External Validation of a Model for Persistent Perfusion Deficit in Patients With Incomplete Reperfusion After Thrombectomy

EXTEND-PROCEED

Adnan Mujanovic, MD,* Felix C. Ng, MBBS, BMedSc, DCH,* Mattia Branca, PhD, Hannes A. Deutschmann, MD, Thomas R. Meinel, MD, PhD, Leonid Churilov, PhD, Oliver Nistl, MD, Peter J. Mitchell, MBBS, MMed, Nawaf Yassi, MD, PhD, Mark W. Parsons, MD, FRACP, PhD, Gagan J. Sharma, MCA, MITPM, Thomas Gattringer, MD, PhD, Marcel Arnold, MD, Fabiano Cavalcante, MD, Eike I. Piechowiak, MD, Timothy J. Kleinig, MD, PhD, David J. Seiffge, MD, Tomas Dobrocky, MD, Jan Gralla, MD, MSc, Urs Fischer, MD, MSc, Markus Kneihsl, MD, PhD, Bruce C.V. Campbell, MD,† and Johannes Kaesmacher, MD, PhD†

Correspondence
Dr. Kaesmacher
johannes.kaesmacher@
insel.ch

Abstract

Background and Objectives
We recently developed a model (PROCEED) that predicts the occurrence of persistent perfusion deficit (PPD) at 24 hours in patients with incomplete angiographic reperfusion after thrombectomy. This study aims to externally validate the PROCEED model using prospectively acquired multicenter data.

Methods
Individual patient data for external validation were obtained from the Endovascular Therapy for Ischemic Stroke with Perfusion-Imaging Selection, Tenecteplase versus Alteplase Before Endovascular Therapy for Ischemic Stroke part 1 and 2 trials, and a prospective cohort of the Medical University of Graz. The model's primary outcome was the occurrence of PPD, defined as a focal, wedge-shaped perfusion delay on 24-hour follow-up perfusion imaging that corresponds to the capillary phase deficit on last angiographic series in patients with <Thrombolysis in Cerebral Infarction 3 reperfusion after thrombectomy. The model's performance was evaluated with discrimination, calibration accuracy, and clinical decision curves.

Results
We included 371 patients (38% with PPD). The externally validated model had good discrimination (C-statistic 0.81, 95% CI 0.77–0.86) and adequate calibration (intercept 0.25, 95% CI 0.21–0.29 and slope 0.98, 95% CI 0.90–1.12). Across a wide range of probability thresholds (i.e., depending on the physicians' preferences on how the model should be used), the model shows net benefit on clinical decision curves, informing physicians on the likelihood of PPD. If a physician's attitude toward false-positive and false-negative ratings is equal, the model would reduce 13 in 100 unnecessary interventions by correctly predicting complete delayed reperfusion, without missing a patient with PPD.

---

*These authors contributed equally to this work as co-first authors.

†These authors contributed equally to this work as co-senior authors.

From the Departments of Diagnostic and Interventional Neuroradiology (A.M., E.I.P., T.D., J.G., J.K.), Neurology (T.R.M., M.A., D.J.S., U.F.), University Hospital Bern, Inselspital, Graduate School for Health Sciences (A.M.), and CTU Bern (M.B.), University of Bern, Switzerland; Department of Medicine and Neurology, Melbourne Brain Centre (F.C.N., N.Y., G.J.S., B.C.V.C.), Melbourne Medical School (L.C.), and Department of Radiology (P.J.M.), Royal Melbourne Hospital, University of Melbourne, Parkville; Department of Neurology (F.C.N.), Austin Health, Heidelberg, Australia; Division of Neuroradiology, Vascular and Interventional Radiology Department of Radiology (H.A.D., O.N., T.G., M.K.), and Department of Neurology (T.G., M.K.), Medical University of Graz, Austria; Population Health and Immunity Division (N.Y.), The Walter and Eliza Hall Institute of Medical Research, Parkville; Department of Neurology (M.W.P.), Liverpool Hospital, University of New South Wales, Sydney, Australia; Department of Radiology and Nuclear Medicine (F.C.), Amsterdam University Medical Centers, University of Amsterdam, Amsterdam Neuroscience, the Netherlands; Department of Neurology (T.J.K.), Royal Adelaide Hospital, Australia; and Department of Neurology (U.F.), University Hospital Basel, University of Basel, Switzerland.

Go to Neurology.org/N for full disclosures. Funding information and disclosures deemed relevant by the authors, if any, are provided at the end of the article.

The Article Processing Charge was funded by the authors.

Glossary

aOR = adjusted odds ratio; **ASITN/SIR** = American Society of Intervention and Therapeutic Neuroradiology and Society of Interventional Radiology; **DR** = delayed reperfusion; **eTICI** = expanded Thrombolysis in Cerebral Infarction; **EXTEND-IA** = Endovascular Therapy for Ischemic Stroke with Perfusion-Imaging Selection; **EXTEND-IA TNK** = Tenecteplase versus Alteplase Before Endovascular Therapy for Ischemic Stroke; **PPD** = persistent perfusion deficit; **PROCEED** = PeRfusion OutComE prEDiction; **RCT** = randomized controlled trial; **TICI** = Thrombolysis in Cerebral Infarction.

Discussion

The externally validated model had adequate predictive accuracy and discrimination. Depending on the acceptable threshold probability, the model accurately predicts persistent incomplete reperfusion and may advise physicians whether additional reperfusion attempts should be performed.

Introduction

The goal of endovascular therapy for patients with acute ischemic stroke is complete reperfusion, defined as Thrombolysis in Cerebral Infarction (TICI) score 3.[1,2] However, the benefit of additional reperfusion attempts among patients who achieve technically successful but incomplete angiographic reperfusion (TICI2b) varies considerably because of different prognostic factors and unknown natural evolution of the distal occlusion (i.e., spontaneous delayed reperfusion [DR] vs persistent deficit progressing to infarction).[3-5] More than half of endovascularly treated stroke patients with incomplete reperfusion have complete DR of remaining vessel occlusions, with no or very few minor infarct evolutions and a good clinical outcome.[4] Among these patients, adjuvant reperfusion rescue strategies (e.g., secondary mechanical thrombectomy, intra-arterial thrombolytics, antithrombotics) are likely unnecessary or potentially harmful.[3,6]

To inform interventionalists on the chances of favorable natural evolution of the distal occlusion, that is, DR vs an unfavorable natural disease evolution, that is, persistent perfusion deficit (PPD), we have previously developed and internally validated the PeRfusion OutComE prEDiction (PROCEED) model.[7] PROCEED is a prognostic model that predicts the occurrence of PPD at 24 hours after incomplete reperfusion (TICI <3) has been observed on the final digital subtraction angiography runs. The model consists of 7 predictors and has been validated internally. However, being validated only on a single-center stroke registry data limits the generalizability of the model.[7] Therefore, it remains uncertain whether the PROCEED model could be applied in other settings and health care systems.[8-10] If the PROCEED model performs well on an external data set, it could inform the decision making whether to pursue additional reperfusion attempts or stop the procedure once incomplete reperfusion with remaining distal vessel occlusions is encountered.

To evaluate the performance of the model further, we performed an external validation study using pooled international multicenter data of patients who received perfusion imaging on the 24-hour follow-up examination after endovascular treatment.

Methods

Standard Protocol Approvals, Registrations, and Patient Consents

External validation has been performed using pooled individual-patient data from 3 randomized controlled trials (RCTs) and a prospective single-center cohort. This study was performed according to the Declaration of Helsinki and was approved by the ethics committee (KEK ID 231/2014). Informed consent was obtained from all participants or legal representatives. Study results have been reported according to the transparent reporting of multivariable prediction models developed or validated using clustered data (TRIPOD-Cluster) statement.[11]

External Data Set

Individual patient data for external validation were obtained from the Endovascular Therapy for Ischemic Stroke with Perfusion-Imaging Selection (EXTEND-IA), Tenecteplase Versus Alteplase Before Endovascular Therapy for Ischemic Stroke (EXTEND-IA TNK) part 1 and part 2 trials (ClinicalTrials.gov, unique identifier: NCT01492725, NCT02388061, and NCT03340493, respectively).[12-14] Methodology of these RCTs has been described previously.[12-14] For all of these RCTs, written informed consent was obtained from patients or legal representatives and institutional ethics committee approved the study at each recruitment site. We also pooled the data from the prospective study cohort of the Medical University Graz, a tertiary-level hospital serving as a reference center for patients with stroke in southern Austria. This particular data set was used for external validation only and consists of prospectively enrolled consecutive patients from 2018 to 2022 undergoing endovascular therapy with perfusion imaging on scheduled follow-up (ClinicalTrials.gov, unique identifier: NCT05273216).[15] Only patients who presented

with acute ischemic stroke due to large vessel occlusions in the anterior circulation, final expanded TICI (eTICI) grade 2a–2c, and available perfusion imaging on 24 ± 12 hour follow-up were included in the final analysis.

Primary Outcome

The primary prediction outcome of the model is perfusion status on follow-up perfusion maps, which was dichotomized into PPD or DR. Perfusion imaging outcome was evaluated by an independent core laboratory, blinded to clinical and technical details, using the same methodology that was outlined in the charter of the original model.[7] In brief, PPD refers to a focal, wedge-shaped perfusion delay on follow-up perfusion maps that corresponds to the capillary phase deficit on the last angiography series. If the perfusion deficit was smaller than the angiographic deficit, but still suggestive of a distal vessel occlusion within the angiographically hypoperfused territory, it was rated as PPD. Conversely, DR was defined as the absence of any focal perfusion deficit on follow-up perfusion maps, despite incomplete reperfusion on the final angiography series (Figure 1). PPD was chosen to be the predicted outcome, because decision curves are usually shown with respect to prediction of the disease (e.g., cancer).[16-18]

Perfusion Postprocessing Protocols

In all 3 RCTs, perfusion imaging was performed on both admission and the 18- to 30-hour follow-up examination, as part of the prespecified study protocol.[12-14] Raw CT perfusion maps from the admission in these RCTs were postprocessed with the same software as the CT perfusion maps in the original model (syngo.via; Siemens, Erlangen, Germany).[7] Perfusion maps from the follow-up were postprocessed with a commercially available software (RAPID; iSchemaView, Inc., Menlo Park, CA), as per the RCT protocols. Patients from the

Figure 1 Perfusion Imaging Outcome in Patients With Incomplete Reperfusion
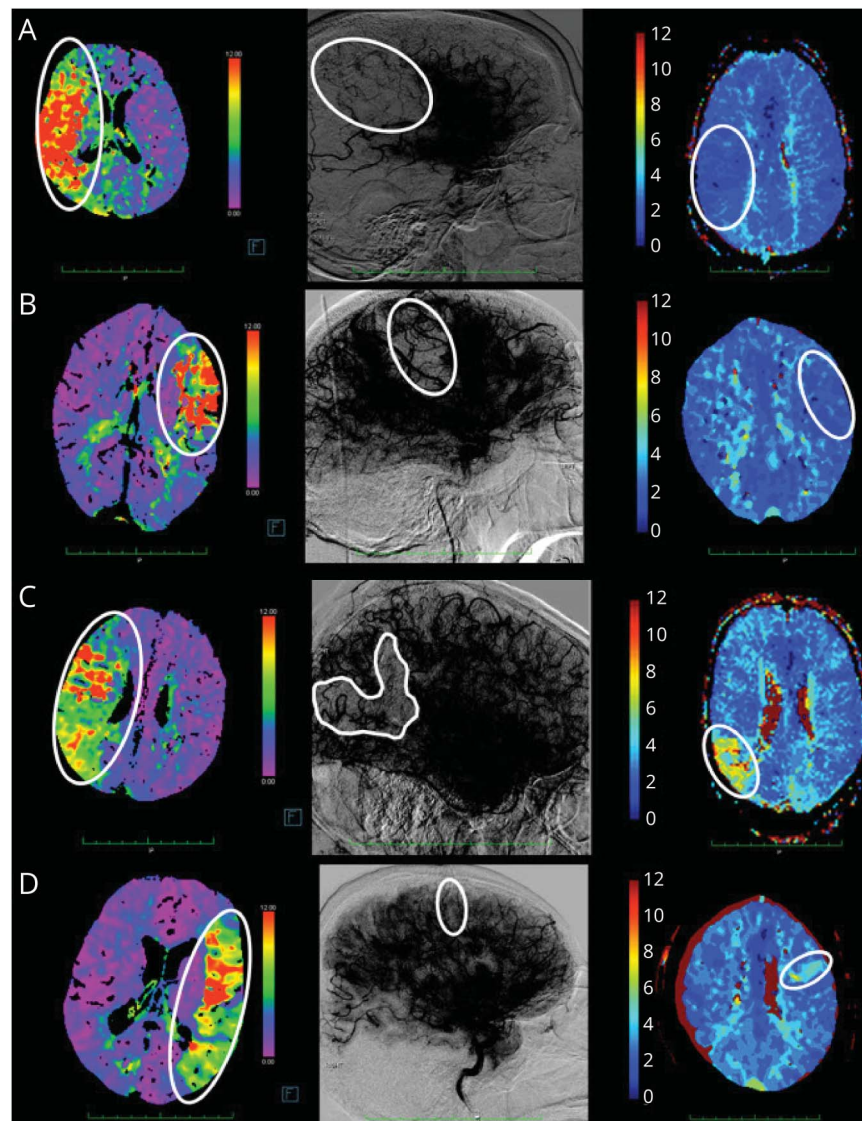
Perfusion imaging outcome was evaluated on time-to-peak and time-to-maximum perfusion maps, because of their high sensitivity for vessel reperfusion status. Final angiography runs are displayed with high contrast to emphasize the capillary phase deficits. (A) Preinterventional perfusion imaging shows a right-side internal carotid artery occlusion (left panel). Final DSA run shows incomplete reperfusion (eTCI 2b50, middle panel). On the follow-up perfusion imaging, there is no visible deficit despite incomplete reperfusion on the final DSA run (right panel). (B) Preinterventional perfusion imaging shows left-side M2 occlusion (left panel). Final DSA run shows incomplete reperfusion (eTICI 2b50, middle panel). On the follow-up perfusion imaging, there is no visible deficit despite incomplete reperfusion on the final DSA run (right panel). (C) Preinterventional perfusion imaging shows right-side M1 occlusion (left panel). Final DSA run shows incomplete reperfusion (eTICI 2b67, middle panel). On the follow-up perfusion imaging, there is a visible perfusion deficit directly corresponding to the area of incomplete reperfusion on the final DSA run (right panel). (D) Preinterventional perfusion imaging shows left-side M1 occlusion (left panel). Final DSA run shows incomplete reperfusion (eTICI 2c, middle panel). On the follow-up perfusion imaging, there is a visible perfusion deficit directly corresponding to the area of incomplete reperfusion on the final DSA run (right panel) DSA = digital subtraction angiography; eTICI = expanded Thrombolysis in Cerebral Infarction.

prospective single-center cohort from Graz underwent contrast-enhanced MR perfusion imaging as per the institution's protocol.[15] The same postprocessing software for MRI perfusion was used in this center as in the original model (Olea Sphere; Olea Medical, La Ciotat, France).[7] All anonymized imaging sequences used for the analysis were transferred using a secure cloud-based server (eTable 1). Concordance between perfusion maps postprocessed with syngo.via and Olea Sphere was already reported in the original model[7]; however, the original model did not use the RAPID software for perfusion postprocessing. Therefore, concordance of ratings for PPD vs DR on perfusion maps processed with a different software was evaluated and reported with the inter-rater agreement on a random patient sample (n = 60, eFigure 1 and respective caption for methodology).

Image Analysis

Patient demographic and baseline clinical characteristics were extracted from electronic case report forms of randomized trial data or from prospective databases (in case of the cohort from the Medical University Graz). To reduce the likelihood for inconsistent readings, all imaging and interventional data were re-evaluated centrally according to the methods outlined in the original model.[7] In short, collateral grading was performed with the American Society of Intervention and Therapeutic Neuroradiology and Society of Interventional Radiology (ASITN/SIR) Collateral Flow Grading System on pretreatment angiography images, and angiographic reperfusion was graded with the expanded TICI (eTICI) scale on the final angiography series. Intervention-to-follow-up time was defined as the time frame beginning with the final angiography series run and ending with the first follow-up imaging series.

Statistical Analysis

Statistical handling of all variables is reported in eTable 2. Categorical and continuous variables were compared between 2 groups with Fisher exact and Mann-Whitney U tests, respectively. Statistical evaluation of heterogeneity of baseline risk for PPD across the studies was performed in 2 independent steps.[11,19] First, fixed-effects logistic regressions were used to estimate study-specific risk estimates of baseline risk for PPD. Afterward, mixed-effects logistic regression models were used to account for the heterogeneity between the studies by using a random intercept adjusted for clustering at the study level. Heterogeneity is reported with the estimated random-effect variance for each variable, which quantifies the variance of variable-specific effect sizes across the studies.[11,19] Only cases with all available variables were included in the final analysis. All statistical analyses were performed in R statistical software (version 4.0.0).

Performance on External Data Set

All the predictors from the original model were used in the external validation: age, sex, atrial fibrillation, device passes, final eTICI score, and ASITN/SIR collateral score.[7] Point estimates of these predictors from the original model were applied to the external data set, and performance is reported with cross tables, accuracy, discrimination, calibration, and clinical decision curves. Cross tables were calculated for different threshold probabilities: 0.2, 0.3, 0.33, 0.5, 0.66, 0.7, and 0.8. Discrimination refers to the model's probability of correctly distinguishing patients with and without the primary end point of interest (i.e., patient with and without PPD). Discrimination was quantified with concordance statistics (C-statistics) that varies from 0.5 (noninformative) to 1 (perfect discrimination).[8-10] In case of binary outcomes, C-statistics corresponds to the area under the receiver-operating characteristic curve.[8-10] Calibration reflects the extent to which predicted risks align with proportions of observed outcomes.[8-10] Bootstrapped resampling (n = 1,000) was performed to obtain the 95% CI for all model metrics. Shrinkage of prediction uncertainty was performed by calculating prediction intervals for the occurrence of having PPD across predefined time thresholds after the intervention. We used precision-based sample size analysis by Pavlou et al.[20] to calculate the required number of patients for external validation. Based on previous reports, we assumed C-statistic = 0.8 with standard error = 0.025 and PPD prevalence = 40%.[4,7,21] This led to the recommended sample size of n = 315 (including 126 event cases, i.e., patients with PPD).

Decision Curves

Clinical decision curves are used to advise on the clinical relevance of the model and can be quantified with net benefit and net reduction across different threshold probabilities (R).[16-18] The unit of net benefit is true positives, which is correctly classified PPD in the present context. A net benefit of 0.10 means that, compared with assuming that all patients do not have PPD, 10 true-positive patients for every 100 patients in the target population can be identified, that is, correctly identifying 10 patients per 100, all of whom would develop PPD if the interventionalist does not perform any additional reperfusion efforts after observing incomplete reperfusion at the end of an intervention. A model is considered superior to another model or scenario if the net benefit of the former surpasses that of the latter at the predefined clinically meaningful threshold probability.[16-18]

Net reduction is weighted with true negatives, which are the rates of DR in the present context. A net reduction of 0.10 means 10 true-negative patients for every 100 patients in the target population, that is, correctly identifying 10 patients per 100, all of whom would develop DR if the interventionalist does not pursue further reperfusion attempts. In other words, at the given probability threshold, omitting the intervention based on the advice of the model could reduce the number of interventions in 10 of 100 patients without missing a single patient who would develop a PPD. In general, high net benefit and net reduction are preferred. To calculate net benefit and net reduction, a clinically relevant range of threshold probabilities need to be defined.[16-18]

In the current analysis, the PROCEED model is used as a prognostic model. Although PPD is a serious event associated

with infarct growth, the benefit of additional reperfusion attempts is somewhat unclear and could be followed by side effects (intracranial hemorrhage, perforation, additional thrombus formation). Owing to differences in options concerning the value of additional reperfusion attempts, as well as differences between patients and remaining occlusion patterns, the threshold probability used to determine additional reperfusion attempts may vary. This threshold will depend on what the treating team is more concerned about: If the team is more concerned about missing and not treating a patient with PPD, the team would choose a low threshold. If the team is more concerned about the risk of additional reperfusion attempts and intervening on patients who will not have PPD, the team would choose a high threshold probability.

For example, choosing a threshold of 50% would mean that missing a PPD and falsely informing the physician not to perform a potentially beneficial intervention is equally worse to predicting a PPD—although the patient would have DR—thus falsely informing the physicians to perform an unnecessary and potentially harmful intervention. Owing to the present lack of evidence on this topic, we evaluated the model across a range of clinically reasonable threshold probabilities (R = 33%–66%). For example, a threshold probability of 33% would reflect a physician's opinions that it is twice as worse to miss a PPD (i.e., falsely assume no PPD and not treat a patient who develops PPD) vs falsely predicting PPD (i.e., falsely assume PPD and treat a patient who does not develop PPD). Vice versa, a threshold probability of 66% would reflect that it is twice as worse to false-positively predict PPD (i.e., falsely assume PPD and treat a patient who does not develop PPD) vs missing a true PPD (i.e., falsely assume no PPD and not treat a patient who develops PPD). Decision curves are presented graphically across a range of threshold probabilities in comparison with "Assume PPD in All," equivalent to a "Treat All" scenario, and "Assume PPD in None," equivalent to a "Treat None" scenario.

Sensitivity Analysis

For sensitivity purposes, we have performed 2 separate pre-specified analyses. In the first analysis, we have compared performance of the PROCEED model with that of model 2. Model 2 assumes that all patients with final reperfusion grade eTICI2b50-67 will develop PPD and those with eTICI2c will develop DR on follow-up perfusion imaging. Model 2 is based on the argument that interventionalists tend to pursue additional reperfusion attempts when only 50%–89% reperfusion of the target downstream territory is achieved, as present guidelines state that complete reperfusion (eTICI3) should be achieved whenever possible.[1,2] Conversely, interventionalists might not pursue any additional device passes once 90% of reperfusion is established because of the higher risk-to-benefit ratio in this subgroup.[22,23] Performances of model 2 is reported with the same metrics used for the PROCEED model. In the second sensitivity analysis, we have excluded 2 predictors from the PROCEED model (atrial fibrillation and intervention-to-follow-up time) and compared this version of

the PROCEED model (named "PROCEED 2" further below) against model 2. The rationale for excluding atrial fibrillation as a predictor was that it includes few patients (estimated <20%) who had newly detected atrial fibrillation during their acute stay—and this would happen after the intervention, that is, this predictor would not be available at the time point when the prediction model is meant to be used. In addition, timing of the follow-up perfusion imaging after the intervention is different across centers and not helpful for real-time decision making.

Data Availability

Anonymized study data are available from the corresponding author upon reasonable request after receipt of a research plan and clearance by the appropriate ethics committee.

Results

We included 371 patients in the final analysis. From the initially screened cohort, 371 of 518 (72%) patients fulfilled the inclusion criteria. Most common exclusion reasons were lack of follow-up imaging; perfusion imaging outside of the 24 ± 12 hour window; or final eTICI score 0, 1, and 3 (eFigure 2). Median age was 73 years (interquartile range 64–80), 52% were male, and 38% had PPD. On average, patients with DR were more likely to have a lower number of device passes (DR vs PPD: 1 [1–2] vs 2 [1–3]; $p < 0.001$), better collateral status (DR vs PPD: 2 [1–3] vs 1 [1–2]; $p < 0.001$), better final reperfusion score (DR vs PPD for eTICI2c: 48.9% vs 15%; $p < 0.001$), and longer intervention-to-follow-up time (DR vs PPD: 23 hours 16 minutes vs 21 hours 23 minutes; $p < 0.001$; Table 1). There was no heterogeneity between the studies used for the internal and external validation (eTable 3 and eFigure 3). Inter-rater agreement for evaluating perfusion imaging outcome on different perfusion processing software was good (Krippendorf alpha 0.84, 95% CI 0.76–0.90 for a random sample of 60 cases). Notably, patients with incomplete reperfusion at the end of an intervention who experienced PPD had lower rates of good clinical outcome than patients with DR (53.6% vs 66.7%; $p = 0.016$ for modified Rankin Scale 0–2). This association was present irrespective of the final reperfusion score (eFigure 4). Of the 371 included patients, 11 (3%) were from EXTEND IA, 103 (28%) from EXTEND IA TNK part 1, 153 (41%) from EXTEND IA TNK part 2, and 104 (28%) from the University Hospital Graz. Patient characteristics stratified across pooled studies are presented in eTable 4.

Predictor effects were comparable between the population in the internal and external validation. There were significant associations between PPD, eTICI (adjusted odds ratio [aOR] 0.3, 95% CI 0.2–0.4), collateral score (aOR 0.6, 95% CI 0.4–0.8), and intervention-to-follow-up time (aOR 0.9, 95% CI 0.8–0.9). We also found an association between PPD, device passes (aOR 1.8, 95% CI 1.4–2.3), and age (aOR 1.02, 95% CI 1.0–1.1); however, no association was found between

**Table 1** Population Baseline and Interventional Characteristics

	Overall	Delayed reperfusion	Persistent perfusion deficit	*p* Value
n	371	231	140	
Age, y, median (IQR)	73 (64–80)	73 (62–80)	73 (65–80)	0.643
Sex, male, n (%)	193 (52.0)	119 (51.5)	74 (52.9)	0.886
Atrial fibrillation, yes, n (%)	124 (33.4)	81 (35.1)	43 (30.7)	0.455
Device passes, median (IQR)	2 (1–2)	1 (1–2)	2 (1–3)	<0.001
ASITN/SIR collateral score, median (IQR)	2 (1–2)	2 (1–3)	1 (1–2)	<0.001
eTICI, n (%)				<0.001
2a	26 (7.0)	5 (2.2)	21 (15.0)	
2b50	93 (25.1)	30 (13.0)	63 (45.0)	
2b67	118 (31.8)	83 (35.9)	35 (25.0)	
2c	134 (36.1)	113 (48.9)	21 (15.0)	
Intervention-to-follow-up, h, median (IQR)	22.7 (20.1–25.0)	23.3 (20.8–25.8)	21.4 (19.1–23.3)	<0.001

Abbreviations: ASITN/SIR = American Society of Intervention and Therapeutic Neuroradiology and Society of Interventional Radiology; eTICI = expanded Treatment in Cerebral Infarction; IQR = interquartile range.

PPD and atrial fibrillation (aOR 0.8, 95% CI 0.4–1.4, eTable 5).

The model had good discrimination (C-statistics 0.81, 95% CI 0.77–0.86) and adequate calibration, as seen through calibration intercept 0.25 (95% CI 0.21–0.29) and slope 0.98 (95% CI 0.90–1.12, Figure 2). At a 50% threshold probability, the model correctly classified most of the patients, with moderate sensitivity (69%, 95% CI 60%–76%), specificity (79%, 95% CI 73%–84%), and accuracy (75%, 95% CI 70%–79%; Figure 3). At extreme threshold probabilities, for example, 20% or 80%, the model has high sensitivity (94%) at the cost of low specificity (39%) or vice versa—low sensitivity (27%) and high specificity (95%, eFigure 5).

Figure 2 Calibration Plot
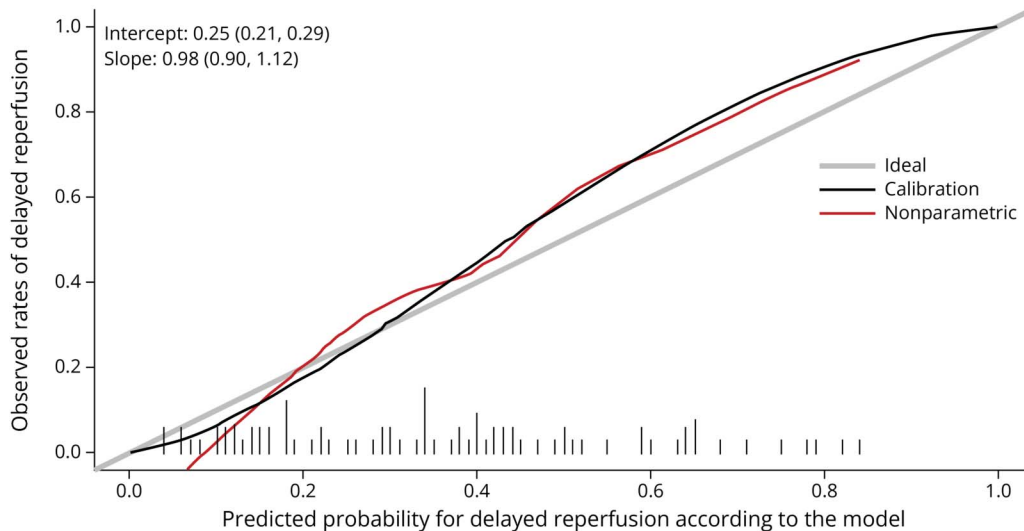
Calibration is usually reported with intercept and slope. Intercept compares the means between all predicted and observed risks and informs on the value of predictors. The slope provides information on strength of the model's predictors. Perfect predictors align directly on the calibration line with an intercept having a value of 0 and slope value of 1. The model had adequate calibration with intercept having a value of 0.25 (95% CI 0.21–0.29) and slope 0.98 (95% CI 0.90–1.12)

Figure 3 Performance of the Model at the Threshold Probability of 50%
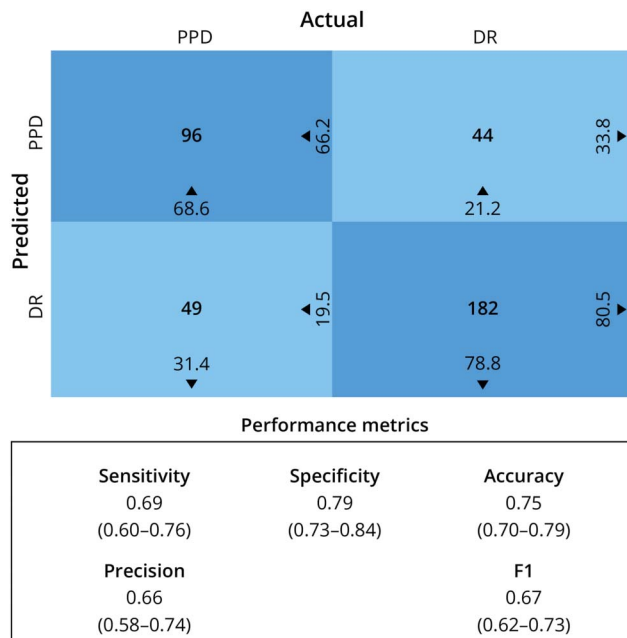
Confusion matrix can be divided into 4 fields: true positives (upper left corner, n = 96), false positives (lower left corner, n = 49), false negatives (upper right corner, n = 44), and true negatives (lower right corner, n = 182). At the bottom of each field are the column percentages, and on the right side are the row percentages. DR = delayed reperfusion; PPD = persistent perfusion deficit.

On clinical decision curves, the PROCEED model performed better than "Assume PPD in All" scenario (corresponding to a "Treat All" scenario) or "Assume PPD in None" scenario (corresponding to a "Treat None" scenario) across a wide range of threshold probabilities. Using the model at a very low threshold probability of R <20% would provide no or only marginal additional benefit compared with a "Assume PPD in All"/"Treat All" scenario (see overlap of the curve "PRO-CEED" and "Assume PPD in All" in Figure 4 top panel). At a threshold probability of R = 50%, the model has a net benefit of 0.13, that is, the model correctly identifies 13 of 100 patients, all of whom would develop PPD and could potentially benefit from adjunctive reperfusion attempts, without missing a PPD. In other words, one could identify 13–0 = 13 patients more (of 100) with correctly predicted PPD when using the model as opposed to the "Assume PPD in All" or "Assume PPD in None" scenario. Improvements in net benefit can also be seen across other threshold probabilities. For example, at R = 40% or 70%, one could identify 17 or 7 (of 100) patients more with correctly predicted PPD, respectively (Table 2). In terms of standardized net reduction, using the model with the threshold probability of R >80% would provide no or only marginal additional information, compared with assuming PPD in none (see overlap of the curve "PROCEED" and "Assume PPD in None" in Figure 4 bottom panel). When using the threshold probability of R = 50%, the model could reduce the number of interventions in 13 of 100 patients

without missing any patient who would eventually develop a PPD. Comparable trends in net reduction were observed across other threshold probabilities, for example, of R = 40% or R = 70% (Table 2).

When compared with model 2, the PROCEED model had significantly better discrimination (0.61, 95% CI 0.56–0.65 vs 0.81, 95% CI 0.77–0.86) and calibration (calibration intercept −0.52, 95% CI −0.91 to −0.21 vs 0.25, 95% CI 0.21–0.29). The PROCEED model also performed better across a wide range of threshold probabilities with higher net benefit and net reduction than model 2 (eTable 6 and eFigure 6). Similarly, PROCEED 2 had significantly better performance than model 2 across all relevant metrics (C-statistics: 0.80, 95% CI 0.76–0.85 vs 0.61, 95% CI 0.56–0.65; calibration intercept: 0.16, 95% CI 0.02–0.50 vs −0.52, 95% CI −0.91 to −0.21) with higher net benefit and net reduction (eTable 7 and eFigure 7). An open-access online web tool for complementing clinical decision making has been made available at: proceed.shinyapps.io/model/.

Discussion

The PROCEED model shows good discrimination and adequate calibration for prognosticating the occurrence of PPD in a pooled international multicenter data set comprising 12 comprehensive thrombectomy centers across Europe, Australia, and New Zealand. The predictors' contributions to the model for PPD were comparable between the initial and external cohorts. Depending on the acceptable threshold probability, this model could correctly identify patients who are more likely to benefit from additional reperfusion efforts and thus can be used to complement clinical decision making.

Continuous improvements in medicine have led to an increased number of prediction models and clinical decision support systems.[8] However, a systematic review reported that most published prediction models suffer methodological shortcomings, the most often cited one being lack of validation on an external data set.[24] Only 5% of all published prediction models in medicine provide information on external validation.[8,25] In this analysis, we have performed an independent external validation, where the external population was assembled in a completely separate manner from the initial cohort. Our external data set comprised pooled data from 3 RCTs and 1 prospective observational tertiary-level single-center cohort.[12-15] Patients enrolled into RCTs have to fulfill a list of inclusion and exclusion criteria as defined by the trial protocol. Therefore, they represent a highly selected patient subgroup that often differs from routinely treated stroke patients, who were used in the original creation of the model.[26,27] All 3 of these RCTs have randomized patients in hospitals across Australia and New Zealand, which have substantially different health care systems than the one in Switzerland, where the initial study population originated from.[7] By contrast, the prospective cohort study of

Figure 4 Clinical Decision Curves
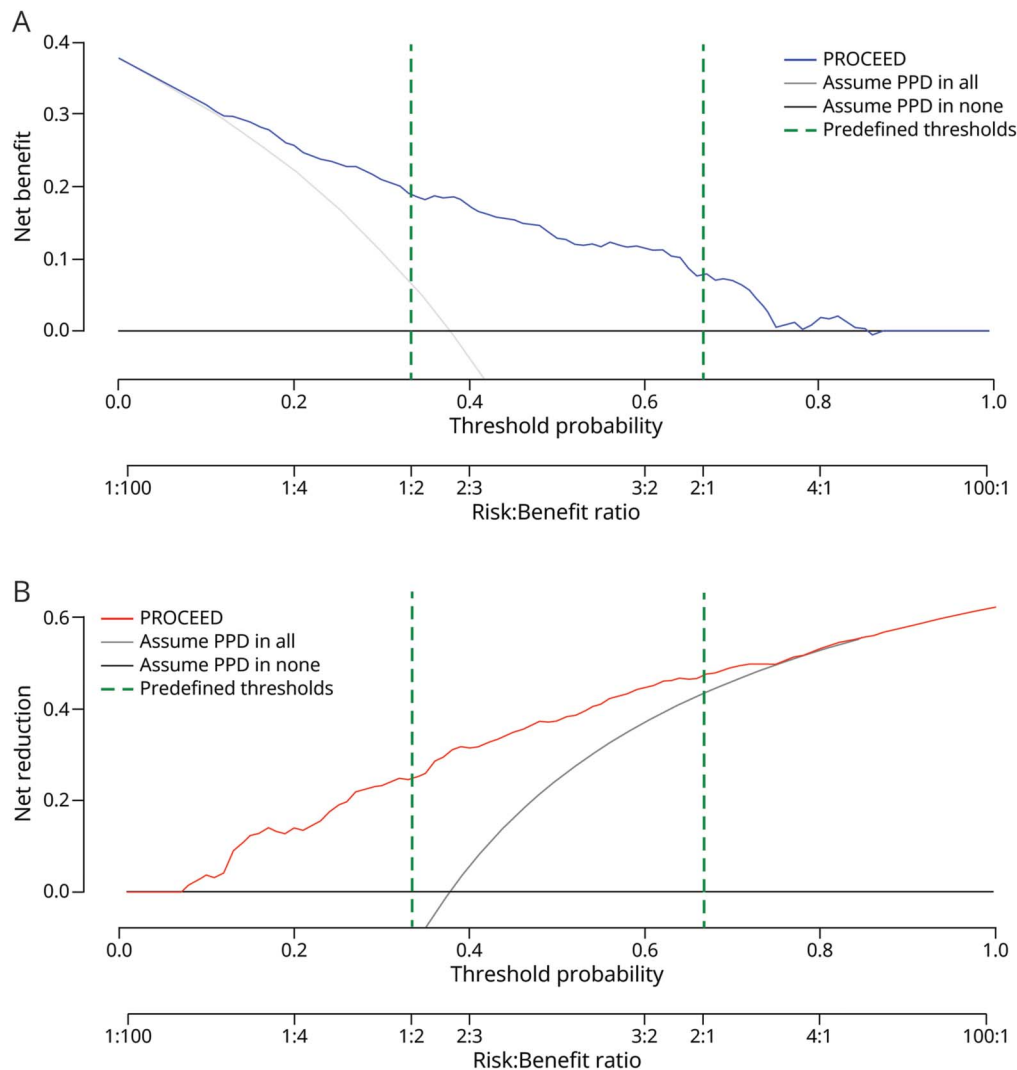
(Top panel) Net benefit of the prediction model and "Assume PPD in All" or "Treat All" option overlap in the threshold probability range from 0% to 20%. Using the PROCEED model, in this threshold probability range would have no added value for the classification of perfusion outcome. At a threshold probability of R = 50%, the model would have a net benefit of 0.13. (Bottom panel) In comparison with the "Assume PPD in All" or "Assume PPD in None" scenario and across a large range of threshold probabilities, the use of the model shows higher net reduction and could reduce the number of interventions performed. For example, at a threshold probability of 50%, the intervention could be avoided in 13 of 100 patients without missing a single patient who would eventually develop a PPD.

the Medical University Graz is reflective of real-world patients and practices in a different health care system when compared with that in the initial model.[15] These geographic disparities, different study time frames, and distinct methods used for patient inclusion between the internal and external cohorts offer a strong foundation for robust validation of the model.[10] Despite RCTs' strict inclusion criteria and difference in routinely treated clinical patients, we did not observe substantial changes between model's internal and external performance. This highlights the robustness of the model's performance in various clinical contexts and the applicability of its findings across different health care systems. Having a model externally validated on a population comprising RCT and real-world stroke patients might help bridge the gap between guideline recommendations and individualized patient approach.[28]

Evidence on further management of incomplete reperfusion (<TICI3) in patients at the end of an intervention is lacking.[3-5] These patients are presently being treated on an ad hoc basis at the discretion of the operator.[6] However, more than half of <TICI3 patients will have favorable natural evolution of incomplete reperfusion (i.e., DR) with good clinical outcome.[4] We have observed the same trend in the present cohort and saw a close association between the final eTICI score and the occurrence of PPD on follow-up. However, it is difficult to say whether the patient will subsequently develop DR or PPD when the intervention ends with incomplete reperfusion and pursuing additional reperfusion efforts among these patients may be unwarranted and even potentially harmful.[4] Evolution of tissue reperfusion is a complex dynamic dependent on many factors, such as patients' baseline status, final reperfusion score,

Table 2 Net Benefit and Net Reduction of the PROCEED Model

| | PROCEED vs assume PPD in all | | | | |
| | Net benefit | | | Advantage of PROCEED | |
R (%)	PROCEED	Treat all	Treat none	ΔNet benefit[a]	Net reduction
20	25.67385	22.16981	0	3.50404	14.01617
30	20.98575	11.66761	0	9.31814	23.18059
33	19.14551	7.01832	0	12.12719	24.52013
40	17.25067	−0.37735	0	17.25067	28.71159
50	12.93801	−2.45283	0	12.93801	12.93801
60	11.45553	−4.65691	0	11.45553	7.637017
66	7.61059	−7.79358	0	7.61059	3.920608
70	7.00809	−10.75471	0	7.00809	3.003466
80	1.88679	−21.13207	0	1.88679	0.471698

R = threshold probability. Net benefit and net reduction are given per 100 patients. Using the model in the threshold probability range of R <20% or R >80% would provide no or only marginal additional information for the treating team because of the overlap of PROCEED with the "Assume PPD in All"/"Treat All" and "Assume PPD in None"/"Treat None" scenarios (see Methods and Results).
[a] ΔNet benefit was calculated by subtracting net benefit of either "Treat All" or "Treat None" scenarios from the net benefit of the PROCEED model, whichever one was higher for a given threshold probability.

collateral status, number of device passes, to name a few. Further management of these cases might be challenging and potentially harmful when only one of these factors is considered (e.g., all TICI2b patients will develop PPD and, therefore, should be treated further). In scenarios where multiple patient characteristics ought to be considered, having a validated prognostic model on natural evolution of incomplete reperfusion could provide potentially relevant information to the treating operator.[7-9]

Overall, the PROCEED model seems to perform well in an external population with good discrimination and adequate calibration. The model also showed high net benefit across a wide array of clinically relevant threshold probabilities (e.g., R = 33%–66%). Whether it is more dangerous for a PPD to be falsely predicted as DR, than a DR being falsely predicted as PPD, is still a matter of debate. Large heterogeneity on threshold probabilities can be expected among treating physicians,[29] particularly because evidence on the management of secondary distal vessel occlusions is more unclear than the treatment of primary distal vessel occlusions.[3-6] Therefore, the main results of the model are reported under the assumption that missing a high-likelihood PPD and not performing a potentially beneficial intervention vs falsely predicting PPD and prompting the physicians to perform a potentially harmful intervention are equally worse (i.e., R = 50%). A crucial understanding on how to use and interpret the model depends on what threshold probabilities would an operator consider for proceeding or stopping with additional reperfusion attempts after seeing incomplete reperfusion. Lower threshold probabilities carry a risk of exposing patients to potentially unnecessary and harmful additional reperfusion attempts whereas with higher threshold probabilities, interventions may be omitted for patients who could likely benefit from them. Owing to the present lack of evidence on when to proceed or stop with additional reperfusion attempts,[1,2] we have also created model 2 and compared the performance between the PROCEED model and model 2. Model 2 assumes the decision-making process of an interventionalist given the current guideline recommendations and reported standards of practice, once half (TICI2b) or near-complete (TICI2c) reperfusion has been achieved.[1,2,22,23] When comparing the 2 models, the PROCEED model demonstrated better performance than model 2 across all relevant metrics for prognostic models (discrimination, calibration, decision curves). Comparable results were shown even when excluding predictors from the PROCEED model that may not be known at the time point of admission (e.g., new findings of atrial fibrillation). All these additional subanalyses suggest that the PROCEED model may offer clinically relevant information on the management of incomplete reperfusion, given appropriate threshold probabilities.

This study has several limitations. Patients without perfusion imaging on follow-up were excluded from our analysis. In the initial study population, 508 of 914 (55%) patients underwent follow-up perfusion imaging after endovascular therapy, and in the EXTEND-IA, EXTEND-IA TNK parts 1 and 2 RCTs, and Medical University of Graz, these numbers were 11 of 34 (32%), 115 of 152 (76%) and 157 of 185 (85%), and 170 of 482 (35%), respectively. Multimorbidity, worse baseline profile, or sudden clinical deterioration make it less likely for a

patient to undergo perfusion imaging on follow-up.[4,21] Therefore, rates of reported PPD should be handled cautiously because they are likely to differ from the absolute PPD rates among patients without follow-up perfusion imaging. Furthermore, reference standard for the management of incomplete reperfusion does not exist; therefore, performance of this model could not be compared with an appropriate reference threshold. However, in such cases, it is recommended to measure the performance of the model by comparing it with the "All" and "None" scenarios.[16-18] Patients in the external data set have received tenecteplase, which might affect the rates of DR; however, presently no data exist on the association between tenecteplase and DR. Full performance of the model should be further tested by upcoming RCTs that are investigating the effect of adjuvant reperfusion strategies among patients with incomplete reperfusion, as these patients would most likely be the target population of the model (TECNO, NCT05499832; ALLY, NCT05172934; INSIST-TNK; NCT04201964).

The externally validated model had adequate predictive accuracy and discrimination. Depending on the acceptable threshold probability, the model reasonably accurately predicts persistent incomplete reperfusion and, therefore, may advise physicians on the decision to perform additional rescue maneuvers once incomplete reperfusion is encountered.

Study Funding

Disclosure

Publication History

Appendix Authors

Name	Location	Contribution
Adnan Mujanovic, MD	Department of Diagnostic and Interventional Neuroradiology, University Hospital Bern, Inselspital, and Graduate School for Health Sciences, University of Bern, Switzerland	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; study concept or design; analysis or interpretation of data
Felix C. Ng, MBBS, BMedSc, DCH	Department of Medicine and Neurology, Melbourne Brain Centre at the Royal Melbourne Hospital, University of Melbourne; Department of Neurology, Austin Health, Heidelberg, Australia	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; study concept or design
Mattia Branca, PhD	CTU Bern, University of Bern, Switzerland	Study concept or design; analysis or interpretation of data
Hannes A. Deutschmann, MD	Division of Neuroradiology, Vascular and Interventional Radiology, Department of Radiology, Medical University of Graz, Austria	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data
Thomas R. Meinel, MD, PhD	Department of Neurology, University Hospital Bern, Inselspital, University of Bern, Switzerland	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data
Leonid Churilov, PhD	Melbourne Medical School, The Royal Melbourne Hospital, University of Melbourne, Parkville, Australia	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data
Oliver Nistl, MD	Division of Neuroradiology, Vascular and Interventional Radiology, Department of Radiology, Medical University of Graz, Austria	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data
Peter J. Mitchell, MBBS, MMed	Department of Radiology, The Royal Melbourne Hospital, University of Melbourne, Parkville, Australia	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data
Nawaf Yassi, MD, PhD	Department of Medicine and Neurology, Melbourne Brain Centre at the Royal Melbourne Hospital, University of Melbourne; Population Health and Immunity Division, The Walter and Eliza Hall Institute of Medical Research, Parkville, Australia	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data
Mark W. Parsons, MD, FRACP, PhD	Department of Neurology, Liverpool Hospital, University of New South Wales, Sydney, Australia	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data

Appendix *(continued)*

Name	Location	Contribution
Gagan J. Sharma, MCA, MITPM	Department of Medicine and Neurology, Melbourne Brain Centre at the Royal Melbourne Hospital, University of Melbourne, Australia	Drafting/revision of the manuscript for content, including medical writing for content; analysis or interpretation of data
Thomas Gattringer, MD, PhD	Division of Neuroradiology, Vascular and Interventional Radiology, Department of Radiology, and Department of Neurology, Medical University of Graz, Austria	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data
Marcel Arnold, MD	Department of Neurology, University Hospital Bern, Inselspital, University of Bern, Switzerland	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data
Fabiano Cavalcante, MD	Department of Radiology and Nuclear Medicine, Amsterdam University Medical Centers, University of Amsterdam, Amsterdam Neuroscience, the Netherlands	Drafting/revision of the manuscript for content, including medical writing for content; analysis or interpretation of data
Eike I. Piechowiak, MD	Department of Diagnostic and Interventional Neuroradiology, University Hospital Bern, Inselspital, University of Bern, Switzerland	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data
Timothy J. Kleinig, MD, PhD	Department of Neurology, Royal Adelaide Hospital, Australia	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data
David J. Seiffge, MD	Department of Neurology, University Hospital Bern, Inselspital, University of Bern, Switzerland	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data
Tomas Dobrocky, MD	Department of Diagnostic and Interventional Neuroradiology, University Hospital Bern, Inselspital, University of Bern, Switzerland	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data
Jan Gralla, MD, MSc	Department of Diagnostic and Interventional Neuroradiology, University Hospital Bern, Inselspital, University of Bern, Switzerland	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data
Urs Fischer, MD, MSc	Department of Neurology, University Hospital Bern, Inselspital, University of Bern; Department of Neurology, University Hospital Basel, University of Basel, Switzerland	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data

Appendix *(continued)*

Name	Location	Contribution
Markus Kneihsl, MD, PhD	Division of Neuroradiology, Vascular and Interventional Radiology, Department of Radiology, and Department of Neurology, Medical University of Graz, Austria	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; study concept or design
Bruce C.V. Campbell, MD	Department of Medicine and Neurology, Melbourne Brain Centre at the Royal Melbourne Hospital, University of Melbourne, Australia	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; study concept or design; analysis or interpretation of data
Johannes Kaesmacher, MD, PhD	Department of Diagnostic and Interventional Neuroradiology, University Hospital Bern, Inselspital, University of Bern, Switzerland	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; study concept or design; analysis or interpretation of data

References

1. Turc G, Bhogal P, Fischer U, et al. European Stroke Organisation (ESO) – European Society for Minimally Invasive Neurological Therapy (ESMINT) guidelines on mechanical thrombectomy in acute ischaemic stroke endorsed by stroke alliance for Europe (SAFE). *Eur Stroke J.* 2019;4(1):6-12. doi:10.1177/2396987319832140
2. Powers WJ, Rabinstein AA, Ackerson T, et al. Guidelines for the early management of patients with acute ischemic stroke: 2019 update to the 2018 guidelines for the early management of acute ischemic stroke a guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke.* 2019;50(12):e344–e418. doi:10.1161/STR.0000000000000211
3. Kaesmacher J, Ospel JM, Meinel TR, et al. Thrombolysis in cerebral infarction 2b reperfusions: to treat or to stop? *Stroke.* 2020;51(11):3461-3471. doi:10.1161/STROKEAHA.120.030157
4. Mujanovic A, Jungi N, Kurmann CC, et al. Importance of delayed reperfusions in patients with incomplete thrombectomy. *Stroke.* 2022;53(11):3350-3358. doi:10.1161/STROKEAHA.122.040063
5. Cimflova P, Singh N, Kappelhof M, et al. Effect of incomplete reperfusion patterns on clinical outcome: insights from the ESCAPE-NA1 trial. *J Neurointerv Surg.* 2023;9:jnis-2023-020553. doi:10.1136/jnis-2023-020553
6. Mujanovic A, Kurmann CC, Serrallach BL, et al. Intra-arterial thrombolysis is associated with delayed reperfusion of remaining vessel occlusions following incomplete thrombectomy. *AJNR Am J Neuroradiol.* 2023;44(9):1050-1056. doi:10.3174/ajnr.A7943
7. Mujanovic A, Brigger R, Kurmann CC, et al. Prediction of delayed reperfusion in patients with incomplete reperfusion following thrombectomy. *Eur Stroke J.* 2023;8(2):456-466. doi:10.1177/23969873231164274
8. Ramspek CL, Jager KJ, Dekker FW, Zoccali C, Van DIepen M. External validation of prognostic models: what, why, how, when and where? *Clin Kidney J.* 2021;14(1):49-58. doi:10.1093/ckj/sfaa188
9. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J.* 2014;35(29):1925-1931. doi:10.1093/eurheartj/ehu207
10. Steyerberg EW, Harrell FE. Prediction models need appropriate internal, internal-external, and external validation. *J Clin Epidemiol.* 2016;69:245-247. doi:10.1016/j.jclinepi.2015.04.005
11. Debray TPA, Collins GS, Riley RD, et al. Transparent reporting of multivariable prediction models developed or validated using clustered data: TRIPOD-Cluster checklist. *BMJ.* 2023;380:e071018. doi:10.1136/bmj-2022-071018
12. Campbell BCV, Mitchell PJ, Kleinig TJ, et al. Endovascular therapy for ischemic stroke with perfusion-imaging selection. *N Engl J Med.* 2015;372(11):1009-1018. doi:10.1056/nejmoa1414792
13. Campbell BCV, Mitchell PJ, Churilov L, et al. Tenecteplase versus alteplase before thrombectomy for ischemic stroke. *N Engl J Med.* 2018;378(17):1573-1582. doi:10.1056/nejmoa1716405
14. Campbell BCV, Mitchell PJ, Churilov L, et al. Effect of intravenous tenecteplase dose on cerebral reperfusion before thrombectomy in patients with large vessel occlusion ischemic stroke: the EXTEND-IA TNK Part 2 randomized clinical trial. *JAMA.* 2020;323(13):1257-1265. doi:10.1001/jama.2020.1511
15. Kneihsl M, Hinteregger N, Nistl O, et al. Post-reperfusion hyperperfusion after endovascular stroke treatment: a prospective comparative study of TCD versus MRI. *J Neurointerv Surg.* 2023;15(10):983-988. doi:10.1136/jnis-2022-019213

16. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making*. 2006;26(6):565-574. doi:10.1177/0272989x06295361

17. Vickers AJ, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. *BMJ*. 2016;352:i6. doi:10.1136/bmj.i6

18. Vickers AJ, van Calster B, Steyerberg EW. A simple, step-by-step guide to interpreting decision curve analysis. *Diagn Progn Res*. 2019;3(1):18-8. doi:10.1186/s41512-019-0064-7

19. Kent DM, Steyerberg E, Van Klaveren D. Personalized evidence based medicine: predictive approaches to heterogeneous treatment effects. *BMJ*. 2018;363:k4245. doi:10.1136/bmj.k4245

20. Pavlou M, Qu C, Omar RZ, et al. Estimation of required sample size for external validation of risk models for binary outcomes. *Stat Methods Med Res*. 2021;30(10):2187-2206. doi:10.1177/09622802211007522

21. Mujanovic A, Kammer C, Kurmann CC, et al. Association of intravenous thrombolysis with delayed reperfusion after incomplete mechanical thrombectomy. *Clin Neuroradiol*. 2023;33(1):87-98. doi:10.1007/s00062-022-01186-7

22. Dargazanli C, Fahed R, Blanc R, et al. Modified Thrombolysis in Cerebral Infarction 2C/Thrombolysis in Cerebral Infarction 3 reperfusion should be the aim of mechanical thrombectomy: insights from the ASTER trial (contact aspiration versus stent retriever for successful revascularization). *Stroke*. 2018;49(5):1189-1196. doi:10.1161/STROKEAHA.118.020700

23. Goyal M, Fargen KM, Turk AS, et al. 2C or not 2C: defining an improved revascularization grading scale and the need for standardization of angiography outcomes in stroke trials. *J Neurointerv Surg*. 2014;6(2):83-86. doi:10.1136/neurintsurg-2013-010665

24. Collins GS, De Groot JA, Dutton S, et al. External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. *BMC Med Res Methodol*. 2014;14(1):40-11. doi:10.1186/1471-2288-14-40

25. Hameed M, Yeung J, Boone D, Mallett S, Halligan S. Meta-research: how many diagnostic or prognostic models published in radiological journals are evaluated externally? *Eur Radiol*. 2024;34(4):2524-2533. doi:10.1007/s00330-023-10168-3

26. Yaghi S, Siegler JE, Nguyen TN. Pitfalls of randomized controlled trials in stroke: how can we do better? *Stroke Vasc Interv Neurol*. 2023;3(4):e000807. doi:10.1161/svin.123.000807

27. Nguyen TN, Raymond J, Nogueira RG, Fischer U, Siegler JE. The problem of restrictive thrombectomy trial eligibility criteria. *Stroke*. 2022;53(9):2988-2990. doi:10.1161/STROKEAHA.122.040006

28. Fröhlich H, Balling R, Beerenwinkel N, et al. From hype to reality: data science enabling personalized medicine. *BMC Med*. 2018;16(1):150. doi:10.1186/s12916-018-1122-7

29. Kappelhof M, Ospel J, Kashani N, et al. Influence of intravenous alteplase on endovascular treatment decision-making in acute ischemic stroke due to primary medium-vessel occlusion: a case-based survey study. *J Neurointerv Surg*. 2022;14(5):439-443. doi:10.1136/neurintsurg-2021-017471