



Prospective external validation of the automated PIPRA multivariable prediction model for postoperative delirium on real-world data from a consecutive cohort of non-cardiac surgery inpatients

Kelly A Reeve,^{1,2} Nayeli Schmutz Gelsomino,^{1,3} Michela Venturini,¹ Felix Buddeberg,⁴ Martin Zozman ,⁵ Reto Stocker,⁴ Mary-Anne Kedda,¹ Philipp Meier,⁴ Marius Möller,⁴ Simone Pascale Wildhaber,⁴ Benjamin T Dodsworth ¹

To cite: Reeve KA, Schmutz Gelsomino N, Venturini M, *et al.* Prospective external validation of the automated PIPRA multivariable prediction model for postoperative delirium on real-world data from a consecutive cohort of non-cardiac surgery inpatients. *BMJ Health Care Inform* 2025;**32**:e101291. doi:10.1136/bmjhci-2024-101291

► Additional supplemental material is published online only. To view, please visit the journal online (<https://doi.org/10.1136/bmjhci-2024-101291>).

Received 05 September 2024
Accepted 26 March 2025



© Author(s) (or their employer(s)) 2025. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ Group.

For numbered affiliations see end of article.

Correspondence to

Dr Benjamin T Dodsworth;
ben@pipra.ch

ABSTRACT

Objectives Postoperative delirium (POD) is a common complication in surgical patients over 60, increasing morbidity, mortality and hospital stays. While international guidelines recommend risk screening, resource constraints limit implementation. This study externally validated the Pre-Interventional Preventive Risk Assessment (PIPRA) algorithm, a CE-certified tool for identifying high-risk patients to enable targeted prevention.

Methods A prospective validation study was conducted at a 335-bed Swiss hospital as part of a quality improvement initiative. Data from 866 patients aged ≥60 undergoing non-cardiac, non-intracranial surgery (May–June 2023) were analysed. The PIPRA model's performance was assessed on discrimination (Area Under the Receiver Operating Characteristic Curve (AUROC)) and calibration.

Results POD occurred in 11.5% (n=100) of patients. The PIPRA model showed good discrimination (AUROC=0.77, 95% CI: 0.72 to 0.82) and generally accurate calibration, though slightly overpredicting risk in high-risk patients. POD was associated with higher mortality, prolonged intensive care unit (ICU)/hospital stays and increased nursing care needs. The model effectively stratified patients for targeted interventions.

Discussion The PIPRA algorithm demonstrated robust performance in a real-world setting, affirming its utility for POD risk prediction. The study highlighted the model's applicability across diverse clinical environments, despite differences in patient populations and screening protocols.

Conclusions The PIPRA algorithm is a reliable tool for identifying surgical patients at risk of POD, supporting early intervention strategies to improve patient outcomes. Its integration into clinical workflows may enhance POD prevention efforts and optimise resource allocation in perioperative care.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Postoperative delirium (POD) is a prevalent and serious complication among older surgical patients, associated with increased morbidity, mortality and healthcare costs. While early risk identification is recommended, routine screening remains challenging due to resource constraints. The Pre-Interventional Preventive Risk Assessment (PIPRA) algorithm was developed to address this gap and validated on clinical trial data, but its real-world clinical performance required further validation.

WHAT THIS STUDY ADDS

⇒ This study confirms that the PIPRA algorithm effectively predicts POD risk in a real-world setting, demonstrating strong discrimination (AUROC=0.77) and generally accurate calibration. It also highlights the association between POD and worse clinical outcomes, reinforcing the need for targeted preventive strategies.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ The findings support the integration of PIPRA into clinical workflows to enhance POD prevention. Its use could enable more efficient resource allocation, improve patient outcomes and inform future policy decisions on POD risk assessment in perioperative care.

INTRODUCTION

Postoperative delirium (POD) is the most common complication affecting surgical patients over the age of 60.¹ POD is characterised by altered consciousness, agitation, confusion and cognitive decline, and long-term adverse outcomes including increased

postoperative morbidity and mortality, functional and cognitive decline, prolonged length of stay, higher readmission rates and postoperative neurocognitive disorders.^{2–5}

International guidelines and medical societies recommend screening for POD risk prior to surgery so that preventive measures can be introduced for patients at risk.^{6–11} Non-pharmacological multicomponent interventions^{12 13} and programmes such as the Hospital Elder Life Program¹⁴ have been shown to reduce delirium incidence;^{15–17} however, these measures are not routinely applied by all healthcare professionals, often due to limited awareness or resources.

We have developed a preoperative POD risk prediction algorithm called PIPRA (Pre-Interventional Preventive Risk Assessment),¹⁸ following a comprehensive individual patient data meta-analysis (IPDMA).¹⁹ PIPRA is an automated, CE-certified POD risk prediction tool designed to identify at-risk patients over 60 years of age, thereby enabling clinicians to implement targeted preventive strategies for those patients at highest risk.

This study describes the first external validation of PIPRA in a real-world clinical setting.

METHODS

Source of data and participants

The delirium risk calculated using the PIPRA model was routinely collected, together with the delirium screening results, during a quality improvement project (QIP) in a Swiss, private 335-bed hospital.²⁰ The project included all inpatients aged 60 and above undergoing non-cardiac, non-intracranial surgery with an admission date from 1 May to 30 June 2023. In 2023/2024, the hospital had a mix of 43.0% public and 57.0% private or partially privately insured patients, 79.4% of patients were outpatients and, of the inpatients, 82.4% were elective. The QIP aimed to increase delirium screening, treatment and prevention.

Comparison of validation to original development data

The development data originated from an IPDMA of clinical studies^{18 19 21} performed in several countries, whereas the validation data for this study consisted of real-world data collected as part of a QIP in a single private hospital in Switzerland. The eligibility criteria in this QIP were generally broader than those in the studies captured by the IPDMA. The delirium outcome was also measured heterogeneously in the IPDMA data, with only one study (contributing less than 10% of patients) that used the Delirium Observation Screening Scale (DOSS) as a diagnosis tool, although in combination with Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) criteria. In the QIP presented here, the DOSS was used as the primary diagnostic tool, in combination with any routine diagnosis performed by the physician according to DSM-V criteria. In the development dataset, the predictors were collected specifically to investigate predictors of

POD, while here they were collected during the routine preanaesthesia consultation.

Outcome

The outcome predicted by the model was POD, which was defined as delirium occurring up to 7 days after surgery. In this dataset, a patient was deemed delirious if they had a score of 3 points or more for the DOSS or had the ICD-10 diagnosis code for delirium for their hospital stay. The DOSS is a 13-point checklist for the identification of delirium.²² The data were collected in routine clinical practice, with delirium risk present in the electronic health record of the patient and viewable by the nurses performing the DOSS. The nurses were not informed about the validation project and were instructed to perform the DOSS on all patients as part of the QIP. Compliance with DOSS screening was assessed by comparing the observed number of DOSS screenings to the expected number. We anticipated one DOSS screening per patient per shift, with three shifts per day. Thus, compliance was calculated as a percentage:

$$\text{Compliance} = \frac{\text{number of DOS screenings}}{\text{total LOS (days)} \times 3}$$

Predictors

All PIPRA variables (age, body mass index, American Society of Anesthesiologists Physical Status Classification System (ASA) score, number of prescribed medications, cognitive impairment, history of delirium, surgery risk, laparotomy/thoracotomy, optional preoperative C reactive protein (CRP) value) except CRP were routinely collected. CRP is optional and therefore was only used where available.

Predictors were not collected specifically for the delirium risk. Instead, they were all part of clinical routine and were collected by the anaesthesiologist during the preanaesthesia consultation.

In addition to the main predictors, we recorded patient sex and the Self-Care Index (SPI) for further characterisation. The SPI, assessed daily, measures self-care abilities (eg, hygiene, mobility, elimination), producing a total score from 10 (maximum impairment) to 40 (full self-care).²³

Sample size

A sample size was not calculated since it was a QIP. However, with a POD incidence of 10% and a targeted area under the curve (AUC) of 0.75, a 95% CI for the AUC would have a width of 0.12 with the 866 subjects enrolled in the study. We consider this to be precise enough to provide meaningful information about model performance.

Missing data

Missing data were only imputed for PIPRA predictors. Following the use proposed in the development paper, mean/mode imputation was used. When CRP was not available, the PIPRA submodel without CRP was used for

prediction rather than imputing CRP, as in the development paper.

Statistical analysis methods

Subject and procedure characteristics were summarised by median (IQR) or frequency (percentage), dependent on data type. Differences in clinical parameters across groups (eg, POD/no POD) were explored using t-tests or χ^2 tests, accordingly. All analyses were performed using R Statistical Software (V.4.3.2).²⁴

This external validation was performed using data completely independent of the model development. To compute the predicted risk on the new data, the following equation was applied: $p = 1 / (1 + \exp(-lp))$, where p is the predicted risk and lp is the linear combination of the individual predictor variables multiplied by the log odds coefficients (including the intercept).

The participant and procedure characteristics were summarised and compared across the development and validation studies using the same methods as for comparing across POD. They were also visualised using violin plots or bar charts.

Both model discrimination (AUC) and calibration (calibration-in-the-large, calibration slope, calibration plot) were assessed on the validation data.

Risk groups

Patients were stratified into low (PIPRA <10%), intermediate (PIPRA from 10% to 19.9%), high (PIPRA from 20% to 34.9%) and very high (PIPRA >35%) risk groups as per the original publication.¹⁸ Patients with high risk or very high risk have more chance of developing delirium than the average elderly population. For this population, it is essential that some preventive perioperative measures are taken. An exploratory analysis was added, in which a stratified table was created based on risk group. The number of delirious patients was identified, and the number of potentially preventable cases was estimated, assuming prevention effectiveness rates reported in the literature.¹⁷

RESULTS

Participants

5279 patients were admitted during the timeframe of the validation study. Excluding patients under 60 years of age, outpatients and non-surgical or cardiac surgery patients, 866 patients remained (figure 1). The characteristics of the patients showed worse outcomes for delirium patients, including on average sixfold more time spent in the intensive care unit (ICU), twice the length of stay and threefold more nursing time (table 1). Delirium patients had, on average, a 10-fold higher mortality; however, only seven patients died in this cohort. Patients experiencing POD were generally older, with more cognitive impairment, a greater history of delirium, less able to take care of themselves (lower SPI), with higher number of prescribed medications, higher CRP values and higher ASA status (table 1).

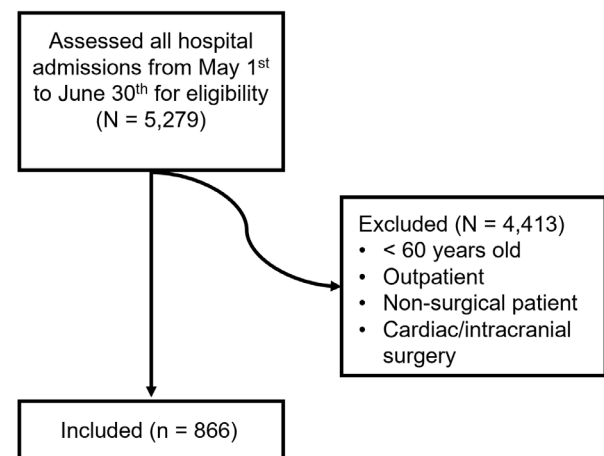


Figure 1 Patient flow through the study. 5279 patients were admitted during the time frame of the QIP. Excluding patients under 60 years of age, outpatients and non-surgical or cardiac surgery patients, 866 patients remained. QIP, quality improvement project.

While the participants in the development study and this validation did not display clinically meaningful differences in age, BMI or history of delirium, they did differ in other ways (figure 2, online supplemental table S1). In the validation, fewer patients underwent higher risk or open procedures, and the patients themselves were more frail (higher ASA). All patients in the validation study were on at least one medication.

The mean compliance to delirium screening was 60.6%. 79 patients (9%) received no screening, and these patients were younger, with lower ASA, smaller surgeries and lower length of stay (online supplemental table S2). Anecdotal feedback suggested the night shift had the lowest compliance, though we were not able to retrieve this level of data due to ethical considerations.

Overall, data were mostly missing for CRP (table 1). However, this was anticipated in the model design, and the submodel without CRP was used for these patients. Missingness was also relatively high for cognitive impairment, history of delirium and number of medications (table 1).

Model performance

The validation analysis included 866 eligible subjects observed from May 2023 to the end of the study. Of these, 100 subjects (11.5%, 95% CI 9.6% to 13.8%) were identified as having POD. According to the risk groups created at development, 59.2% of subjects were considered low risk, 21.6% intermediate risk, 10.4% high risk and 8.8% very high risk (figure 3a).

At external validation, the PIPRA model was found to discriminate well between those with and without POD (AUC=0.77 (95% CI 0.72 to 0.82)). There were no major violations to calibration, although there was a tendency to slight overprediction for higher risk patients (figure 3b,c). The diagnostic accuracy was dependent on the chosen threshold and is displayed in detail in table 2.

Table 1 Description of included subjects and outcomes

	Overall	No POD	POD	P value	Missing
Number of patients	866	766	100		
Age in years (IQR)	72 (66, 78)	71 (66, 77)	79.00 (73, 83)	<0.001	0.0
Female (%)	399 (46.1)	358 (46.7)	41 (41.0)	0.329	0.0
BMI (IQR)	25.14 (22.73, 28.01)	25.24 (22.79, 28.09)	24.30 (21.51, 27.66)	0.078	5.0
Cognitive impairment (%)	24 (3.7)	9 (1.6)	15 (18.8)	<0.001	25.2
History of delirium (%)	24 (3.7)	14 (2.4)	10 (12.7)	<0.001	24.5
ASA status (%)				<0.001	10.7
1	36 (4.7)	36 (5.3)	0 (0.0)		
2	395 (51.1)	365 (53.5)	30 (33.0)		
3	324 (41.9)	270 (39.6)	54 (59.3)		
4	18 (2.3)	11 (1.6)	7 (7.7)		
SPI (IQR)	38.00 (33.00, 40.00)	39.00 (34.00, 40.00)	36.00 (30.75, 40.00)	0.003	5.0
Number of medications (IQR)	4 (2, 7)	4 (2, 6)	5 (4, 8)	<0.001	21.2
Log (CRP)* (IQR)	0.97 (−0.69, 2.81)	0.67 (−0.69, 2.59)	2.16 (1.12, 3.45)	0.001	72.5
Surgical risk† (%)				0.011	0.1
1	253 (29.2)	234 (30.6)	19 (19.0)		
2	565 (65.3)	494 (64.6)	71 (71.0)		
3	47 (5.4)	37 (4.8)	10 (10.0)		
Laparotomy/thoracotomy (%)	51 (6.2)	42 (5.8)	9 (9.6)	0.233	5.7
Outcomes					
Death (%)	7 (0.8)	3 (0.4)	4 (4.0)	0.001	0.0
Length of stay in days:					
Mean (SD)	4.84 (4.72)	4.21 (3.80)	9.69 (7.47)	<0.001	0.0
Median (IQR)	3.15 (2.01, 6.00)	3.03 (1.96, 5.24)	7.55 (4.87, 11.90)	<0.001	0.0
Nursing time in hours					
Mean (SD)	23.58 (31.48)	19.10 (20.05)	57.84 (64.84)	<0.001	0.1
Median (IQR)	14.73 (8.73, 25.35)	13.53 (8.18, 21.82)	33.75 (21.63, 71.83)	<0.001	0.1
ICU admission (%)	71 (8.20)	47 (6.14)	24 (24.00)	<0.001	0.0
ICU length of stay in hours					
Mean (SD)	3.61 (22.51)	2.29 (19.69)	13.72 (36.25)	<0.001	0.0
Median (IQR)	0.00 (0.00, 0.00)	0.00 (0.00, 0.00)	0.00 (0.00, 0.00)	<0.001	0.0

Description of included subjects and outcomes. Data shown are n (%-percentage of total) or median (IQR) or average (SD). The length of stay, in particular the ICU length of stay, can be considered both outcome and precipitating factor.

*Preoperative.

†Cardiac risk for non-cardiac surgery.

ASA, American Society of Anesthesiologists Physical Status Classification; BMI, body mass index; CRP, C reactive protein; ICU, intensive care unit; POD, postoperative delirium; SPI, Selbstpflegeindex = self-care index.

For example, at the threshold of 10% or ‘medium’ risk, the sensitivity was 74% and specificity 64%.

Potential effect

An exploratory analysis was performed to estimate the preventative potential. This analysis suggested that three-quarters of delirium cases could be prevented, under the assumption of a constant prevention effect from the literature, and when focusing exclusively on patients classified as medium risk or higher (online supplemental table S3).

DISCUSSION

The delirium incidence at 11.5% was slightly lower compared with the development dataset (19.7%) and that reported in the literature. However, since the average delirium risk at 13% is similar to the observed incidence, it suggests that the population differences are accounted for in the PIPRA model. These differences might have arisen as the real-world data were collected at a private

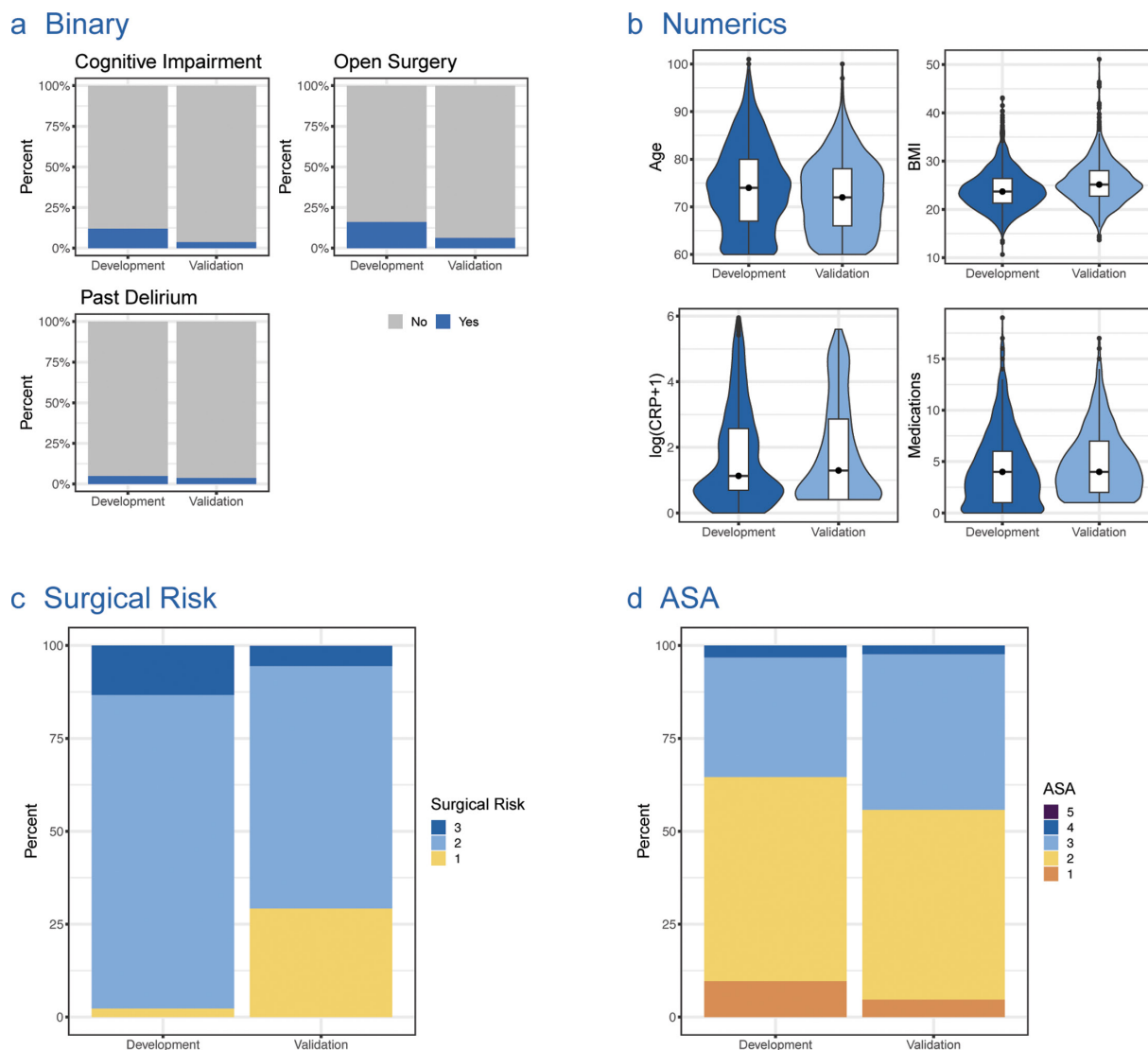


Figure 2 Comparison of predictor variable distributions between the development and validation datasets. BMI, body mass index; CRP, C reactive protein; ASA, American Society of Anesthesiologists Physical Status Classification.

hospital, while the development data originated from university and public hospitals.

Limitations

The QIP was not a clinical trial and, therefore, there were no sample size calculations, and there were strict limitations on exploring the data in depth due to ethical considerations.

While adherence to delirium screening protocols meets acceptable standards for clinical practice (91% of patients were screened), higher compliance is anticipated in research settings, with only 61% of the total expected screenings performed. The DOSS is simple and does not require direct patient interaction; it has limitations in both sensitivity and specificity.²⁵ To mitigate this issue, we provided nurses with training to accurately identify and diagnose delirium, stressing that the DOSS is a mandatory supplementary tool. Additionally, we emphasised in the training that the tool is prone to missing hypoactive delirium.

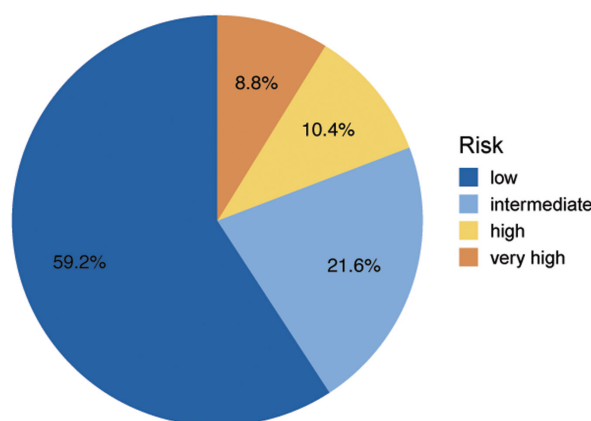
The risk prediction model was originally developed on data collected by IPDMA. In the development data, delirium was predominantly assessed by CAM. Only one study used the DOSS; however, it was used in conjunction with DSM-IV criteria. Differences in setting, outcome measurements and patient characteristics between the development and validation data could have caused the model to underperform; however, the model performed well, and this further confirms that the risk prediction model is robust.

Since the validation was performed in normal clinical practice, there was no blinding of the nurses to the prediction. While they were not involved or informed of the validation project, it is still possible that this introduced bias.

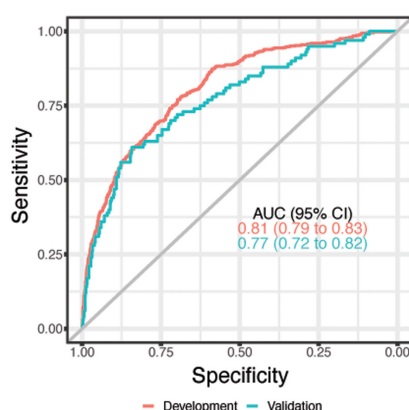
Strengths

Despite the limitations discussed above, we see the model performing well on real-world data, showing the model to be robust. This was expected, since the original model was built on a diverse dataset of IPDMA data. To the best

a Risk



b ROC



c Calibration

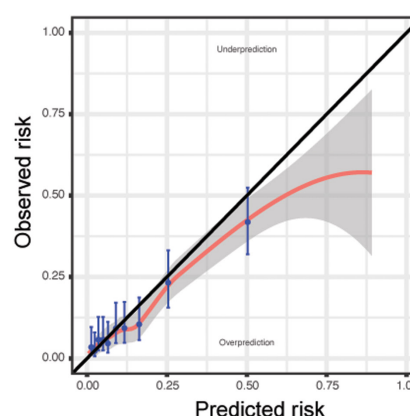


Figure 3 Results and performance of the PIPRA model. (A) The PIPRA model stratifies most patients as low risk. (B) The AUC of the PIPRA model is 0.77 (95% CI 0.72 to 0.82). (C) The model shows good calibration, with a slight overprediction for the higher risk patients. AUC, area under the curve; ROC, Receiver operating characteristic; PIPRA, Pre-Interventional Preventive Risk Assessment.

of our knowledge, this is the only POD risk prediction model that has been built on IPDMA data and so far the only POD risk prediction model that is CE-marked and approved for clinical use.

A key strength of using real-world data is the broader demographic representation for subpopulations often under-represented in trials. All patients were included in

the study, and there was no bias introduced through a consenting process, where often those patients who are most vulnerable are lost (eg., those with low education or mild cognitive impairment—known risk factors for delirium).¹⁹ However, more than half the patients were privately insured, which is not representative of most hospitals.

Table 2 Diagnostic accuracy

Measure	Medium risk	High risk	Very high risk
Sensitivity	0.74 (0.64, 0.82)	0.56 (0.46, 0.66)	0.34 (0.25, 0.44)
Specificity	0.64 (0.6, 0.67)	0.86 (0.83, 0.88)	0.95 (0.93, 0.96)
PPV	0.21 (0.17, 0.26)	0.34 (0.27, 0.41)	0.45 (0.33, 0.57)
NPV	0.95 (0.93, 0.97)	0.94 (0.92, 0.95)	0.92 (0.89, 0.93)
LR+	2.03 (1.75, 2.36)	3.9 (3.05, 4.98)	6.2 (4.15, 9.26)
LR–	0.41 (0.29, 0.57)	0.51 (0.41, 0.64)	0.7 (0.61, 0.8)

Diagnostic accuracy using three different risk thresholds: medium (>10%), high (>20%) and very high (PIPRA>35%). Values in brackets are the 95% CIs.

LR, likelihood ratio; NPV, negative predictive value; PIPRA, Pre-Interventional Preventive Risk Assessment; PPV, positive predictive value.

Interpretation

The AUC from this external validation was excellent in comparison to other external validations for POD prediction. In the original study, a pilot external validation on 359 patients showed an AUC of 0.74, which is close to the 0.77 observed here. A project undertaken by Wong *et al*, using a dataset of 292 patients to test various delirium prediction models head-to-head revealed C-indices ranged from 0.52 to 0.74, where even the highest scoring algorithm was below the AUC of 0.77 observed here.²⁶ Strikingly, in that review, the highest self-reported AUC from the assessed models was 0.94;²⁷ however, this depreciated to 0.61 in the external validation. This illustrates the importance of external validation on new patient cohorts from different settings. It also compares favourably to a risk prediction model using only age, in both discrimination and calibration (online supplemental figure S1).

The use of real-world data has opposing strengths and limitations to the use of data from clinical trials. There was a preliminary external validation in the original study,¹⁸ and a clinical trial to validate the algorithm further is underway, which will complement this study well.

Implications

Identification of patients at risk of delirium is essential for effective, targeted prevention strategies and early treatment of delirium. The PIPRA tool was developed as a consensus tool, using data from many authors and a number of large studies through a collaborative IPDMA, and this real-world validation study confirms that we have a robust and reliable POD risk prediction algorithm.

Author affiliations

¹PIPRA AG, Zurich, Switzerland

²Zurich University of Applied Sciences, Zurich, Switzerland

³Department of Clinical Research, University of Basel, Basel, Switzerland

⁴Hirslanden Klinik Hirslanden, Zurich, Switzerland

⁵Kusnacht Practice AG, Zurich Zollikon, Switzerland

X Benjamin T Dodsworth @btdodsworth

Acknowledgements We would like to thank the nurses who performed the delirium screening and the delirium team for their efforts in implementing the delirium screening and supporting the staff. We would also like to thank Fabian Gautschi and Victoria Engler for supporting the data acquisition.

Contributors Conceptualisation (NSG, FB, MZ, RS, PM and BT), data acquisition and data collection (FB, MZ, MM and SPW), methodology and formal analysis (KAR, MV and BT), project administration (PM and BT), supervision and funding (RS and BT), writing—original draft (KAR, NSG, MAK and BT), writing—review, editing and approval of the final version (KAR, NSG, MV, FB, MZ, RS, MAK, PM, MM, SPW and BT). BT is the guarantor.

Funding This study was funded by PIPRA AG, Stiftung Quality of Life Switzerland and EIT Health. EIT Health is supported by the EIT, a body of the European Union.

Competing interests KAR is a shareholder of PIPRA AG. BT and NSG are founders, shareholders and employees of PIPRA AG. MAK is a shareholder and employee of PIPRA AG. The remaining authors have no competing interests to declare.

Patient consent for publication Not applicable.

Ethics approval The quality improvement project received a waiver Req-2023-00307 from the Zurich Cantonal Ethics committee as it uses anonymous data.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement No data are available. Research data are not shared, but requests for collaboration are encouraged.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Martin Zozman <http://orcid.org/0000-0002-6096-0973>

Benjamin T Dodsworth <http://orcid.org/0000-0001-7908-7855>

REFERENCES

- Inouye SK. Delirium in older persons. *N Engl J Med* 2006;354:1157–65.
- Gleason LJ, Schmitt EM, Kosar CM, *et al*. Effect of Delirium and Other Major Complications on Outcomes After Elective Surgery in Older Adults. *JAMA Surg* 2015;150:1134–40.
- Goldberg TE, Chen C, Wang Y, *et al*. Association of Delirium With Long-term Cognitive Decline: A Meta-analysis. *JAMA Neurol* 2020;77:1373–81.
- Kunicki ZJ, Ngo LH, Marcantonio ER, *et al*. Six-Year Cognitive Trajectory in Older Adults Following Major Surgery and Delirium. *JAMA Intern Med* 2023;183:442–50.
- Raats JW, van Eijdsen WA, Crolla RMPH, *et al*. Risk Factors and Outcomes for Postoperative Delirium after Major Surgery in Elderly Patients. *PLoS One* 2015;10:e0136071.
- Aldecoa C, Bettelli G, Bilotta F, *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;41:81–108.
- Chow WB, Rosenthal RA, Merkow RP, *et al*. Optimal preoperative assessment of the geriatric surgical patient: a best practices guideline from the American College of Surgeons National Surgical Quality Improvement Program and the American Geriatrics Society. *J Am Coll Surg* 2012;215:453–66.
- American Society of Anesthesiologists (ASA). Perioperative brain health initiative 2023. 2024. Available: <https://www.asahq.org/brainhealthinitiative>
- Berger M, Schenning KJ, Brown CH 4th, *et al*. Best Practices for Postoperative Brain Health: Recommendations From the Fifth International Perioperative Neurotoxicity Working Group. *Anesth Analg* 2018;127:1406–13.
- American Geriatrics Society Expert Panel on Postoperative Delirium in Older Adults. American Geriatrics Society abstracted clinical practice guideline for postoperative delirium in older adults. *J Am Geriatr Soc* 2015;63:142–50.
- National Institute for Health and Care Excellence (NICE). Evidence standards framework for digital health technologies. 2019.
- Hsieh TT, Yue J, Oh E, *et al*. Effectiveness of multicomponent nonpharmacological delirium interventions: a meta-analysis. *JAMA Intern Med* 2015;175:512–20.
- Godfrey M, Green J, Smith J, *et al*. Process of implementing and delivering the Prevention of Delirium system of care: a mixed method preliminary study. *BMC Geriatr* 2019;20:1.
- Hsieh TT, Yang T, Gartaganis SL, *et al*. Hospital Elder Life Program: Systematic Review and Meta-analysis of Effectiveness. *Am J Geriatr Psychiatry* 2018;26:1015–33.
- Zhao Q, Liu S, Zhao H, *et al*. Non-pharmacological interventions to prevent and treat delirium in older people: An overview of systematic reviews. *Int J Nurs Stud* 2023;148:104584.
- Janssen TL, Alberts AR, Hooft L, *et al*. Prevention of postoperative delirium in elderly patients planned for elective surgery: systematic review and meta-analysis. *Clin Interv Aging* 2019;14:1095–117.

- 17 Burton JK, Craig LE, Yong SQ, *et al.* Non-pharmacological interventions for preventing delirium in hospitalised non-ICU patients. *Cochrane Database Syst Rev* 2021;7:CD013307.
- 18 Dodsworth BT, Reeve K, Falco L, *et al.* Development and validation of an international preoperative risk assessment model for postoperative delirium. *Age Ageing* 2023;52:1–10.
- 19 Sadeghirad B, Dodsworth BT, Schmutz Gelsomino N, *et al.* Perioperative Factors Associated With Postoperative Delirium in Patients Undergoing Noncardiac Surgery: An Individual Patient Data Meta-Analysis. *JAMA Netw Open* 2023;6:e2337239.
- 20 Dodsworth BT, Reeve KA, Zozman M, *et al.* Benefits of an automated postoperative delirium risk prediction tool combined with non-pharmacological delirium prevention on delirium incidence and length of stay: a before-after analysis based on a quality improvement project. *Age Ageing* 2024;53:afae219.
- 21 Buchan TA, Sadeghirad B, Schmutz N, *et al.* Preoperative prognostic factors associated with postoperative delirium in older people undergoing surgery: protocol for a systematic review and individual patient data meta-analysis. *Syst Rev* 2020;9:261.
- 22 Schuurmans MJ, Shortridge-Baggett LM, Duursma SA. The Delirium Observation Screening Scale: a screening instrument for delirium. *Res Theory Nurs Pract* 2003;17:31–50.
- 23 Schlarmann JG. Der CMS© im ePA©, Verschiedene Qualitätsdimensionen eines Instruments Eine empirische Analyse. Private Universität Witten/Herdecke gGmbH, Fakultät Medizin, Institut für Pflegewissenschaft; 2007.
- 24 R Core Team. R: a language and environment for statistical computing. Vienna, Austria R Foundation for Statistical Computing; 2021. Available: <https://www.R-project.org>
- 25 Gavinski K, Carnahan R, Weckmann M. Validation of the delirium observation screening scale in a hospitalized older population. *J Hosp Med* 2016;11:494–7.
- 26 Wong CK, van Munster BC, Hatseras A, *et al.* Head-to-head comparison of 14 prediction models for postoperative delirium in elderly non-ICU patients: an external validation study. *BMJ Open* 2022;12:e054023.
- 27 Kim MY, Park UJ, Kim HT, *et al.* DELirium Prediction Based on Hospital Information (Delphi) in General Surgery Patients. *Medicine (Baltimore)* 2016;95:e3072.