medRxiv preprint doi: https://doi.org/10.1101/2024.12.30.24319760; this version posted December 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Development and internal validation of a gradient-boosted trees model for prediction of delirium after surgery and anesthesia (the BioCog study)

Florian Lammers-Lietz, MD[1,2], Levent Akyuez, PhD[3,4], Diana Boraschi, PhD[5], Friedrich Borchers, MD[1], Jeroen de Bresser, PhD[6], Sreyoshi Chatterjee, PhD[7,8], Marta M. Correia[9], Nikola M. de Lange[7], Thomas Bernd Dschietzig, PhD[10], Soumyabrata Ghosh, PhD[7], Insa Feinkohl, PhD[11,12], Izabela Ferreira da Silva[7], Marinus Fislage[1], Anna Fournier, PhD[7,13], Jürgen Gallinat, PhD[14], Daniel Hadzidiakos, MD[1], Sven Hädel, MSc[2], Fatima Halzl-Yürek, MD[1], Stefanie Heilmann-Heimbach, PhD[15], Maria Heinrich, MD[1,16], Jeroen Hendrikse, PhD[17], Per Hoffmann, PhD[15,18,19], Jürgen Janke, PhD[11,20], Ilse M. J. Kant, PhD[22], Angelie Kraft, MSc[23,24], Roland Krause, PhD[7], Jochen Kruppa-Scheetz, PhD[25,26], Simone Kühn, PhD[1,14], Gunnar Lachmann, MD[1,16], Markus Laubach, MD, PhD, MBA[1,24,27], Christoph Lippert, PhD[28], David K. Menon[29,30], Rudolf Mörgeli, MD[1], Anika Müller, MD, PhD[1], Henk-Jan Mutsaerts, PhD[32,33], Markus Nöthen, MD[15], Peter Nürnberg, PhD[34-36], Kwaku Ofosu, MD[1], Malte Pietzsch, PhD[37], Sophie K. Piper, PhD[25,39], Tobias Pischon, MPH, PhD[11,20,21], Jacobus Preller, FRCP[29,31], Konstanze Scheurer, PhD[1], Reinhard Schneider, PhD[7], Kathrin Scholtz, PhD[1], Peter H. Schreier, PhD[24,35], Arjen J. C. Slooter, PhD[22], Emmanuel A. Stamatakis, PhD[30,39], Clarissa von Haefen, PhD[1], Simone J. T. van Montfort, PhD[22], Edwin van Dellen, PhD[22,40], Hans-Dieter Volk, MD, PhD[3,4], Simon Weber, PhD[37], Janine Wiebach[25,41], Anton Wiehe, MSc[23,24], Jeanne M. Winterer, MSc[2,42,43], Alissa Wolf, MD[1], Norman Zacharias, PhD[1,24,44], Claudia Spies, MD, PhD*[1], Georg Winterer, MD, PhD*[1,2,24], **on behalf of the BioCog consortium**

* Dr. Claudia Spies and Dr. Georg Winterer are equally contributing senior authors

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2024.12.30.24319760; this version posted December 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

1. Charité–Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Department of Anesthesiology and Intensive Care Medicine | CCM | CVK, Augustenburger Platz 1, 13353 Berlin, Germany
2. Pharmaimage Biomarker Solutions GmbH, Berlin, Germany
3. Berlin Institute of Health at Charité –Universitätsmedizin Berlin, BIH Center for Regenerative therapies (BCRT), and Charité–Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Institute of Medical Immunology, CVK, Augustenburger Platz 1, 13353 Berlin, Germany
4. CheckImmune GmbH, Berlin, Germany
5. Institute of Protein Biochemistry, Consiglio Nazionale delle Ricerche (CNR) di Pisa, Pisa, Italy
6. Department of Radiology, Leiden University Medical Center, Leiden, Netherlands
7. Bioinformatics Core, Luxembourg Center for Systems Biomedicine (LCSB), University of Luxembourg, Belvaux, Luxembourg
8. Deutsches Zentrum für Luft und RaumFahrt (DLR), Cologne, Germay
9. MRC Cognition and Brain Sciences Unit, University of Cambridge, Cambridge, United Kingdom
10. Immundiagnostik AG, Bensheim, Germany
11. Molecular Epidemiology Research Group, Max-Delbrück-Center for Molecular Medicine in the Helmholtz Association (MDC), Berlin, Germany
12. Faculty of Health at Department of Medicine, Witten/Herdecke University
13. Swiss Data Science Center (SDSC), ETH Zurich, Zurich, Switzerland
14. Department of Psychiatry, University Medical-Center Hamburg-Eppendorf, Hamburg, Germany
15. Institute of Human Genetics, University of Bonn, Bonn, Germany
16. Berlin Institute of Health at Charité–Universitätsmedizin Berlin, BIH Academy, (Digital) Clinician Scientist Program, Berlin, Germany

medRxiv preprint doi: https://doi.org/10.1101/2024.12.30.24319760; this version posted December 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

17. Department of Radiology and Brain Center Rudolf Magnus, University Medical Center Utrecht (UMC), Utrecht, Netherlands

18. Division of Medical Genetics, University Hospital, Basel, Switzerland

19. Human Genetics Research Group, Department of Biomedicine, University of Basel, Basel, Switzerland

20. Biobank Technology Platform, Max-Delbrück-Center for Molecular Medicine in the Helmholtz Association (MDC), Berlin, Germany

21. Berlin Institute of Health at Charité–Universitätsmedizin Berlin, Core Facility Biobank, Berlin, Germany

22. Department of Intensive Care Medicine and Brain Center, University Medical Center Utrecht (UMC), Utrecht University, Utrecht, the Netherlands

23. AdaLab UG, Hamburg, Germany

24. Pharmaimage Biomarker Solutions Inc., Cambridge, USA

25. Charité–Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Institute of Biometry and Clinical Epidemiology, Charitéplatz 10117 Berlin

26. Hochschule Osnabrück, University of Applied Sciences, Osnabrück, Germany

27. Department of Orthopaedics and Trauma Surgery, Musculoskeletal University Center Munich (MUM), LMU University Hospital, LMU Munich, Munich, Germany

28. Hasso-Plattner Institute, University of Potsdam, Potsdam, Germany

29. Neurosciences/Trauma Critical Care Unit, Cambridge University Hospitals NHS Foundation Trust, Cambridge, United Kingdom

medRxiv preprint doi: https://doi.org/10.1101/2024.12.30.24319760; this version posted December 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

30. Division of Anaesthesia, Department of Medicine, University of Cambridge, Cambridge, United Kingdom

31. Department of Medicine, University of Cambridge

32. Department of Radiology and Nuclear Medicine, Amsterdam University Medical Center, Amsterdam, The Netherlands

33. Amsterdam Neuroscience, Amsterdam, The Netherlands

34. Cologne Center for Genomics, University of Cologne, Faculty of Medicine and University Hospital Cologne, Cologne, Germany

35. Institute for Genetics of the University of Cologne, Cologne, Germany

36. Atlas Biolabs GmbH, Berlin, Germany

37. Cellogic GmbH (Cellogic), Berlin, Germany

38. Charité–Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Institute of Medical Informatics, Charitéplatz 10117 Berlin

39. Department of Clinical Neurosciences, School of Clinical Medicine, University of Cambridge, Cambridge, United Kingdom

40. Department of Psychiatry and UMC Utrecht Brain Center, University Medical Center Utrecht (UMC), Utrecht, Netherlands

41. BIH Berlin Institute of Health at Charité–Universitätsmedizin Berlin, 10117 Berlin, Germany, Core Unit Metabolomics

42. Charité–Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Department of Psychiatry |CCM, Charitéplatz 1, 10117 Berlin, Berlin, Germany

medRxiv preprint doi: https://doi.org/10.1101/2024.12.30.24319760; this version posted December 30, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

43. Department of Education and Psychology, Freie Universität Berlin, Berlin, Germany

44. Department of Otorhinolaryngology, Head and Neck Surgery, Charité-Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt Universität zu Berlin, Berlin, Germany

Corresponding author: Florian Lammers-Lietz, MD

Charité–Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin,
Department of Anesthesiology and Intensive Care Medicine | CCM | CVK
Augustenburger Platz 1, 13353 Berlin
Phone: +49 30 450 651 458 (Dr. Lammers-Lietz)
E-mail: florian.lammers@charite.de

medRxiv preprint doi: https://doi.org/10.1101/2024.12.30.24319760; this version posted December 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

IMPORTANCE: Postoperative delirium (POD) is a multietiological condition and affects 20% of older surgical patients. It is associated with poor clinical outcome and increased mortality.

OBJECTIVE: We aimed to develop and validate a risk prediction algorithm for POD based on a multimodal biomarker database exploiting preoperative data (predisposing factors) and procedural factors as well as perioperative molecular changes associated with POD (precipitating factors).

DESIGN: BioCog is a prospective cohort study conducted from November 2014 to April 2017. Patients were followed up for seven postoperative days after surgery for POD. Gradient-boosted trees (GBT) with nested cross-validation was used for POD prediction.

SETTING: Patients aged ≥65 years were enrolled at the anesthesiologic departments of two tertiary care centers.

EXPOSURE: All patients underwent surgery with an expected duration of at least 60min. Clinical, neuropsychological, neuroimaging data and blood were collected and clinically well established as well as non-established biomarkers (e.g., gene expression profiling) were measured pre- and postoperatively.

MAIN OUTCOME: POD according to DSM 5 until the seventh postoperative day

RESULTS: 184 of 929 (20%) patients experienced POD. A GBT algorithm using both preoperative data, characteristics of the intervention and postoperative changes in laboratory parameters achieved the highest area under the curve (0.83, [0.79; 0.86]) with a Brier score of 0.12 (0.12; 0.13).

CONCLUSIONS AND RELEVANCE: Models combining predisposing factors with precipitating factors predict POD best. Non-routine laboratory data provide useful information for POD risk prediction, providing relevant results for future studies on the molecular factors of POD. In addition, possibly relevant molecular mechanisms contributing to the development of POD were identified, mostly indicating a dysregulated postoperative immune response. This

medRxiv preprint doi: https://doi.org/10.1101/2024.12.30.24319760; this version posted December 30, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

study constitutes the basis for future hypothesis-driven analyses or implementation of prediction expert system for clinical practice.

KEYWORDS: *postoperative delirium, postoperative complications, anesthesia, cohort study, neuroimaging, risk factors, transcriptome, blood specimen collection*

medRxiv preprint doi: https://doi.org/10.1101/2024.12.30.24319760; this version posted December 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

1 Background

Delirium is an acute disturbance in attention, awareness, cognition, psychomotor behavior and emotional state because of another medical condition. The incidence of postoperative delirium (POD) ranges from 5-50% (1), but is most frequent in older patients (2, 3). POD incidence is assumed to rise in aging populations (4), challenging healthcare systems since it is associated with poor cognitive outcome, hospitalization, treatment costs, re-institutionalization, and mortality (3, 5).

Prehabilitation effectively mitigates postoperative neurocognitive disorders but is time consuming and prediction algorithms are necessary to carefully weigh POD risk against a delay of surgery (6).

Various previous studies have tried to build machine learning-based prediction tools for POD, usually based on retrospective analyses (7-11). The only two prospective studies achieved AUC values of 71% (12) and 74% (13). The prospective Biomarker Development for Postoperative Cognitive Impairment in the Elderly (BioCog) study was conducted with the main goal to improve POD-prediction. We were taking a systems medicine approach with focus on inflammatory alterations and the immune system, the cholinergic system and metabolic changes as well as indicators for early dementia based on an in-depth systematic review (1). Investigations included a wide range of perioperative clinical and neuropsychological parameters, neuroimaging, laboratory investigations and gene expression. Furthermore, the incorporation of precipitating factors may have additional value to predisposing factors.

The primary aim of this study was to develop and internally validate a POD risk index based on multimodal non-routine data intended for use by healthcare professionals to advise patients during medical decision making and allocating healthcare resources.

medRxiv preprint doi: https://doi.org/10.1101/2024.12.30.24319760; this version posted December 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

2 Methods

2.1 Study design

BioCog (www.biocog.eu, clinicaltrials.gov: NCT02265263, study protocol: (5)) is a prospective observational cohort study with the aim of identifying POD risk factors. The model was developed and internally validated in this cohort. All procedures were approved by the local ethics committees in Berlin, Germany (EA2/092/14) and Utrecht, Netherlands (14-469) and conducted in line with the declaration of Helsinki. All participants gave written informed consent prior to inclusion.

2.2 Participants

Male and female patients were enrolled in two tertiary care centers at the Charité–Universitätsmedizin Berlin, Germany, and the University Medical Center Utrecht, Netherlands. Consenting patients aged ≥65 years presenting for elective surgery with an expected duration >60min were included. Patients meeting one of the following criteria were excluded:

- positive screening for pre-existing major neurocognitive disorder defined as a Mini-Mental Status Examination (MMSE) score ≤23 points
- any condition interfering with neurocognitive assessment (severe sensory impairment, neuropsychiatric illness including alcohol and drug dependence, intracranial surgery)
- unavailability for follow-up assessment
- accommodation in an institution due to official or judicial order
- inability to give informed consent

2.3 Study procedures

The preoperative data were collected at least one day before surgery including medical history and clinical assessments, neuropsychological testing, blood collection and neuroimaging. Postoperative study visits took place twice daily until the seventh postoperative day.

medRxiv preprint doi: https://doi.org/10.1101/2024.12.30.24319760; this version posted December 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

2.4 Outcome

POD during the first seven days after surgery was the primary endpoint. Independently of the routine hospital procedures, POD screening was started in the recovery room and repeated twice per day at 8:00am and 7:00pm (±1h) up to seven days after surgery, by or under supervision of a study physician. POD was defined according to DSM-5 criteria and assessed by prospective screening with three validated tools which were recorded at each visit in accordance with current guidelines (2, 3), to mitigate the known tendency of physicians to underdiagnose POD. Patients were considered delirious if at least one of the following criteria was positive:

- ≥2 points on the Nursing Delirium Screening Scale (Nu-DESC),
- positive Confusion Assessment Method (CAM) score on a general ward,
- positive CAM for the Intensive Care Unit (CAM-ICU) score on an intensive care unit (ICU),
- chart review showing descriptions of delirium.

2.5 Clinical assessments

Before surgery, the study team recorded sociodemographic data and information on medication according to Carnahan's anticholinergic drug scale, health-related quality of life (EQ5D), Mini-Nutritional Assessment (MNA) and Body Mass Index (BMI), tobacco and hazardous alcohol consumption (AUDIT). A functional and physical assessment battery including frailty and walking speed was conducted. Precipitating factors were recorded: duration of surgery and anesthesia, type of anesthetic procedure (regional and/or general anesthesia), type of surgery (intracranial, intrathoracic/-abdominal/-pelvic surgery or peripheral), postoperative pain, prescription of anticholinergic medication daily until the seventh postoperative day, length of hospital and ICU stay as well as complications and postoperative mortality until the 90th postoperative day (eChapters 1.1-1.6).

medRxiv preprint doi: https://doi.org/10.1101/2024.12.30.24319760; this version posted December 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

2.6 Neuropsychological data

The preoperative cognitive assessment consisted of a comprehensive screen-based neuropsychological test battery (CANTAB, Cambridge Cognition Ltd., Cambridge, UK) and additional tests (Trail-Making-Test Parts A and B). MMSE score at the screening visit, CANTAB test scores and overall preoperative cognitive impairment (PreCI) were analyzed as risk factors.

PreCI is a dichotomous variable defined through comparison of cognitive test performance with a control group. We used multiple cognitive test parameters with moderate-to-good retest-reliability in the control group (14) and calculated z-scores of the baseline measurement in each test parameter assessed in the control group. The same z-transformation was then applied to the surgical cohort. Z-scores <-1·96 in at least two cognitive test parameters or an averaged z-score <-1·96 was used to define PreCI (eTable 2, eChapter 1.7).

2.7 Laboratory parameters

Preoperative serum and plasma samples were collected in supine position immediately before induction of anesthesia after eight hours of fasting and on the morning of the first postoperative day. Blood sampling was performed by trained clinic staff according to a standard operating procedure adapted from the German National Cohort Study (15). Samples were immediately sent to laboratories adjacent to the respective hospital site for analysis, or frozen at -80 °C and shipped to a central biobank at the Molecular Epidemiology Group, Max-Delbrück Center (MDC), Berlin for sample processing and storage. This group distributed samples for additional analyses to Atlas Biolabs GmbH as well as to several partners (Immundiagnostik AG in Bernsheim, Germany, Institute of Protein Biochemistry at Consiglio Nazionale delle Ricerche di Pisa, Immune Study Lab of Institute of Medical Immunology and BIH Center for Regenerative therapies at Charité-Universitätsmedizin Berlin). See eChapter 1.8 for a list of measured molecules. Whenever necessary, values were adjusted for laboratory.

medRxiv preprint doi: https://doi.org/10.1101/2024.12.30.24319760; this version posted December 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

2.8 Transcriptomics

Samples for transcriptomic analysis were collected in PAXgene tubes (Qiagen) at the same timepoints as other blood samples. Analyses were performed with Affymetrix Clariom S human microarray for RNA and Affymetrix® Flash Tag™ Biotin HSR (miRNA 4.1 Array Plates) for microRNA analyses (Thermo Fischer, Santa Clara, CA, USA) in a GeneTitan™ Multi-Channel Instrument by Atlas Biolabs GmbH (Berlin).

2.9 Neuroimaging

The magnetic resonance imaging (MRI) protocol included whole brain T1-weighted and T2-weighted high-resolution hippocampus imaging and diffusion tensor imaging (DTI). In addition, functional MRI and arterial spin labeling, but have not been considered for prediction due to low between-scanner agreement (inter-class correlation coefficient of 0·36-0·54 for functional connectivity in default mode, salience executive and dorsal attention networks and 0·17-0·39 for quantified cerebral blood flow). We calculated global and regional brain volumes including hippocampal subregions, cortical thickness and curvature from T1-weighted imaging, mean diffusivity, kurtosis and fractional anisotropy from DTI (eChapter 1.10).

2.10 Statistics

2.10.1 Estimation of sample size

The rule of thumb of Harrell was used to plan an appropriate number of POD events for a stable prediction model, i.e., ≥10 events per independent variable in logistic regression (16), which was considered adequate for machine learning. Requiring 260 patients with POD for analysis of up to 26 independent predictor variables and expecting a 25% incidence of POD, number of required patients was N=1040. Assuming a drop-out rate of 15%, a total number of N=1200 patients was planned. The initial analysis plan stipulated a training/test split approach for internal validation due to its computational efficiency. Since the study finally achieved a

medRxiv preprint doi: https://doi.org/10.1101/2024.12.30.24319760; this version posted December 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

lower cohort size, nested k-fold cross validation was used instead which works more efficiently on small samples.

2.10.2 Analysis of single parameters

For descriptive purposes, associations of pre-/perioperative parameters with POD were analyzed using simple logistic regression. We report odds ratios (OR) with 95% confidence intervals (CI) for the depending variable POD (reference category: no POD). To improve the interpretability of single parameter analyses, standardizing transformations were applied to the raw variables (eChapter 1.11.1) or dichotomized according to clinically relevant cut-off values for presentation of interpretable ORs. Analyses were conducted in R (R Core Team, R Foundation for Statistical Computing, Vienna, Austria) and SPSS (IBM, Armonk, NY). No adjustments for multiple testing were made and therefore, results should be considered exploratory, and we abstain from reporting p-values.

2.10.3 Machine learning

We applied machine learning (gradient boosted trees, GBT) to explore how the interplay of a larger set of predictors would benefit the prediction of POD risk in a bottom-up, data-driven fashion to allow unforeseen predictor-prediction relationships. Data available before surgery as well as data available on the first postoperative day by the latest were eligible for inclusion in machine learning, since these data were deemed useful for preoperative POD risk prediction as well as postoperative re-evaluation of further management.

Variables were assembled into blocks, i.e., preoperative data from the clinical assessment (“Clinical”), characteristics of the surgical intervention (“Precipitants” an “Pain”), preoperative neuroimaging data (“Imaging”), preoperative values) and perioperative difference in laboratory parameters measured in whole blood, plasma or serum (“Blood” and “Blood periop.”), preoperative RNA and µRNA abundance (“RNA” and “µRNA”), as well as perioperative difference in transcript abundance (“RNA periop.”). Different GBT models were built on combinations of various variable blocks. Combinations were selected sequentially, starting with

medRxiv preprint doi: https://doi.org/10.1101/2024.12.30.24319760; this version posted December 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

simple models (i.e., using only variables from one block) and then adding further blocks based on the AUC, assumptions on feasibility and relevance for clinical routine. Models using RNA data were evaluated separately since transcript abundance was only available for a subgroup of patients.

The GBT algorithm takes a set of decision trees as weak classifiers and combines them to form a strong classifier. It does so by incrementally adding decision trees during training to steadily improve its previous performance. The sampling of input cases is focused on those cases that were hard to classify before training and individual tree predictions are weighted. During inference, the output is computed through sequential application of each tree. GBT provides a continuous output parameter bounded between 0 and 1, allowing the choice of a clinically relevant cut-off which can be flexibly adapted to address various clinical questions and is inherently able to handle missing data. Area under the receiver operating curve and the Brier score with 95% CI are provided. The Brier score measures the difference between predicted probabilities and actual outcomes, ranging between 0, for perfect prediction, and 1. Models were validated using nested cross-validation. This approach allows model hyperparameter optimization and model selection while avoiding model overfitting. While each of the training datasets is provided to a hyperparameter optimized procedure, the evaluation of hyperparameters is performed using another cross-validation procedure that splits up the each of the provided train dataset into another set of k-folds (see eChapter 1.11.2). Sex-specific analyses have been conducted for the best-performing model. The funding source was not involved in study design, data collection, analysis, interpretation, writing or submitting the manuscript.

3 Findings

We recruited 933 patients between November 2014 and April 2017. Table 1 characterizes the sample. The patient flow chart is given in figure 1. Additional details on excluded patients are

medRxiv preprint doi: https://doi.org/10.1101/2024.12.30.24319760; this version posted December 30, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

given in eChapter 2.1. POD assessments were available for 929 patients. 184/929 (20%) patients developed POD.

83/184 (45%) cases of POD were identified in the bedside screening only, 13/184 (7%) cases of POD were diagnosed from chart review only, and 88/184 (48%) cases were proven in both chart review and bedside screening. eFigure 4 and eTable 3 give an overview of daily POD incidence.

medRxiv preprint doi: https://doi.org/10.1101/2024.12.30.24319760; this version posted December 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Table 1: Sample description (N=929)

	All		Women		Men	
	Median (IQR)	**Min.-max.**	**Median (IQR)**	**Min.-max.**	**Median (IQR)**	**Min.-max.**
Age (years)	72 (69-76)	65-91	72 (68-76)	65-87	72 (68-76)	65-91
BMI (kg/m²)	26.6 (24.0-29.4)	14.7-46.8	27.0 (23.6-30.5)	16.0-46.8	26.3 (24.2-28.7)	14.7-44.3
MMSE score (points)	29 (28-30)	24-30	29 (28-30)	24-30	29 (28-30)	24-30
GDS	1 (0-3)	0-13	2 (1-3)	0-13	1 (0-2)	0-10
EQ5D	0.88 (0.76-1.00)	-0.14-1.00	0.83 (0.69-0.92)	0.17-1.00	0.91 (0.79-1.0)	-0.14-1.00
Charlson's comorbidity index (p)	1 (0-2)	0-10	1 (0-2)	0-7	1 (0-2)	0-10
Hemoglobin (g/dL)	13.1 (11.9-14.3)	5.4-17.9	12.6 (11.6-13.6)	7.0-16.2	13.5 (12.3-14.7)	5.4-17.9
CRP (mg/L)	3.4 (1.4-8.3)	0.1-232.0	4.4 (1.7-10.0)	0.1-232.0	2.9 (1.1-7.2)	0.1-174.4
Leukocytes (nL^{-1})	6.2 (5.0-7.5)	1.6-24.6	6.2 (5.9-7.3)	2.7-19.4	6.2 (5.1-7.7)	1.6-24.6
Albumine (g/L)	40.7 (37.8-43.2)	15.5-51.7	40.1 (37.1-43.0)	22.2-51.7	41.0 (38.3-43.3)	15.5-51.6
Creatinine (µmol/L)	76.0 (64.5-90.2)	32.7-529.5	66.3 (58.3-79.1)	35.4-251.9	82.2 (71.6-95.0)	32.7-529.5
NT-proBNP (pmol/L)	6.1 (2.9-21.4)	2.9-617.2	6.9 (2.9-23.7)	2.9-617.2	5.4 (2.9-18.9)	2.9-397.8
LDL cholesterol (mmol/L)	3.0 (2.3-3.7)	0.1-7.7	3.1 (2.5-3.8)	0.1-7.7	2.9 (2.3-3.6)	0.5-6.4
Duration of anesthesia (Utrecht, min)[a]	265 (213-390)	10-1669	265 (225-400)	39-1663	260 (209-384)	10-1669
Duration of anesthesia (Berlin, min)[a]	167 (106-279)	25-753	169 (110-285)	30-676	162 (103-269)	25-753
Duration of surgery (Berlin, min)[b]	102 (55-191)	3-594	105 (58-191)	3-543	100 (55-140)	6-594
Duration of hospital	7	1-131	7	1-87	6	1-131

medRxiv preprint doi: https://doi.org/10.1101/2024.12.30.24319760; this version posted December 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

stay (days)		(4-11)		(4-12)		(3-9)	
Duration of ICU stay (days)		0 (0-0)	0-55	0 (0-0)	0-55	0 (0-0)	0-45
		Absolute n[b]	**Relative freq.[b]**	**Absolute n[b]**	**Relative freq.[b]**	**Absolute n[b]**	**Relative freq.[b]**
PreCI		122/924	13%	60/391	15%	62/533	12%
POD		184/929	20%	85/394	22%	99/535	19%
Mortality at 3 months (Berlin)[c]		29/683	4%	19/316	6%	10/367	3%
Compli-cations (Berlin)[c]	**death**	17/684	3%	10/316	3%	7/368	2%
	non-fatal	351/684	51%	167/316	53%	184/368	50%
Site of surgery	**Intracranial[d]**	10/911	1%	3/388	1%	7/523	1%
	Intratho-racic, -ab-dominal, -pelvic	397/91	44%	166/388	43%	231/523	44%
	peripheral	505/911	55%	219/388	56%	286/523	55%
Type of anesthe-sia	**general**	687/912	75%	283/385	74%	404/527	77%
	regional	57/912	6%	22/385	6%	35/527	7%
	combined	168/912	18%	80/385	21%	88/527	17%
ASA-PS	**I**	36/929	4%	14/394	4%	22/535	4%
	II	557/929	60%	241/394	62%	316/535	59%
	III	335/929	36%	138/394	35%	197/535	37%
	IV	1/929	<1%	1/394	<1%	0/535	0%
Women		394/929	42%	n.a.	n.a.	n.a.	n.a.
MNA	**Normal**	662/911	73%	266/382	70%	396/529	75%
	At risk	200/911	22%	91/382	24%	109/529	21%
	Malnour-ishement	49/911	5%	25/382	7%	24/529	5%
Frailty (Fried)	**robust**	354/631	56%	139/277	50%	215/354	61%
	prefrail	175/631	28%	80/277	29%	95/354	27%
	frail	102/631	16%	58/277	21%	44/354	12%
Smoker		90/903	10%	39	10%	51	10%
Hazardous alcohol con-sumption (AUDIT)		62/862	7%	21	6%	41	8%
ISCED level	**1+2**	150/839	18%	80/358	22%	70/481	15%
	3+4	343/839	41%	177/358	49%	166/481	35%
	5+6	346/839	41%	101/358	28%	245/481	51%

medRxiv preprint doi: https://doi.org/10.1101/2024.12.30.24319760; this version posted December 30, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

BMI: body mass index; IQR: interquartile range; MMSE: Mini-Mental Status Examination; GDS: geriatric depression scale; CRP: C-reactivel protein; LDL: low-density lipoprotein; NT-proBNP: N-terminal pro-brain natriuretic peptide; freq.: frequency; ICU: intensive care unit; PreCI: preoperative cognitive impairment; POD: postoperative delirium; ASA-PS: American Society of Anesthesiologists Physical Status; MNA: Mini-nutritional assessment; ISCED: International Standard Classification for Education

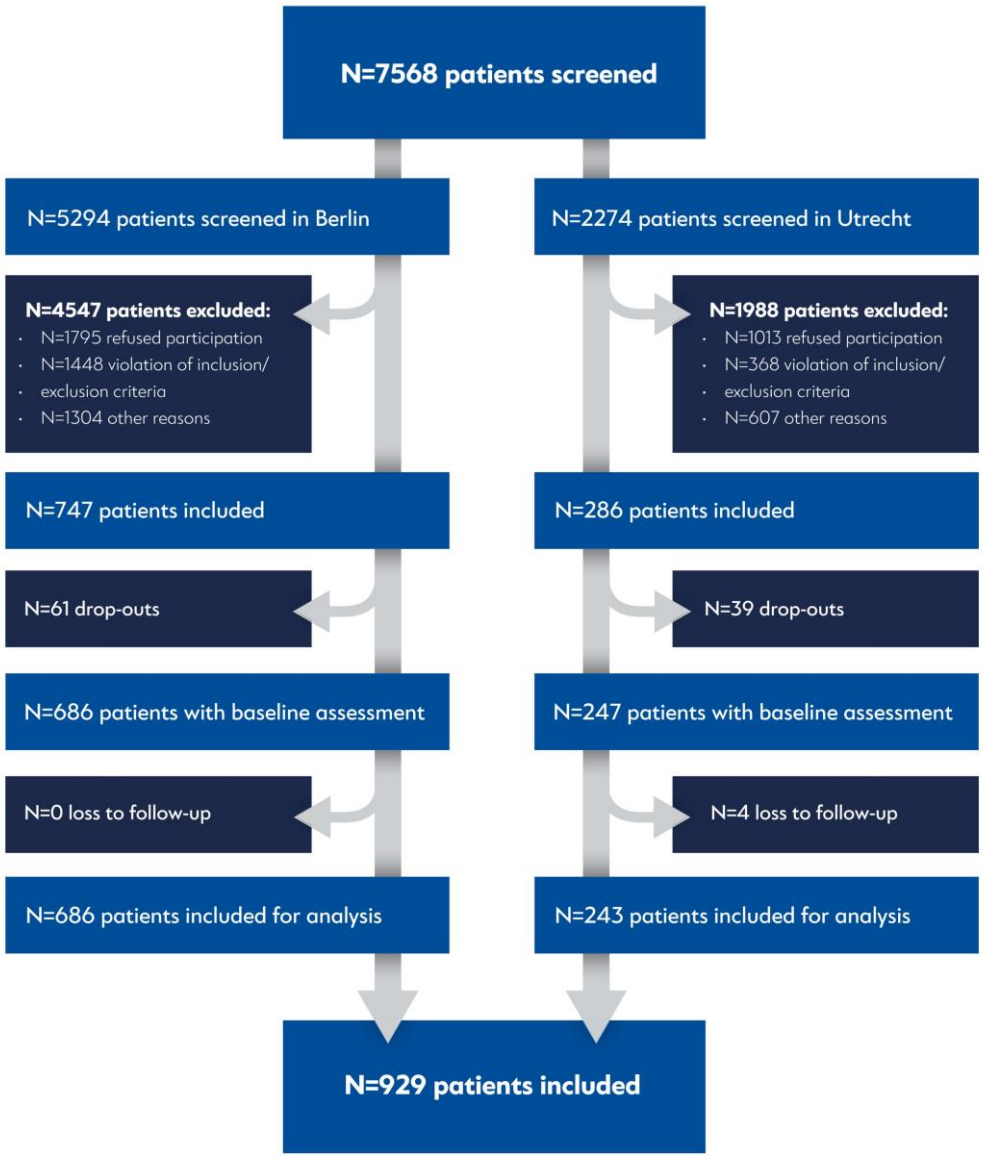
[a] end of anesthesia was assessed differentially in both study centers

[b] relative frequencies are calculated after correction for missing values

[c] data are only available for the study center in Berlin

[d] intracranial surgery not affecting brain parenchyma (e.g. meningioma)

medRxiv preprint doi: https://doi.org/10.1101/2024.12.30.24319760; this version posted December 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .



345 ***Figure 1:*** *Patient flow chart.*

medRxiv preprint doi: https://doi.org/10.1101/2024.12.30.24319760; this version posted December 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

3.1 POD risk factors

Figure 2 displays unadjusted OR with 95% CIs for preoperative parameters with CIs excluding unity. Sample and effect sizes for all parameters are given in the online-only material.

Age was directly associated with POD. Among age-related conditions, frailty had the strongest association with POD (OR 1·90 [1·49; 2·44] for category change in Fried's phenotype), as well as slow walking speed (1·80 [1·22; 2·63] for TUG>10s), malnutrition (OR 1·67 [1·29; 2·16] for category change in MNA-SF), any functional impairment according to Barthel index or IADL assessment (OR 1·59 [1·12; 2·24]) and depressive symptoms (OR 1·57 (1·05; 2·41) for GDS>0).

An MMSE score <27points had a higher OR (3·10 [1·96; 4·85]) for POD than PreCI (OR 2·57 [1·69; 3·88], see also eTable 8 and eFigure 5)

Preoperatively higher levels of cholesterol (standardized, adjusted OR 0·79 [0·65; 0·95]) and associated lipoproteins (HDL and LDL) were protective against POD. A postoperative decrease in triglycerides, cholesterol and LDL were associated with higher POD incidence.

Four inflammatory parameters were positively associated with POD: IL6 (standardized OR 1·19 [1·03; 1·38]), whole blood IL8 (standardized OR 1·42 [1·02; 1·98]) (17), CRP (standardized, adjusted OR 1·20 [1·03; 1·41]), immature granulocyte fraction (standardized OR=1·34 [1·10; 1·63]) and neutrophil count (standardized, adjusted OR 1·22 [1·03; 1·46]). An increase of inflammatory parameters on the first postoperative assessment was associated with higher likelihood of POD (CRP: standardized, adjusted OR 1·59 [1·14; 2·21], IL6: standardized OR 1·76 [1·48; 2·09], and IL8: standardized OR 1·96 [1·18; 3·24]). Cellular immune response showed a more complex association with POD: Whereas a postoperative increase in leukocytes (standardized, adjusted OR 1·36 [1·12; 1·64]) and neutrophiles (standardized, adjusted OR 1·47 [1·2; 1·81]) was associated with POD, an increase in lymphocytes lowered the odds for POD (standardized, adjusted OR 0·66 [0·54; 0·81]).

medRxiv preprint doi: https://doi.org/10.1101/2024.12.30.24319760; this version posted December 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Higher levels tryptophan (standardized, laboratory-adjusted OR 0·74 [0·62; 0·89]) and albumin (standardized, adjusted OR 0·68 [0·56; 0·81]) also lowered the odds for POD. A higher plasma β-amyloid 42/40-ratio was found to be related to a lower POD likelihood (standardized, adjusted OR 0·74 [0·56; 0·93]), but this association seemed to be driven by increased POD risk in patients with higher levels of β-amyloid 40 (standardized, adjusted OR 1·20 [1·02; 1·41]).

Both higher preoperative γ-glutamyltransferase levels (standardized, adjusted OR 1·24 [1·06; 1·47]) and a postoperative decrease (standardized, adjusted OR 0·81 [0·68; 0·98]) were associated with POD. A postoperative increase in transaminases was associated with POD. Postoperatively decreasing levels of oxidative stress indicated by nitrotyrosine levels (standardized OR 0·72 [0·53; 0·98]) and nitrous oxide production indicated by homoarginine levels (standardized OR 0·48 [0·31; 0·73]) were associated with increased POD risk.

Longer duration of anesthesia (OR 4·42 [3·15; 6·27] for >4h) and surgery (OR 7·44 [4·84; 11·50] for >4h) as well as blood loss (standardized, adjusted OR for perioperative changes in Hb: 0·76 [0·63; 0·91], thrombocytes: 0·57 [0·46; 0·69], and albumin: 0·66 [0·54; 0·81]) were associated with POD. Compared to general anesthesia, surgery performed in regional anesthesia was associated with lower rates of POD (0·29 [0·09; 0·72]). Surgery with opening of thorax, abdomen or pelvis was associated with increased rates of POD compared to peripheral surgery (OR 3·00 [2·13; 4·25]). Pain (OR 2·16 [1·55; 3·01]) or intake of any anticholinergic medication (OR 2·35 [1·50; 3·84]) at least once during follow-up until the seventh postoperative day were both associated with POD (see also eTables 11-12, eFigures 6 and 7).

Various associations of structural MRI-derived parameters were observed (complete results: eFigure 8, eTable 15), we would like to emphasize a protective association of POD with global brain volume (standardized OR 0·71 [0·55; 0·92]) as well as hippocampus volume (standardized OR [0·67 (0·53; 0·86]).

medRxiv preprint doi: https://doi.org/10.1101/2024.12.30.24319760; this version posted December 30, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

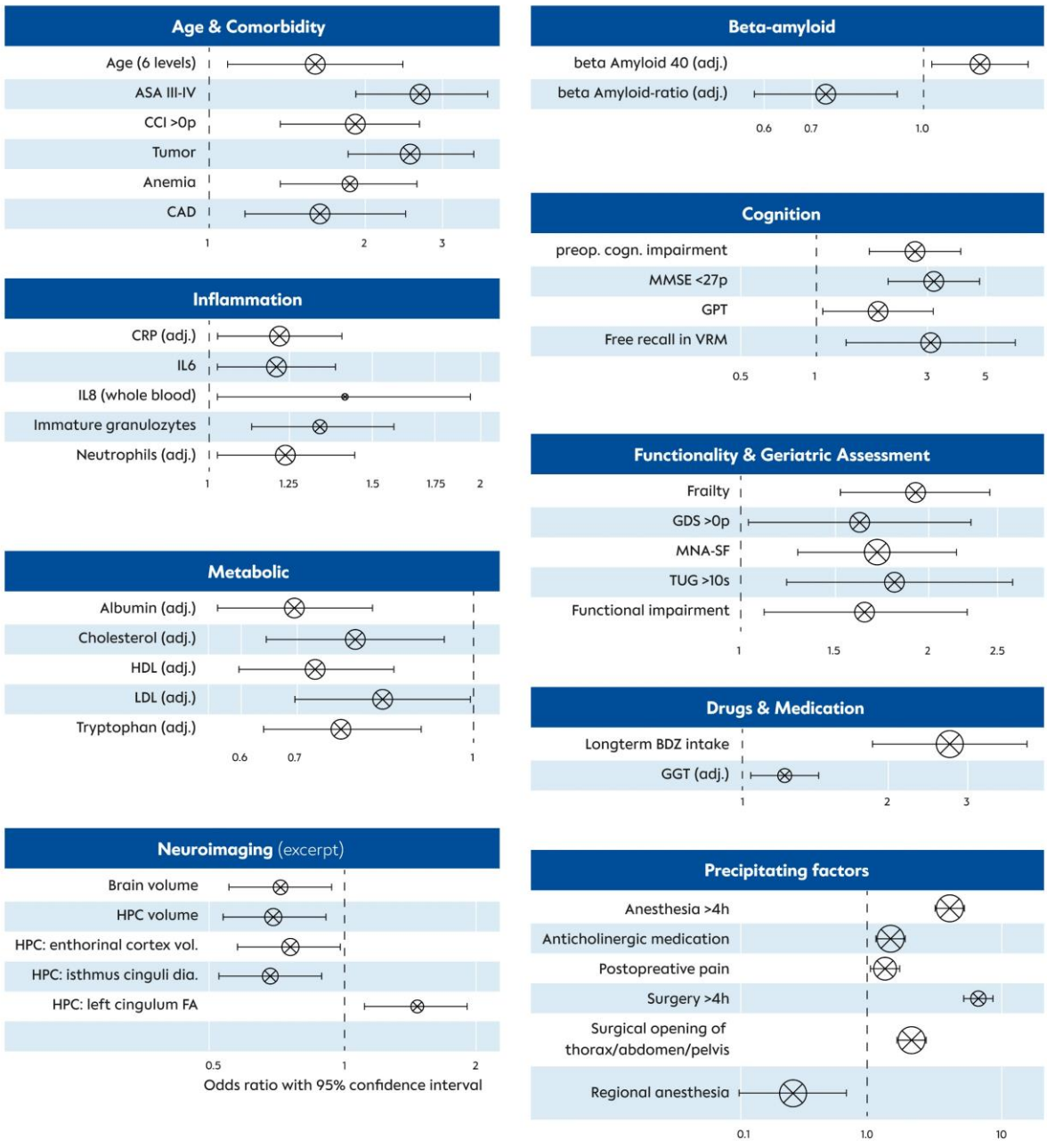


Figure 2: *Summary of parameters that were significantly associated with POD. Odds ratios (OR) with 95% confidence interval (95% CI) are shown (only parameters are depicted with CI excluding unity). The diameter of the circle corresponds to the number of available datasets. See also supplementary material 2.*

The term tumor includes diagnoses of solid malignoma, leukemia and lymphoma.

Abbreviations:

adj.: adjusted for assessment in different study centers, p: points

medRxiv preprint doi: https://doi.org/10.1101/2024.12.30.24319760; this version posted December 30, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

age & comorbidity: ASA: American Society of Anesthesiologists Physical Status, CAD: coronary artery disease CCI: Charlson comorbidity index

inflammation: CRP: C-reactive protein, IL: interleukin

cognition: GPT: Grooved Pegboard Test (completion time), MMSE: Mini-mental status examination, preop. cogn. impairment: preoperative cognitive impairment, VRM: Verbal Recognition Memory

functionality & geriatric assessment: GDS: Geriatric depression scale, MNA-SF: Mini-nutritional assessment short form, TUG: Timed up-and-go test, frailty refers to Fried's frailty phenotype

metabolic: BMI: body mass index, HDL: high density lipoprotein, LDL: low density lipoprotein

drugs & medication: GGT: γ-glutamyltransferase, BDZ: preoperative longterm prescription of benzodiazepines

neuroimaging: CA: cornu ammonis, dia.: diameter (cortical thickness), FA: fractional anisotropy, HPC: hippocampus, vol.: volume

medRxiv preprint doi: https://doi.org/10.1101/2024.12.30.24319760; this version posted December 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

3.2 Machine learning-prediction of POD

Figure 3 displays AUC for the GBT models built from the most relevant combinations of variable blocks and eTable 16 provides details on the model performance. Among the models using only preoperative data, the model using only clinical data performed best (AUC 0·76 [0·69; 0·81], Brier score: 0·14 [0·13; 0·16]). Adding preoperative blood or RNA data to the model did not improve the AUC.

The model AUC was increased considerably by adding characteristics of the intervention ("Precipitants") and perioperative changes in laboratory parameters ("Blood periop.") to the clinical data, and the highest overall AUC was achieved by a model using these three blocks of data (AUC 0·83 [0·79; 0·86], Brier score 0·12 [0·12; 0·13]). The most important variables in this model are displayed in figure 4, and eFigure 10 displays sex-stratified Receiver-Operating Curves. This model also provided the highest performance in the subgroup of patients with RNA data (AUC 0·78 [0·73; 0·83], Brier score 0·15 [0·14; 0·16]), and adding transcript data to the model did not improve AUC. However, a model exploiting only pre- and and postoperative RNA data ("RNA+RNA periop.") showed almost identical performance (AUC 0·77 [0·71; 0·78], Brier Score 0·15 [0·14; 0·16]).

eFigure 10 displays the relevant transcripts from the "RNA+RNA periop." model: The perioperative changes in mRNA abundance were more often predictive of POD than preoperative abundance. Most important transcripts were *BTN3A1, LAP3, DSN1, HPGD* and *KIF4B*. Notably, both preoperative *JAK2* and circular *JAK2* mRNA were predictive of POD. In an exploratory Cox regression analysis, we found that a considerable number of transcripts (*HPGD, BTN3A1, LAP3, JAK2* and circular *JAK2*) were also associated with postoperative mortality (eTable 17, eFigure 11).

medRxiv preprint doi: https://doi.org/10.1101/2024.12.30.24319760; this version posted December 30, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

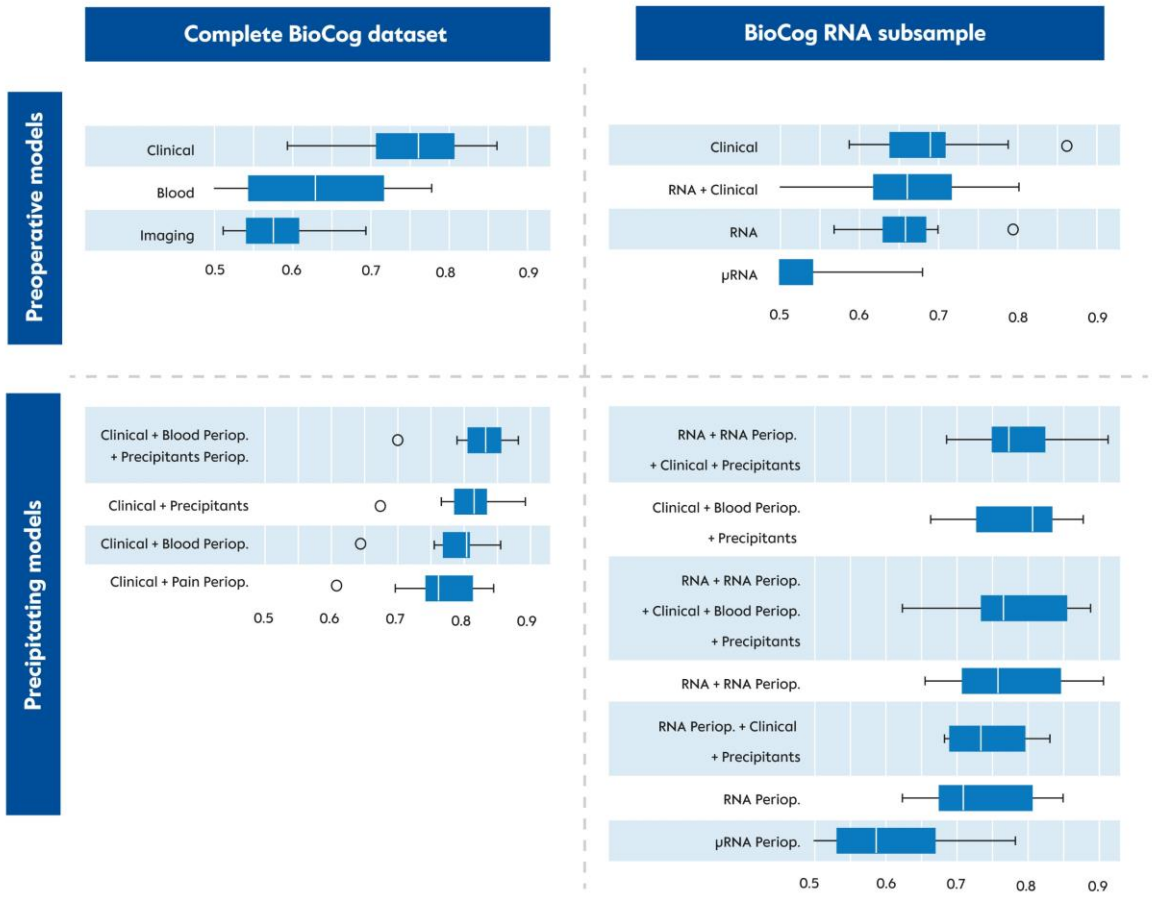


Figure 3: *Boxplot displaying area under the curve (AUC) of the receiver-operating characteristic (ROC). A value of 1 indicates 100% sensitivity at 100% specificity, whereas a value of 0.5 indicates indiscriminability of the model for POD. Each model evaluates a different combination of available datasets, as indicated on the y-axis.*

Abbreviations: periop.: perioperative (referring to precipitating factors, e.g., pain or medication, and perioperative changes in molecule abundance), RNA: transcriptomic data features.

medRxiv preprint doi: https://doi.org/10.1101/2024.12.30.24319760; this version posted December 30, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

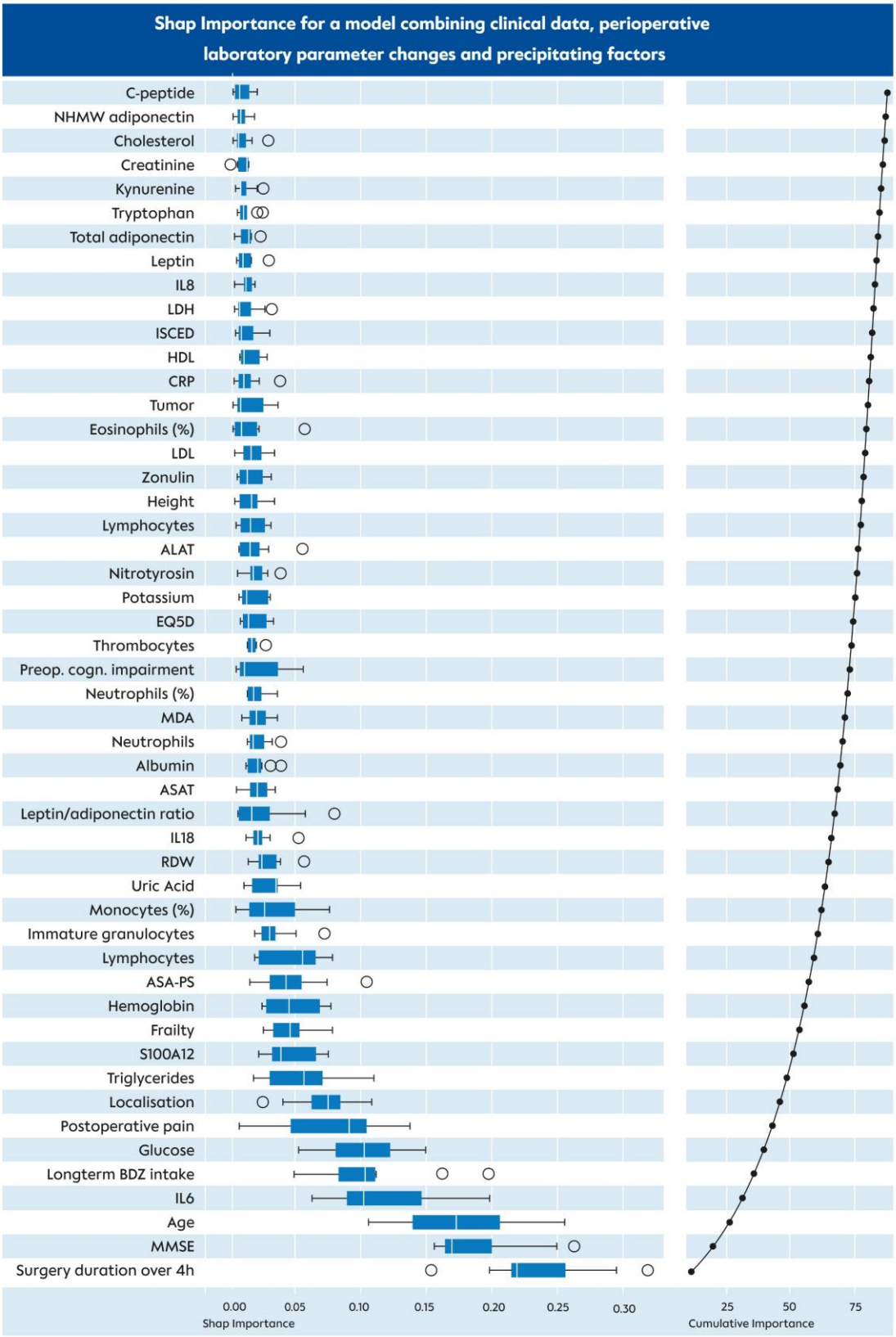


Figure 4: *Feature importance of the model with the highest predictive performance.*

medRxiv preprint doi: https://doi.org/10.1101/2024.12.30.24319760; this version posted December 30, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Tumor diagnosis includes solid malignancies, lymphoma and leukemia. Abbreviations: ALAT: alanin-aminotransferase, ASAT: aspartat-aminotransferase, BDZ: benzodiazepine, CRP: C-reactive protein, HDL: high-density lipoprotein, IL: interleukin, ISCED: International Standard Classification of Education, LDH: lactate dehydrogenase, LDL: low-density lipoprotein, MDA: malondialdehyde MMSE: Mini-mental status examination, NHWM: non-high molecular weight, postop.: postoperative, preop. cogn. impairment: preoperative cognitive impairment, RDW: red cell distribution width

medRxiv preprint doi: https://doi.org/10.1101/2024.12.30.24319760; this version posted December 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

4 Discussion

We estimated POD prediction algorithms based on clinical, neuropsychological, blood-based and neuroimaging data. This is the first approach to POD prediction using data-driven analysis of a prospectively collected multimodal dataset. By aggregating clinical preoperative data, precipitating factors with preoperative laboratory values and postoperative changes, the model achieved good discriminability (AUC 83%) with good model fit.

Previous approaches used retrospectively collected data or merged heterogenous data from multiple studies (9, 11, 13). The only prospective study (SAGES) achieved an AUC of 71% using machine learning in preoperatively available clinical data (12). Our model solely relying on preoperative clinical data achieved similar performance (AUC 76%), and no improvement by adding preoperative non-routine data was achieved. Hence, thorough preoperative clinical evaluation to identify patients at risk can be considered a suitable approach in clinical routine. However, using algorithms as a diagnostic expert device can support quantifying POD risk and drive the establishment of POD risk assessment in routine clinical practice. Results suggest that information about intervention and postoperative course can improve the model to an AUC of 80%. Although models using precipitating data are intended for risk monitoring rather than prediction, relevant information is usually available before surgery, i.e., estimated duration of intervention and expected postoperative pain, and may be used for prediction as well.

Our analyses suggest that models exploiting precipitating factors and perioperative laboratory assessments can considerably improve POD risk monitoring, but neuroimaging and transcriptomic data do not. However, gene expression data may be of particular interest for further studies, since in the subgroup of patients with RNA data, a model exploiting only mRNA achieved a similar AUC (77%) compared to the best performing model (78%). A perioperative risk monitoring algorithm based on two gene expression analyses could relieve medical staff from extensive clinical assessments, be more cost-effective than using multiple independent laboratory assays and avoids data aggregation from different sources.

medRxiv preprint doi: https://doi.org/10.1101/2024.12.30.24319760; this version posted December 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Many of the most predictive transcripts were mRNA of ubiquitously expressed genes involved in major molecular mechanisms such as cell proliferation (e.g., *DSN1*, *LAP3*, *KIF4B* and *JAK*). This may suggest that POD is a heterogeneous phenomenon originating via distinct molecular pathways. These central molecular nodes may be the common denominator among different POD subgroups, but nevertheless suitable for prediction. Certain transcripts suggest involvement of γδT-cells in neuroinflammation and neuroplasticity (*BTN3A1*)(18), metabolic dysregulation and autophagy (*LAP3*)(19), proliferation (*DSN1*, *KIF4B*)(20), interaction with the immune system (*JAK*)(21, 22), and senescence (*HPGD*)(23). These transcripts are also associated with 3-month mortality (eChapter 2.3). Above-mentioned molecules S100A12 (24), interleukins and zonulin (25) point to certain immune response pathways, which may be related to neurotransmitter dysbalance by tryptophan and kynurenine metabolism (26). Some of the identified molecular targets have already been discussed with respect to neurodegeneration, i.e., malondialdehyde, nitrotyrosine (27), metabolites of the kynurenine pathway (28), S100A12 (24) and zonulin (25).

The BioCog study is small in relation to the wide spectrum of parameters included in our database. To fully exploit the potential of machine learning, larger samples are necessary. The current sample excluded patients with MMSE score ≤23, but brain atrophy may be relevant biomarker in patients with preoperative cognitive impairment. Since external validation in an independent dataset is pending, we have used nested cross validation as an internal validation procedure, which is robust against overfitting. The focus of this manuscript is prediction, whereas a molecular causal model cannot be addressed here. E.g., single variable analyses have not been adjusted for confounders, warranting further analysis. The best model was chosen by ROC-AUC, which is a measure of discrimination in diagnostic testing. For prognostic questions, outcome probability estimation is preferable and more evaluations are needed (29).

POD screening was performed according to the evidence-based standard that measure POD at least twice a day and has a comprehensive geriatric assessment included to describe the

medRxiv preprint doi: https://doi.org/10.1101/2024.12.30.24319760; this version posted December 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

clinical entity of this population. The clinical phenomenology was structured and annotated according to this standard (2).

BioCog has made advancements towards POD prediction and will facilitate comprehensive hypothesis-driven analyses including subgrouping of patients for better understanding of pathophysiological processes and conception of interventional studies. Our dataset can guide prevention strategies to reduce POD, e.g., via the JAK-pathway (22).

5 Acknowledgements

We thank the Koordinierungszentrum für Klinische Studien (KKS Berlin), and especially Alexander Krannich, and ALTA (Siena, Italy) who provided additional administrative and coordinating services throughout the study. Multimodal data management was conducted by Pharmaimage Biomarker Solutions GmbH. We thank our colleagues who participated as MD students and team of study nurses, especially Tuba Gülmez, Emmanuel Keller, Mario Lamping, Helene Michler, Juliane Dörfler, Zdravka Bosancic, Fieras Nosierat, Irene Mergele, Victoria Windmann and Anna Nottbrock. The authors further wish to thank the team of the student apprentices/interns of the Department of Anesthesiology at the Charité-Universitätsmedizin Berlin. Magnet resonance imaging has been supported by the Berlin Center for Advanced Neuroimaging core staff, especially Stefan Hetzer. Henning Krampe supported the study by recruiting and supervising students for neuropsychological testing.

6 Funding

The research leading to these results has received funding from the European Union Seventh Framework Program [FP7/2007-2013] under grant agreement n° 602461.

7 Data availability

Due to the protection of intellectual property, machine learning algorithms will not be made publicly available, but can be obtained from Dr. Winterer (georg.winterer@pi-

medRxiv preprint doi: https://doi.org/10.1101/2024.12.30.24319760; this version posted December 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

pharmaimage.com) after signing a confidentiality agreement. Participant data may be made available upon request following publication to researchers who provide a methodologically sound proposal in accordance with applicable legal and regulatory restrictions. Proposals for data analysis must be directed to both claudia.spies@charite.de and georg.winterer@pi-pharmaimage.com. To gain access, requesting researchers will need to sign a data access agreement. Analyses will be limited to those approved in appropriate ethics and governance arrangements. All study documents which do not identify individuals (e.g. study protocol, informed consent form) will be freely available on request.

8 Statement on sex and gender equity in research

Data on the patients' gender have not been collected, and we refer to "sex" throughout the manuscript. Although the best-performing model was found to have similar discriminability in women and men, analyses in the online-only material have neither been adjusted nor disaggregated for sex, although further in-depth analyses would require this step to yield valid results.

9 Protein products of mentioned genes

BTN3A1	-	butyrophilin
LAP3	-	leucine aminopeptidase 3
DSN1	-	DSN1 component of MIS12 kinetochore complex
HPGD	-	15-hydroxyprostaglandin dehydrogenase
KIF4B	-	kinesin family member 4B

10 Author contributions

Florian Lammers-Lietz: Writing – Original Draft, Investigation, Validation, Visualization; Levent Akyuez; Methodology, Investigation, Resources; Diana Boraschi: Methodology, Formal

medRxiv preprint doi: https://doi.org/10.1101/2024.12.30.24319760; this version posted December 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Analysis, Writing – Review & Editing; Friedrich Borchers: Conceptualization, Investigation, Formal Analysis, Validation, Data Curation, Writing – Review and Editing; Jeroen de Bresser: Methodology, Investigation, Formal Analysis; Sreyoshi Chatterjee: Investigation, Formal Analysis; Marta M. Correia: Formal Analysis; Nikola M. de Lange: Methodology, Investigation, Formal Analysis; Thomas Bernd Dschietzig: Conceptualization, Methodology, Investigation, Formal Analysis, Supervision, Funding Acquisition; Soumyabrata Ghosh: Methodology, Investigation, Formal Analysis; Insa Feinkohl: Conceptualization, Methodology, Formal Analysis, Data Curation; Izabela Ferreira da Silva: Investigation, Formal Analysis; Marinus Fislage: Investigation, Writing – Original Draft, Formal Analysis; Anna Fournier: Methodology, Investigation, Formal Analysis; Jürgen Gallinat: Methodology, Investigation, Formal Analysis, Supervision; Daniel Hadzidiakos: Conceptualization, Methodology, Investigation, Data Curation, Validation; Sven Hädel: Software, Data Curation; Fatima Halzl-Yürek: Investigation; Stefanie Heilmann-Heimbach: Methodology, Investigation, Formal Analysis; Maria Heinrich: Investigation, Data Curation, Validation; Jeroen Hendrikse: Methodology, Investigation, Formal Analysis; Per Hoffmann: Methodology, Investigation, Formal Analysis; Jürgen Janke: Data Curation; Ilse M. J. Kant: Investigation, Methodology, Validation; Angelie Kraft: Formal Analysis; Roland Krause: Methodology, Investigation, Formal Analysis, Supervision; Jochen Kruppa-Scheetz: Methodology, Formal Analysis, Data Curation; Simone Kühn: Conceptualization, Methodology, Investigation, Formal Analysis, Supervision, Funding Acquisition; Gunnar Lachmann: Investigation, Methodology; Markus Laubach: Investigation; Christoph Lippert: Methodology, Investigation, Formal Analysis; David K. Menon: Conceptualization, Resources, Supervision, Writing – Review and Editing; Rudolf Mörgeli: Methodology, Investigation; Anika Müller: Methodology, Investigation; Henk-Jan Mutsaerts: Methodology, Investigation, Formal Analysis; Markus Nöthen: Methodology, Investigation, Formal Analysis; Peter Nürnberg: Conceptualization, Methodology, Supervision, Project Administration, Funding Acquisition; Kwaku Ofosu: Investigation; Malte Pietzsch: Conceptualization, Funding Acquisition, Methodology; Sophie K. Piper: Methodology, Formal Analysis; Tobias Pischon:

medRxiv preprint doi: https://doi.org/10.1101/2024.12.30.24319760; this version posted December 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Conceptualization, Methodology, Resources, Data Curation, Writing – Original Draft, Supervision, Project Administration, Funding Acquisition; Jacobus Preller: Supervision, Conceptualization, Methodology, Investigation, Formal Analysis, Writing – Review and Editing, Project Administration; Konstanze Scheurer: Project Administration; Reinhard Schneider: Methodology, Investigation, Formal Analysis; Kathrin Scholtz: Project Administration; Peter H. Schreier: Writing – Review and Editing; Arjen J. C. Slooter: Conceptualization, Funding Acquisition, Methodology, Writing – Review and Editing, Administration; Emmanuel A. Stamatakis: Supervision, Conceptualization, Funding Acquisition, Investigation, Methodology, Writing – Review and Editing, Project Administration; Clarissa von Haefen: Data Curation, Project Administration; Simone J. T. van Montfort: Investigation, Validation; Edwin van Dellen: Methodology, Investigation; Hans-Dieter Volk: Resources, Writing – Review and Editing; Simon Weber: Conceptualization, Funding Acquisition, Methodology; Janine Wiebach: Methodology, Formal Analysis, Software, Data Curation; Anton Wiehe: Methodology, Formal Analysis, Visualization, Software; Jeanne M. Winterer: Investigation; Alissa Wolf: Methodology, Formal Analysis, Project Administration; Norman Zacharias: Methodology, Formal Analysis, Investigation; Claudia Spies: Conceptualization, Methodology, Validation, Software, Resources, Data Curation, Writing – Original Draft, Supervision, Project Administration, Funding Acquisition; Georg Winterer: Conceptualization, Methodology, Software, Resources, Data Curation, Writing – Original Draft, Supervision, Project Administration, Funding Acquisition; Florian Lammers-Lietz and Georg Winterer have verified the underlying data.

11 Assistance in scientific writing

Neither artificial intelligence tools nor a professional medical writer has assisted in manuscript preparation.

medRxiv preprint doi: https://doi.org/10.1101/2024.12.30.24319760; this version posted December 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

12 Conflict of interest

Georg Winterer is currently licensing a Class IIa medical device (web-based software tool for risk prediction of POD and POCD in clinical practice). Dr. Winterer is CEO of PharmaImage Biomarker Solutions GmbH Berlin (Germany) and President of its subsidiary Pharmaimage Biomarkers Incl. (Cambridge, MA, USA).

Dr. Spies, Dr. Winterer, Dr. Boraschi, Dr. Dschietzig, Dr. Kühn, Dr. Nürnberg, Dr. Pischon, Dr. Pietzsch, Dr. Slooter, Dr. Stamatakis, Dr. Weber, report grants from the European Commission during the conduct of the study. Dr. Winterer reports grants from the Deutsche Forschungsgemeinschaft (DFG)/German Research Society and from the German Ministry of Health. Dr. Spies reports grants from DFG/German Research Society, Einstein Foundation Berlin, Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR)/German Aerospace Center, Projektträger im DLR/Projec Management Agency, Gemeinsamer Bundesausschuss (GBA)/Federal Joint Committee, inneruniversity grants, Stifterverband/Non-Profit Society Promoting Science and Education, European Society of Anesthesiology and Intensive Care, BMWI – Federal Ministry of Economic Affairs and Climate Action, Dr. F. Köhler Chemie GmbH, Sintetica GmbH, Max-Planck-Gesellschaft zur Förderung der Wissenschaft e.V., Metronic, BMBF – Federal Ministry of Education and Research, Robert Koch Institute and payments by Georg Thieme Verlag, board activity for Prothor, Takeda Pharmaceutical Company Ltd., Lynx Health Science GmbH, AWMF (Association of the Scientific Medical Societies in Germany), DFG, Deutsche Akademie der Naturforscher Leopoldina e.V. (German National Academy of Sciences Leopoldina), Berliner Medizinische Gesellschaft, European Society of Intensive Care Medicine (ESICM), European Society of Anaesthesiology and Intensive Care (ESAIC), Deutsche Gesellschaft für Anästhesiologie und Intensivmedizin (DGAI)/German Society of Anaesthesiology and Intensive Care Medicine, German Interdisciplinary Association for Intensive Care and Emergency Medicine (DIVI) as well as patents 15753 627.7, PCT/EP

medRxiv preprint doi: https://doi.org/10.1101/2024.12.30.24319760; this version posted December 30, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

2015/067731, 3 174 588, 10 2014 215 211.9, 10 2018 114 364.8, 10 2018 110 275.5, 50 2015 010 534.8, 50 2015 010 347.7, 10 2014 215 212.7.

Gunnar Lachmann and Maria Heinrich report grants from the BIH Charité Clinician Scientist Program during conduct of the study. Dr. Dschietzig reports personal fees from Immundiagnostik AG during the conduct of the study. Dr. Lammers-Lietz and Anton Wiehe report personal fees from Pharmaimage Biomarker Solutions GmbH during the conduct of the study. Dr. Lachmann reports personal fees from Sobi, the University of Zurich and Thieme outside the submitted work. Dr. Wolf receives fees from the Kompetenz-Centrum Qualitätssicherung. Dr. Stamatakis reports funding from Stephen Erskine Fellowship from Queens' College of the University of Cambridge, UK outside the BioCog study. Dr. Bresser reports funding from Alzheimer Nederland outside of the study. Dr. Gallinat received funding from the German Research Foundation (DFG), Federal Ministry of Education and Research (BMBF) and received payment for five lectures and presentations with about 1.500 € per presentation sponsored by Lundbeck, Janssen-Cilag and Boehringer. Dr. Heilmann-Heimbach receives personal fees from Life&Brain GmbH.

None of the other authors have a conflict of interest to declare.

medRxiv preprint doi: https://doi.org/10.1101/2024.12.30.24319760; this version posted December 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

References

1. Androsova G, Krause R, Winterer G, Schneider R. Biomarkers of postoperative delirium and cognitive dysfunction. Front Aging Neurosci. 2015;7:112.
2. Aldecoa C, Bettelli G, Bilotta F, Sanders RD, Aceto P, Audisio R, et al. Update of the European Society of Anaesthesiology and Intensive Care Medicine evidence-based and consensus-based guideline on postoperative delirium in adult patients. Eur J Anaesthesiol. 2024;41(2):81-108.
3. Aldecoa C, Bettelli G, Bilotta F, Sanders RD, Audisio R, Borozdina A, et al. European Society of Anaesthesiology evidence-based and consensus-based guideline on postoperative delirium. Eur J Anaesthesiol. 2017;34(4):192-214.
4. World report on Ageing and Health. Geneva, Switzerland: World Health Organisation; 2015 2015.
5. Winterer G, Androsova G, Bender O, Boraschi D, Borchers F, Dschietzig TB, et al. Personalized risk prediction of postoperative cognitive impairment - rationale for the EU-funded BioCog project. Eur Psychiat. 2018;50:34-9.
6. Humeidan ML, Reyes JC, Mavarez-Martinez A, Roeth C, Nguyen CM, Sheridan E, et al. Effect of Cognitive Prehabilitation on the Incidence of Postoperative Delirium Among Older Adults Undergoing Major Noncardiac Surgery: The Neurobics Randomized Clinical Trial. JAMA Surg. 2021;156(2):148-56.
7. Yang T, Yang H, Liu Y, Liu X, Ding YJ, Li R, et al. Postoperative delirium prediction after cardiac surgery using machine learning models. Comput Biol Med. 2024;169.
8. Zhao XX, Li JL, Xie XH, Fang ZJ, Feng Y, Zhong Y, et al. Online interpretable dynamic prediction models for postoperative delirium after cardiac surgery under cardiopulmonary bypass developed based on machine learning algorithms: A retrospective cohort study. Journal of Psychosomatic Research. 2024;176.

medRxiv preprint doi: https://doi.org/10.1101/2024.12.30.24319760; this version posted December 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

9. Oosterhoff JHF, Karhade AV, Oberai T, Franco-Garcia E, Doornberg JN, Schwab JH. Prediction of Postoperative Delirium in Geriatric Hip Fracture Patients: A Clinical Prediction Model Using Machine Learning Algorithms. Geriatr Orthop Surg Rehabil. 2021;12:21514593211062277.

10. Rossler J, Shah K, Medellin S, Turan A, Ruetzler K, Singh M, et al. Development and validation of delirium prediction models for noncardiac surgery patients. J Clin Anesth. 2024;93:111319.

11. Bishara A, Chiu C, Whitlock EL, Douglas VC, Lee S, Butte AJ, et al. Postoperative delirium prediction using machine learning models and preoperative electronic health record data. BMC Anesthesiol. 2022;22(1):8.

12. Racine AM, Tommet D, D'Aquila ML, Fong TG, Gou Y, Tabloski PA, et al. Machine Learning to Develop and Internally Validate a Predictive Model for Post-operative Delirium in a Prospective, Observational Clinical Cohort Study of Older Surgical Patients. J Gen Intern Med. 2021;36(2):265-73.

13. Dodsworth BT, Reeve K, Falco L, Hueting T, Sadeghirad B, Mbuagbaw L, et al. Development and validation of an international preoperative risk assessment model for postoperative delirium. Age Ageing. 2023;52(6).

14. Feinkohl I, Borchers F, Burkhardt S, Krampe H, Kraft A, Speidel S, et al. Stability of neuropsychological test performance in older adults serving as normative controls for a study on postoperative cognitive dysfunction. BMC Res Notes. 2020;13(1):55.

15. German National Cohort C. The German National Cohort: aims, study design and organization. Eur J Epidemiol. 2014;29(5):371-82.

16. Harrell F. Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis. New York: Springer; 2015.

17. Gaudreau JD, Gagnon P, Harel F, Tremblay A, Roy MA. Fast, systematic, and continuous delirium assessment in hospitalized patients: the nursing delirium screening scale. J Pain Symptom Manage. 2005;29(4):368-75.

medRxiv preprint doi: https://doi.org/10.1101/2024.12.30.24319760; this version posted December 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

18. Ribot JC, Lopes N, Silva-Santos B. gammadelta T cells in tissue physiology and surveillance. Nat Rev Immunol. 2021;21(4):221-32.

19. Feng L, Chen Y, Xu K, Li Y, Riaz F, Lu K, et al. Cholesterol-induced leucine aminopeptidase 3 (LAP3) upregulation inhibits cell autophagy in pathogenesis of NAFLD. Aging (Albany NY). 2022;14(7):3259-75.

20. Zhu C, Zhao J, Bibikova M, Leverson JD, Bossy-Wetzel E, Fan JB, et al. Functional analysis of human microtubule-based motor proteins, the kinesins and dyneins, in mitosis/cytokinesis using RNA interference. Mol Biol Cell. 2005;16(7):3187-99.

21. Xue C, Yao Q, Gu X, Shi Q, Yuan X, Chu Q, et al. Evolving cognition of the JAK-STAT signaling pathway: autoimmune disorders and cancer. Signal Transduct Target Ther. 2023;8(1):204.

22. Rodriguez S, Hug C, Todorov P, Moret N, Boswell SA, Evans K, et al. Machine learning identifies candidates for drug repurposing in Alzheimer's disease. Nat Commun. 2021;12(1):1033.

23. Sun CC, Zhou ZQ, Yang D, Chen ZL, Zhou YY, Wen W, et al. Recent advances in studies of 15-PGDH as a key enzyme for the degradation of prostaglandins. Int Immunopharmacol. 2021;101(Pt B):108176.

24. Lai Y, Lin P, Lin F, Chen M, Lin C, Lin X, et al. Identification of immune microenvironment subtypes and signature genes for Alzheimer's disease diagnosis and risk prediction based on explainable machine learning. Front Immunol. 2022;13:1046410.

25. Boschetti E, Caio G, Cervellati C, Costanzini A, Rosta V, Caputo F, et al. Serum zonulin levels are increased in Alzheimer's disease but not in vascular dementia. Aging Clin Exp Res. 2023;35(9):1835-43.

26. Romer TB, Jeppesen R, Christensen RHB, Benros ME. Biomarkers in the cerebrospinal fluid of patients with psychotic disorders compared to healthy controls: a systematic review and meta-analysis. Mol Psychiatry. 2023;28(6):2277-90.

medRxiv preprint doi: https://doi.org/10.1101/2024.12.30.24319760; this version posted December 30, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

27. Jomova K, Vondrakova D, Lawson M, Valko M. Metals, oxidative stress and neurodegenerative disorders. Mol Cell Biochem. 2010;345(1-2):91-104.

28. Giil LM, Midttun O, Refsum H, Ulvik A, Advani R, Smith AD, et al. Kynurenine Pathway Metabolites in Alzheimer's Disease. J Alzheimers Dis. 2017;60(2):495-504.

29. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology. 2010;21(1):128-38.