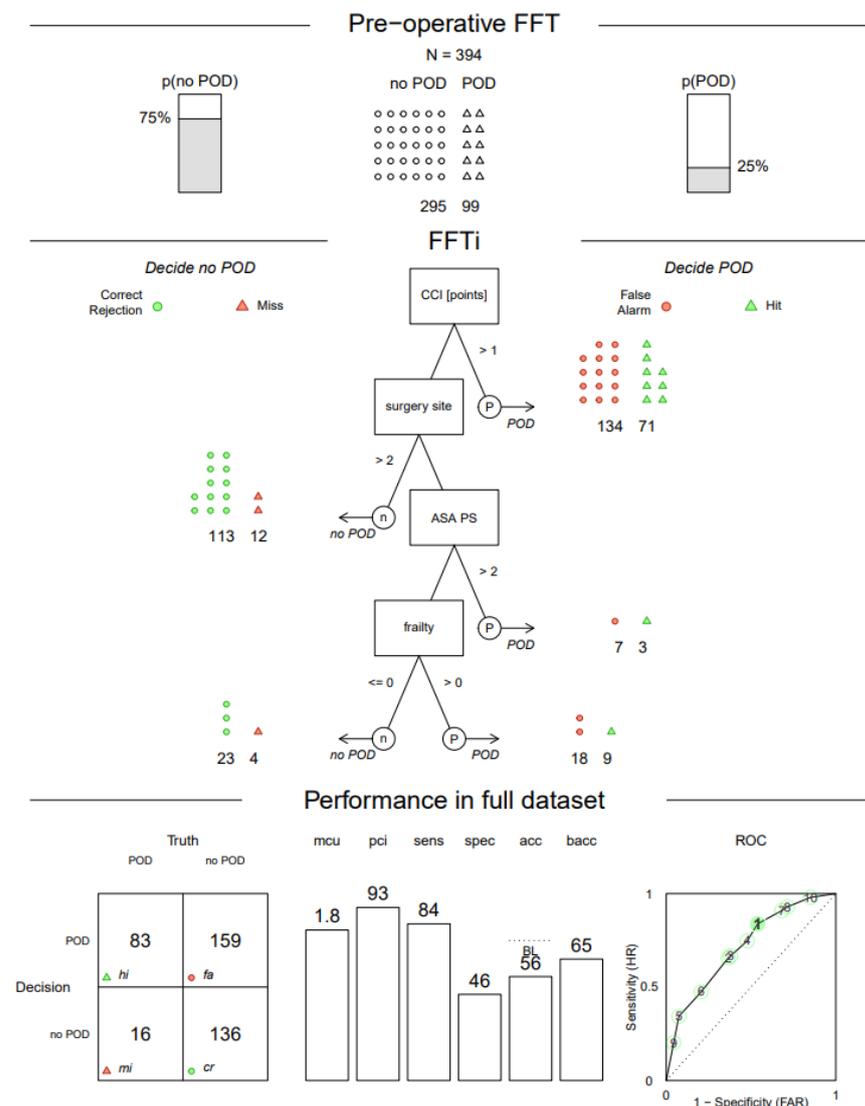


Figure 4. Pre-operative fast-and-frugal tree estimated with the ifan algorithm. POD—postoperative delirium, p(POD)—probability of POD a priori (base rate), p (no POD)—complement of p(POD),

FFTi—fast-and-frugal tree construction using the ifan algorithm, CCI—Charlson Comorbidity Index, surgery site > 2—peripheral, ASA PS—physical status according to the American Society of Anesthesiologists, frailty <= 0—robust, frailty > 0—pre-frail/frail, mcu—mean cues used, pci—percent cues ignored, sens—sensitivity, spec—specificity, acc—unweighted accuracy, bacc—balanced accuracy (sensitivity + specificity)/2, BL—probable BL (the base rate of 75% that could be achieved by classifying all cases as negative), ROC—receiver operating characteristic (shows the performance of all compared trees using the same cue order numbered according to their resulting balanced accuracy (in the training set), each data point shows the false alarm rate (FAR) on the x-axis and sensitivity/hit rate (HR) on the y-axis), hi—hit, mi—miss, cr—correct rejection.

3.3.2. Postoperative FFTree

The construction of the second decision tree (postoperative decision tree) took intraoperative parameters into account in addition to the pre-operative data. Following the results of the model comparison, we chose the dfan algorithm for tree construction in this case. The postoperative decision tree contained three cues (duration of anesthesia, age and CCI). While the maximum depth was set to four cues, the algorithm did not find an improvement by adding an additional layer to the tree, generating a truncated tree (see also [29]).

The decision tree demonstrated a sensitivity of 0.81, a specificity of 0.72 and a balanced accuracy of 0.76 (see Figure 5). It used 2.1 cues on average to make a classification, and it ignored 92% of the provided information on average. The tree was more balanced than the pre-operative tree and achieved an unweighted accuracy of 0.74.

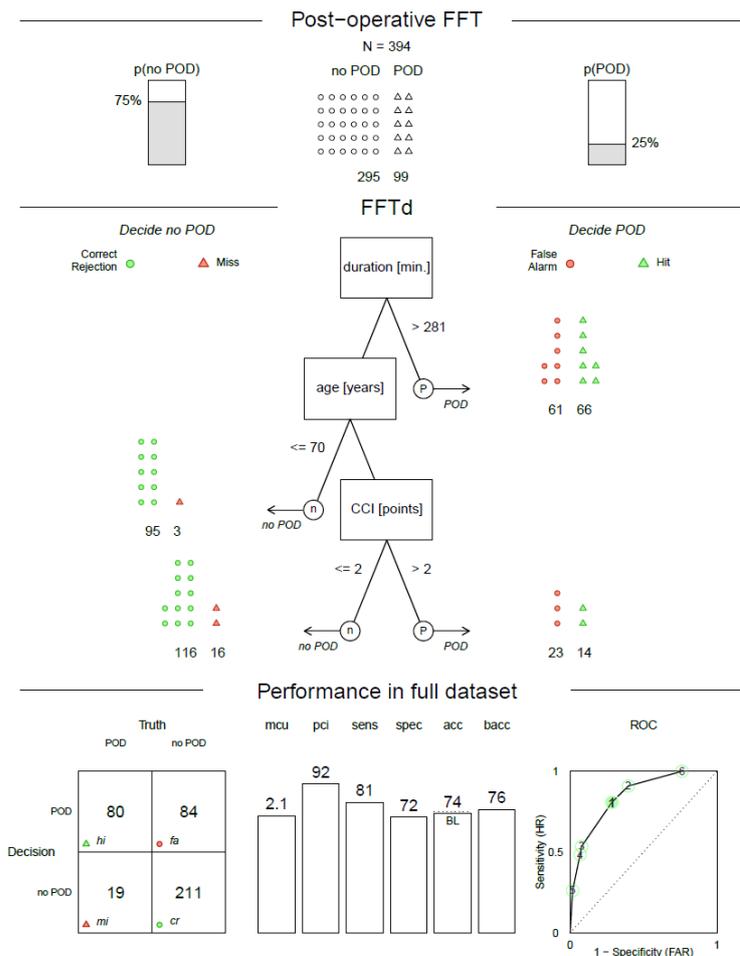


Figure 5. Postoperative fast-and-frugal tree estimated with the dfan algorithm. POD—postoperative delirium, p(POD)—probability of POD a priori (base rate), p (no POD)—complement of p(POD),

FFTd—fast-and-frugal tree construction using the dfan algorithm, duration—duration of anesthesia, CCI—Charlson Comorbidity Index, mcu—mean cues used, pci—percent cues ignored, sens—sensitivity, spec—specificity, acc—unweighted accuracy, back—balanced accuracy (sensitivity + specificity)/2, BL—probable BL (the base rate of 75% that could be achieved by classifying all cases as negative), ROC—receiver operating characteristic (shows the performance of all compared trees using the same cue order numbered according to their resulting balanced accuracy (in the training set), each data point shows the false alarm rate (FAR) on the *x*-axis and sensitivity/hit rate (HR) on the *y*-axis), hi—hit, mi—miss, cr—correct rejection.

4. Discussion

The aim of this work was to develop decision trees that can be used to estimate the risk of developing POD both pre-operatively and postoperatively in older adult patients. We were able to create two decision trees that differed in the parameters included. The pre-operative decision tree contained four cues (CCI, site of surgery, ASA PS and frailty status) and the postoperative contained three cues (duration of anesthesia, age and CCI). Before estimating fast-and-frugal trees (FFTrees), we compared two methods of FFTree construction (the ifan algorithm (FFTi) and the dfan algorithm (FFTd)) with unconstrained classification trees (UDTs, based on CART) and logistic regression. Fast-and-frugal trees are minimal binary classification trees that are constrained in terms of their structure. Various algorithms have been proposed for the construction of fast-and-frugal trees [22,25,29,42]. Here, we chose two algorithms that have been proved most competitive in achieving a high balanced accuracy [22], the ifan algorithm (FFTi) and the dfan algorithm (FFTd). A natural comparison for highly constrained fast-and-frugal trees are unconstrained classification trees (UDTs). Furthermore, we compared FFTrees with binary logistic regression models (LogReg), which predict the probability of an older adult patient being at risk for POD based on a weighted integration of all provided cues. It should be noted that, due to the relatively smaller number of positive cases in the sample, models aimed at achieving a high unweighted accuracy would, in contrast, likely sacrifice sensitivity for specificity, which would not be in line with the aims for the decision tool.

The results were in line with previous model comparisons [22,25,29]. More flexible models generally outperform less flexible models in this type of fitting performance. In line with this, unconstrained decision trees exhibited the best average balanced accuracy in the training set in the pre-operative model comparison, followed by logistic regression and the two FFTree models. The distribution of results across trials demonstrated the variability across trials and put the average differences into context. In the testing set with pre-operative modeling, logistic regression showed the highest balanced accuracy, closely followed by fast-and-frugal trees constructed with the ifan algorithm. All models showed worse performance in the testing set when predicting cases that were not part of the sample used to estimate their parameters. The inflation of predictive accuracy when predicting familiar cases has also been termed “overfitting”, and it is usually more pronounced in more complex and flexible models. As expected, the performance of unconstrained decision trees suffered the most, changing the order of performance in the testing set.

Adding the postoperative variables improved the performance of all models, both in fitting the training set and predicting the test set. Again, the order of performance changed between fitting and prediction, with unconstrained decision trees showing the best fitting performance and the worst prediction performance. In contrast to the previous comparison, the fast-and-frugal trees outperformed both alternative models in prediction. Based on these model comparisons, we chose the ifan algorithm for pre-operative testing and the dfan algorithm for FFTree construction. It should be noted that the model comparison used 50% of the full dataset, providing a training sample that was smaller than the full sample. The advantage of this method is that training samples were less correlated. However, larger sample sizes tend to make logistic regression and CART more competitive. Based on our results for model comparison, we argue that FFTrees performed similarly to logistic regression and were not necessarily superior. From our point of view, the choice of method could be

guided by the availability of data: the more data there are, the better the case for using more complex methods to achieve better predictive capability. In smaller (and often typical) datasets, the case is stronger for FFTrees. Moreover, FFTrees are simpler to apply, easier to communicate and requires less information, but they are competitive in predicting cases.

For all the cues included in the FFTrees, there is strong evidence in the literature [11,43–45] that they are independent risk factors in the development of POD and have to be considered in perioperative care according to guideline recommendations. The fact that age was not considered in the pre-operative decision tree may reflect the relevance of biological age rather than chronological age. In the pre-operative decision trees, this is represented by frailty status. Duration of anesthesia had a strong impact in decision tree development. It can be regarded as a surrogate for extent of surgery and associated inflammation, toxicity of anesthesia or intraoperative complications, such as bleeding or organ damage. All of these factors influence the risk for developing POD. The evaluation of this simple surrogate (duration of anesthesia) is significant for clinical applicability.

To the best of our knowledge, this is the first time decision trees were created for the risk stratification of POD. In the past, more emphasis has been placed on developing predictive models with clinical implications that remain unclear. A number of prediction models have been developed but primarily to predict delirium risk (not POD risk) [17] or for ICU patients [15,18,20,46]. A simple translation to surgical patients is problematic, as the needs of surgical patients are not addressed. Although surgical patients have a baseline risk of developing POD, surgery is such a relevant incident that it requires reassessment of risk. This means that models for POD must be adaptable and must include the conditions associated with surgery.

The oldest prediction model for delirium, developed in the 1990s, included the evaluation of vision and cognitive impairment, severe illness and high urea nitrogen/creatinine ratio [17]. Prediction models for ICU delirium include—besides age parameters, which are primarily related to intensive care treatment, such as for coma—use of sedatives and morphine [46], respiratory failure [20], vasoactive medication use and requirement of continuous renal replacement therapy and mechanical ventilation [15]. Some of them were developed through retrospective analysis [15,18].

There are three noteworthy studies that address modeling for POD risk prediction that have been recently published. A nine-item model for predicting POD risk with an area under the curve (AUC) of 0.77 was developed in a cohort of patients with acute hip fracture [47]. In this model, ASA PS was also considered, as well as functional dependence and pre-operative use of mobility aids, which are surrogates for frailty. Although this model was based on an extensive dataset, there was a crucial limitation, as POD was not determined prospectively but retrospectively by means of a chart review. Another seven-item model with an AUC of 0.82 was developed in older adult orthopedic patients in the ICU [19]. POD was determined prospectively, which reduced the sample size accordingly. Here, intraoperative parameters could be considered. For example, in addition to age, major hemorrhage was also included in the score. This score appears promising but also has the limitations that it is a very specific patient group, not all of the parameters are routine parameters (level of interleukin-6) and pre-operative application of the score is not appropriate (three intraoperative parameters are included). These scores have in common the limitation that eight or nine criteria may be too extensive for routine daily use. In this regard, a score based on four items with an AUC of 0.83 for cardiac surgery patients is more feasible for clinical application [16]. This score includes age, evaluation of MMSE, insomnia needing medical treatment and low physical activity, which is equivalent to one item in our frailty definition. Here, an automatic calculator, which calculates the risk, is actually available. Nevertheless, intraoperative parameters are not considered in this approach either.

We were able to solve this challenge by developing both a tree for pre-operative use and a tree for postoperative use. Furthermore, the datasets for development of the decision trees included both parameters for which there is strong evidence regarding the association with POD and parameters for which there are only hypotheses. With our work,

we were able to select the most relevant parameters and rank them. It has been shown that only parameters with existing strong evidence were relevant in our analysis. Sensitivity appeared to be adequate in both trees. Specificity was very low in the pre-operative tree at 0.46. This does not matter considering how these trees will be applied in clinical care. In theory, all patients should receive a complete delirium screening and predisposing and precipitating factors should be anticipated. Since this is not implemented across the board for various reasons, the application of decision trees was intended to fill this gap, at least for those patients who should receive screening and special attention in any case. There is no disadvantage in giving special attention pre-operatively to a probably higher number of patients with misclassified increased risk of delirium. The postoperative decision tree has a higher specificity, which is relevant for clinical application, since POD screening requires personnel resources that are supposed to be extensive. The next steps include validation of the decision trees and verification of clinical practicability.

Strength and Limitations

A key strength of this study is the prospective design of the POD assessment. POD was assessed through a comprehensive, standardized and validated assessment. While the use of routine data for modeling, as in many of the previously discussed models, has the great advantage that extensive datasets are available, this must be viewed very critically, especially in the case of the clinical picture of postoperative delirium. Especially due to its fluctuating characteristics and the frequent occurrence of hypoactive forms, a comprehensive validated screening is essential. As described above, there is a large gap in POD coverage, so it cannot be assumed that POD screening is implemented adequately. This raises doubts about the quality of the analysis of routine data.

The study database contained extensive information on both parameters for which there is strong evidence regarding the association with POD and parameters for which there are only hypotheses. In addition, we were able to develop the decision trees based on a dataset of patients covering a wide range of surgical disciplines (see Supplementary Table S1), which reflects the conditions that apply to the perioperative risk evaluation settings in clinical practice and gives the translation approach a bit more feasibility.

Nevertheless, some important limitations must be considered. Even though our dataset was very high quality and extensive, the sample size was small. Therefore, the model comparison used 50% of the full dataset, providing a training sample that was smaller than the full sample. Larger sample sizes tend to make unconstrained classification trees and logistic regression more competitive whereas smaller sample sizes do the opposite. Therefore we would like to argue that the choice of method should be guided by the availability of data: the more data there are, the better the case for using more complex methods to achieve a better predictive capability. In smaller (and often typical) datasets, the case is stronger for the FFTrees. Finally, we were able to provide initial insights with our analysis, but these still need to be validated.

5. Conclusions

Within the pre-operative setting, FFTrees outperformed the more complex UDT algorithm with respect to their predictive balanced accuracy, nearing the prediction level of logistic regression. Within the postoperative setting, FFTrees outperformed both complex models. Applying the best-performing algorithms to the full datasets we propose an FFTree using four cues (CCI, site of surgery, ASA PS and frailty status) for the pre-operative setting and an FFTree containing only three cues (duration of anesthesia, age and CCI) for the postoperative setting. Given that both FFTrees contain considerably fewer criteria, which can be easily memorized and applied by health professionals in daily routine, FFTrees could help identify patients requiring intensified POD screening.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/jcm11195629/s1>, Supplementary Information S1: Assessment of

pre-operative cognitive impairment. Supplementary Information S2: Assessment of frailty status. Supplementary Information S3: Further information on the models applied. Supplementary Table S1: Overview of surgical sites of the surgical procedures (n = 394). References [48–54] are cited in the supplementary materials.

Author Contributions: M.H. was involved in data collection and plausibility checks. M.H., C.D.S. and O.W. developed the research question. J.K.W. and O.W. carried out the analyses. The first draft of the manuscript was written by M.H., J.K.W. and O.W., and all authors commented on previous versions of the manuscript. All authors have read and agreed to the published version of the manuscript. A subset of the results was presented by M.H. at the Conference of the German Society of Anti-Aging Medicine e.V. 2021 as an oral presentation.

Funding: The BioCog project (Biomarker Development for Postoperative Cognitive Impairment in the Elderly) from which the data were initially acquired was funded by the European Union Seventh Framework Programme under grant agreement No. 602461.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Ethics Committee of Charité–Universitätsmedizin Berlin (ref: EA2_048_18 16 July 2020).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data available on request due to privacy restrictions.

Acknowledgments: The authors would like to thank all participants who contributed to the preparation of the comprehensive dataset. Maria Heinrich is a participant of the BIH Charité Digital Clinician Scientist Program funded by the Charité–Universitätsmedizin Berlin and the Berlin Institute of Health at Charité (BIH).

Conflicts of Interest: The authors declare no competing interests.

References

- Hewer, W.; Thomas, C.; Drach, L.M. *Delirium beim Alten Menschen*; W. Kohlhammer GmbH: Stuttgart, Germany, 2016; Volume 1.
- Rengel, K.F.; Pandharipande, P.P.; Hughes, C.G. Postoperative delirium. *Presse Med.* **2018**, *47 Pt 2*, e53–e64. [[CrossRef](#)] [[PubMed](#)]
- Bickel, H.; Gradinger, R.; Kochs, E.; Forstl, H. High risk of cognitive and functional decline after postoperative delirium. A three-year prospective study. *Dement. Geriatr. Cogn. Disord.* **2008**, *26*, 26–31. [[CrossRef](#)] [[PubMed](#)]
- Moskowitz, E.E.; Overbey, D.M.; Jones, T.S.; Jones, E.L.; Arcomano, T.R.; Moore, J.T.; Robinson, T.N. Post-operative delirium is associated with increased 5-year mortality. *Am. J. Surg.* **2017**, *214*, 1036–1038. [[CrossRef](#)] [[PubMed](#)]
- Davis, D.H.; Muniz Terrera, G.; Keage, H.; Rahkonen, T.; Oinas, M.; Matthews, F.E.; Cunningham, C.; Polvikoski, T.; Sulkava, R.; MacLulich, A.M.; et al. Delirium is a strong risk factor for dementia in the oldest-old: A population-based cohort study. *Brain A J. Neurol.* **2012**, *135*, 2809–2816. [[CrossRef](#)]
- Daiello, L.A.; Racine, A.M.; Yun Gou, R.; Marcantonio, E.R.; Xie, Z.; Kunze, L.J.; Vlassakov, K.V.; Inouye, S.K.; Jones, R.N.; Alsop, D.; et al. Postoperative Delirium and Postoperative Cognitive Dysfunction: Overlap and Divergence. *Anesthesiology* **2019**, *131*, 477–491. [[CrossRef](#)]
- Sprung, J.; Roberts, R.O.; Weingarten, T.N.; Nunes Cavalcante, A.; Knopman, D.S.; Petersen, R.C.; Hanson, A.C.; Schroeder, D.R.; Warner, D.O. Postoperative delirium in elderly patients is associated with subsequent cognitive impairment. *Br. J. Anaesth.* **2017**, *119*, 316–323. [[CrossRef](#)]
- Wang, J.; Li, Z.; Yu, Y.; Li, B.; Shao, G.; Wang, Q. Risk factors contributing to postoperative delirium in geriatric patients postorthopedic surgery. *Asia-Pac. Psychiatry Off. J. Pac. Rim Coll. Psychiatr.* **2015**, *7*, 375–382. [[CrossRef](#)]
- Berian, J.R.; Zhou, L.; Russell, M.M.; Hornor, M.A.; Cohen, M.E.; Finlayson, E.; Ko, C.Y.; Rosenthal, R.A.; Robinson, T.N. Postoperative Delirium as a Target for Surgical Quality Improvement. *Ann. Surg.* **2017**, *268*, 93–99. [[CrossRef](#)]
- Inouye, S.K. Predisposing and precipitating factors for delirium in hospitalized older patients. *Dement. Geriatr. Cogn. Disord.* **1999**, *10*, 393–400. [[CrossRef](#)]
- Aldecoa, C.; Bettelli, G.; Bilotta, F.; Sanders, R.D.; Audisio, R.; Borozdina, A.; Cherubini, A.; Jones, C.; Kehlet, H.; MacLulich, A.; et al. European Society of Anaesthesiology evidence-based and consensus-based guideline on postoperative delirium. *Eur. J. Anaesthesiol.* **2017**, *34*, 192–214. [[CrossRef](#)]
- Ho, M.H.; Nealon, J.; Igwe, E.; Traynor, V.; Chang, H.R.; Chen, K.H.; Montayre, J. Postoperative Delirium in Older Patients: A Systematic Review of Assessment and Incidence of Postoperative Delirium. *Worldviews Evid. Based Nurs.* **2021**, *18*, 290–301. [[CrossRef](#)] [[PubMed](#)]
- Maldonado, J.R. Neuropathogenesis of delirium: Review of current etiologic theories and common pathways. *Am. J. Geriatr. Psychiatry Off. J. Am. Assoc. Geriatr. Psychiatry* **2013**, *21*, 1190–1222. [[CrossRef](#)]

14. Saller, T.; Hofmann-Kiefer, K. Kenntnis und Umsetzung der S3-Leitlinie zum Delirmanagement in Deutschland. *Anaesthesist* **2016**, *65*, 755–762. [[CrossRef](#)] [[PubMed](#)]
15. Cherak, S.J.; Soo, A.; Brown, K.N.; Ely, E.W.; Stelfox, H.T.; Fiest, K.M. Development and validation of delirium prediction model for critically ill adults parameterized to ICU admission acuity. *PLoS ONE* **2020**, *15*, e0237639. [[CrossRef](#)] [[PubMed](#)]
16. de la Varga-Martínez, O.; Gómez-Pesquera, E.; Muñoz-Moreno, M.F.; Marcos-Vidal, J.M.; López-Gómez, A.; Rodenas-Gómez, F.; Ramasco, F.; Álvarez-Refojo, F.; Tamayo, E.; Gómez-Sánchez, E. Development and validation of a delirium risk prediction preoperative model for cardiac surgery patients (DELIPRECA): An observational multicentre study. *J. Clin. Anesth.* **2021**, *69*, 110158. [[CrossRef](#)] [[PubMed](#)]
17. Inouye, S.K.; Viscoli, C.M.; Horwitz, R.I.; Hurst, L.D.; Tinetti, M.E. A predictive model for delirium in hospitalized elderly medical patients based on admission characteristics. *Ann. Intern. Med.* **1993**, *119*, 474–481. [[CrossRef](#)]
18. Pagali, S.R.; Miller, D.; Fischer, K.; Schroeder, D.; Egger, N.; Manning, D.M.; Lapid, M.I.; Pignolo, R.J.; Burton, M.C. Predicting Delirium Risk Using an Automated Mayo Delirium Prediction Tool: Development and Validation of a Risk-Stratification Model. *Mayo Clin. Proc.* **2021**, *96*, 1229–1235. [[CrossRef](#)]
19. Wang, G.; Zhang, L.; Qi, Y.; Chen, G.; Zhou, J.; Zhu, H.; Hao, Y. Development and Validation of a Postoperative Delirium Prediction Model for Elderly Orthopedic Patients in the Intensive Care Unit. *J. Healthc. Eng.* **2021**, *2021*, 9959077. [[CrossRef](#)]
20. Wassenaar, A.; van den Boogaard, M.; van Achterberg, T.; Slooter, A.J.; Kuiper, M.A.; Hoogendoorn, M.E.; Simons, K.S.; Maseda, E.; Pinto, N.; Jones, C.; et al. Multinational development and validation of an early prediction model for delirium in ICU patients. *Intensive Care Med.* **2015**, *41*, 1048–1056. [[CrossRef](#)]
21. Green, L.; Mehr, D. What alters physicians' decisions to admit to the coronary care unit? *J. Fam. Pract.* **1997**, *45*, 219–226.
22. Phillips, N.; Neth, H.; Woike, J.K.; Gaissmaier, W. FFTrees: A toolbox to create, visualize, and evaluate fast-and-frugal decision trees. *Judgm. Decis. Mak.* **2017**, *12*, 344–368.
23. Wegwarth, O. Deciding the Fast & Frugal Way on the Application of Pharmacodiagnostic Tests in Cancer Care? A Comparative Study of Oncologists', Pathologists', and Cancer Patients' Decision Making in Germany and the USA. Ph.D. Thesis, Humboldt University, Berlin, Germany, 2007.
24. Wegwarth, O.; Day, R.W.; Gigerenzer, G. Decisions on pharmacogenomic tests in the USA and Germany. *J. Eval. Clin. Pract.* **2011**, *17*, 228–235. [[CrossRef](#)] [[PubMed](#)]
25. Woike, J.K.; Hoffrage, U.; Martignon, L. Integrating and testing natural frequencies, naïve Bayes, and fast-and-frugal trees. *Decision* **2017**, *4*, 234–260. [[CrossRef](#)]
26. Pitt, M.A.; Myung, I.J.; Zhang, S. Toward a method of selecting among computational models of cognition. *Psychol. Rev.* **2002**, *109*, 472–491. [[CrossRef](#)] [[PubMed](#)]
27. Gigerenzer, G.; Todd, P.M. *The ABC Research Group. Simple Heuristics that Makes Us Smart*; Oxford University Press: Oxford, UK, 1999.
28. Gigerenzer, G.; Goldstein, D.G. Reasoning the fast and frugal way: Models of bounded rationality. *Psychol. Rev.* **1996**, *103*, 650–669. [[CrossRef](#)] [[PubMed](#)]
29. Martignon, L.; Katsikopoulos, K.V.; Woike, J.K. Categorization with limited resources: A family of simple heuristics. *J. Math. Psychol.* **2008**, *52*, 352–361. [[CrossRef](#)]
30. Heinrich, M.; Müller, A.; Lammers-Lietz, F.; Borchers, F.; Mörgeli, R.; Kruppa, J.; Zacharias, N.; Winterer, G.; Slooter, A.J.C.; Spies, C.D. Radiological, Chemical, and Pharmacological Cholinergic System Parameters and Neurocognitive Disorders in Older Amerurgical Adults. *J. Gerontol. A Biol. Sci. Med. Sci.* **2021**, *76*, 1029–1036. [[CrossRef](#)]
31. American Psychiatric Association Publishing. *American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders*, 5th ed.; American Psychiatric Association Publishing: Washington, DC, USA, 2013. [[CrossRef](#)]
32. Charlson, M.E.; Pompei, P.; Ales, K.L.; MacKenzie, C.R. A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. *J. Chronic Dis.* **1987**, *40*, 373–383. [[CrossRef](#)]
33. UNESCO Institute for Statistics. *International Standard Classification of Education ISCED 2011*; UNESCO Institute for Statistics: Montreal, QC, Canada, 2012. Available online: <http://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-education-isced-2011-en.pdf> (accessed on 2 May 2020).
34. Mahoney, F.I.; Barthel, D.W. Functional evaluation: The barthel index. *Md. State Med. J.* **1965**, *14*, 61–65.
35. Lawton, M.P.; Brody, E.M. Assessment of older people: Self-maintaining and instrumental activities of daily living. *Gerontologist* **1969**, *9*, 179–186. [[CrossRef](#)]
36. Guigoz, Y.; Vellas, B.; Garry, P.J. Assessing the nutritional status of the elderly: The Mini Nutritional Assessment as part of the geriatric evaluation. *Nutr. Rev.* **1996**, *54*, S59–S65. [[CrossRef](#)] [[PubMed](#)]
37. Yesavage, J.A.; Brink, T.L.; Rose, T.L.; Lum, O.; Huang, V.; Adey, M.; Leirer, V.O. Development and validation of a geriatric depression screening scale: A preliminary report. *J. Psychiatr. Res.* **1982**, *17*, 37–49. [[CrossRef](#)]
38. Saunders, J.B.; Aasland, O.G.; Babor, T.F.; de la Fuente, J.R.; Grant, M. Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption—II. *Addiction* **1993**, *88*, 791–804. [[CrossRef](#)]
39. Breiman, L.; Friedman, J.H.; Olshen, R.A.; Stone, C.J. *Classification and Regression Trees*, 1st ed.; Routledge: London, UK, 1984. [[CrossRef](#)]
40. Therneau, T.; Atkinson, B. Recursive Partitioning and Regression Trees. R Package Version 4.1-16. 2019. Available online: <https://CRAN.R-project.org/package=rpart> (accessed on 2 May 2020).

41. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2020. Available online: <https://www.R-project.org/> (accessed on 2 May 2020).
42. Luan, S.; Schooler, L.J.; Gigerenzer, G. A signal-detection analysis of fast-and-frugal trees. *Psychol. Rev.* **2011**, *118*, 316–338. [[CrossRef](#)]
43. Radtke, F.M.; Franck, M.; MacGuill, M.; Seeling, M.; Lütz, A.; Westhoff, S.; Neumann, U.; Wernecke, K.D.; Spies, C.D. Duration of fluid fasting and choice of analgesic are modifiable factors for early postoperative delirium. *Eur. J. Anaesthesiol.* **2010**, *27*, 411–416. [[CrossRef](#)]
44. Robinson, T.N.; Raeburn, C.D.; Tran, Z.V.; Angles, E.M.; Brenner, L.A.; Moss, M. Postoperative delirium in the elderly: Risk factors and outcomes. *Ann. Surg.* **2009**, *249*, 173–178. [[CrossRef](#)]
45. Marcantonio, E.R.; Goldman, L.; Mangione, C.M.; Ludwig, L.E.; Muraca, B.; Haslauer, C.M.; Donaldson, M.C.; Whittemore, A.D.; Sugarbaker, D.J.; Poss, R.; et al. A clinical prediction rule for delirium after elective noncardiac surgery. *Jama* **1994**, *271*, 134–139. [[CrossRef](#)] [[PubMed](#)]
46. van den Boogaard, M.; Pickkers, P.; Slooter, A.J.; Kuiper, M.A.; Spronk, P.E.; van der Voort, P.H.; van der Hoeven, J.G.; Donders, R.; van Achterberg, T.; Schoonhoven, L. Development and validation of PRE-DELIRIC (PREdiction of DELIRium in ICu patients) delirium prediction model for intensive care patients: Observational multicentre study. *BMJ* **2012**, *344*, e420. [[CrossRef](#)]
47. Kim, E.M.; Li, G.; Kim, M. Development of a Risk Score to Predict Postoperative Delirium in Patients With Hip Fracture. *Anesth. Analg.* **2020**, *130*, 79–86. [[CrossRef](#)]
48. Rasmussen, L.S.; Larsen, K.; Houx, P.; Skovgaard, L.T.; Hanning, C.D.; Moller, J.T. The assessment of postoperative cognitive function. *Acta Anaesthesiol. Scand.* **2001**, *45*, 275–289. [[CrossRef](#)]
49. Fried, L.P.; Tangen, C.M.; Walston, J.; Newman, A.B.; Hirsch, C.; Gottdiener, J.; Seeman, T.; Tracy, R.; Kop, W.J.; Burke, G.; et al. Frailty in older adults: Evidence for a phenotype. *J. Gerontol. Ser. A Biol. Sci. Med. Sci.* **2001**, *56*, M146–M156. [[CrossRef](#)] [[PubMed](#)]
50. Vellas, B.; Guigoz, Y.; Garry, P.J.; Nourhashemi, F.; Bennahum, D.; Lauque, S.; Albarede, J.L. The Mini Nutritional Assessment (MNA) and its use in grading the nutritional state of elderly patients. *Nutrition* **1999**, *15*, 116–122. [[CrossRef](#)]
51. Ensrud, K.E.; Ewing, S.K.; Cawthon, P.M.; Fink, H.A.; Taylor, B.C.; Cauley, J.A.; Dam, T.-T.; Marshall, L.M.; Orwoll, E.S.; Cummings, S.R.; et al. A comparison of frailty indexes for the prediction of falls, disability, fractures, and mortality in older men. *J. Am. Geriatr. Soc.* **2009**, *57*, 492–498. [[CrossRef](#)]
52. Rockwood, K.; Andrew, M.; Mitnitski, A. A comparison of two approaches to measuring frailty in elderly people. *J. Gerontol. Ser. A Biol. Sci. Med. Sci.* **2007**, *62*, 738–743. [[CrossRef](#)]
53. Siscovick, D.S.; Fried, L.; Mittelmark, M.; Rutan, G.; Bild, D.; O’Leary, D.H. Exercise intensity and subclinical cardiovascular disease in the elderly. The Cardiovascular Health Study. *Am. J. Epidemiol.* **1997**, *145*, 977–986. [[CrossRef](#)] [[PubMed](#)]
54. Podsiadlo, D.; Richardson, S. The timed “Up & Go”: A test of basic functional mobility for frail elderly persons. *J. Am. Geriatr. Soc.* **1991**, *39*, 142–148. [[CrossRef](#)] [[PubMed](#)]