Towards Data-Driven Precision Neurorehabilitation

Validation and Implementation of a Multivariable Prediction Model for Prognosis of Functional Independence in Young Adults with Acquired Brain Injury

Master of Science Thesis

E.H.M. Lie

Universiteit Leiden
The Netherlands

TU Delft

Erasmus Universiteit Rotterdam

daan theeuwes centrum
voor intensieve neurorevalidatie

This page was intentionally left blank.

Validation and Implementation of a Multivariable Prediction Model for Prognosis of Functional Independence in Young Adults with Acquired Brain Injury

EHM Lie

4651324

25 October 2024

Thesis in partial fulfilment of the requirements for the joint degree of Master of Science in

Technical Medicine

Leiden University; Delft University of Technology; Erasmus University Rotterdam

Master thesis project (TM30004; 35 ECTS)
Dept. of Biomechanical Engineering, TUDELFT
April 2024 – November 2024

Supervisors	
Prof.dr. W.C. Peul	LUMC
Dr. M. Königs	AMC
Dr. R. van der Veen	AMC

Thesis committee members	
Prof.dr. W.C. Peul (Chair)	LUMC
Dr. M. Königs	AMC
Dr.ir. A.H.A. Stienen	TU Delft

Universiteit Leiden

TU Delft Delft University of Technology

Erasmus Universiteit Rotterdam

Preface and Acknowledgements

The completion of this Master of Science thesis for the Master's degree in Technical Medicine marks the end of not only an inspiring and fascinating project but also of seven years of study. Reflecting on these years, I still believe Technical Medicine was the right choice for me. Throughout my master's program, I had the privilege of gaining valuable experiences in various areas that captured my interest. From working on augmented reality in trauma surgery to application development for plastic surgery, 3D imaging in oral and maxillofacial surgery, and automatic analysis of functional MRI in neurosurgery, each experience shaped my skills and interests. Ultimately, I ended up at the Daan Theeuwes Centre (DTC) in Woerden, combined with the neurosurgery department at Haaglanden Medical Centre (HMC), where I focused on implementing prediction models for young adults suffering from Acquired Brain Injury (ABI).

Attending both hospital and rehabilitation settings gave me insights into the full care pathway for young adults with ABI – from complex surgeries and hospital care to neurorehabilitation. These experiences deepened my understanding of the patient group and the field. Combined with an in-depth exploration of prediction modelling, this equipped me with the necessary skills and knowledge to take steps towards implementing these models in clinical care and advancing more personalised, data-driven approaches. This thesis holds a special place for me, as it relates to a dear friend who was admitted to the DTC. I have experienced the uncertainty that accompanies ABI and the challenges of predicting its outcomes and managing expectations as a close friend. Although my work may not directly benefit her now, I feel encouraged to contribute to this field, knowing it may help others in the future.

First, I would like to thank my supervisors: Dr Marsh Konigs, Prof Dr Wilco Peul, and Dr Ruud van der Veen. Marsh, I am grateful for your guidance throughout these months. I truly appreciated our weekly meetings, our collaborative working style, and how you encouraged me to think critically about my results while helping me navigate both clinical and technical challenges. Wilco, thank you for your valuable input during our meetings and for bridging the gap between the hospital and neurorehabilitation. Your open-mindedness and enthusiasm made our discussions truly energising. Ruud, I am thankful for always being available to answer my questions, for your clear explanations of the logic behind the models, and for updating the data whenever I needed it—your support was invaluable.

I also want to thank the multidisciplinary team at the DTC for their time and for giving me an inside view of neurorehabilitation. Special thanks to Shanna for our insightful discussions on implementing technology in healthcare, and to Andries and Amy for providing feedback on our application. I am grateful to everyone at the Centre who took the time to listen to my progress and share stories with me about work and beyond. For further clinical insights, I am grateful to the neurosurgeons, physician assistants, AIOS, and ANIOS at the neurosurgery department at HMC. Thank you for showing me everything—from the ICU to the wards, from interesting surgeries to your favourite lunch sandwiches in the hospital cafeteria.

Lastly, I want to thank my family and friends for all their love and support. I am especially grateful for their encouragement, which has been a constant source of motivation during my studies. I genuinely enjoyed working on this thesis and am grateful for the opportunity to contribute to this field. While this is an ending I have not particularly longed for, I am also excited about what lies ahead and look forward to the future with great anticipation.

Validation and Implementation of a Multivariable Prediction Model for Prognosis of Functional Independence in Young Adults with Acquired Brain Injury

E.H.M. Lie[a,d], M. Königs[c,d], W.C. Peul[b]

[a]*Delft University of Technology, Erasmus Medical Centre, Leiden University Medical Centre*
[b]*Department of Neurosurgery, LUMC, Leiden University, Leiden, Netherlands*
[c]*Emma Neuroscience Group, Emma Children's Hospital, Amsterdam UMC, University of Amsterdam, Netherlands*
[d]*Daan Theeuwes Center for Intensive Neurorehabilitation, Woerden, Netherlands*

Abstract

Introduction
Acquired Brain Injury (ABI) presents a significant public health challenge due to the diverse recovery trajectories resulting from its heterogeneous nature. Prediction models, derived from structured data collection, offer a more personalised approach to neurorehabilitation. However, a substantial gap remains between the development of these models and their successful implementation in clinical practice. This study addresses this gap by focusing on two key components: externally validating prediction models for functional independence in young adults with ABI, and creating a user-friendly interface to support their application in clinical settings. These efforts represent a crucial step toward advancing precision neurorehabilitation, enabling data-driven, individualised care tailored to the unique needs of ABI patients.

Design and Methods
Previously, three multivariable prediction models were developed to predict the prognosis of severe ABI in young adults aged 16 to 35 admitted to the Daan Theeuwes Centre, demonstrating promising performance. These models focused on functional independence, measured by the Barthel Index (BI) at admission, three months later, and the change in independence during this period. This study focused on the implementation of these prediction models by external validation of these models, and the development of a web-based tool to facilitate their implementation in clinical practice. Data for the external validation cohort were sourced from the Measurement Feedback System (MFS). Highly incomplete variables were excluded, and missing data were handled using Predictive Mean Matching (PMM). Model performance was assessed using Coefficient of Determination (R^2), Root Mean Square (RMSE), and Mean Absolute Error (MAE), alongside calibration and correlation analyses. Additionally, results were assessed against the 95% Prediction Interval (PI) of the development cohort. A web-based tool was developed simultaneously to facilitate the practical application of these models in clinical practice, informed by clinician feedback and literature insights.

Results
The validation cohort (n = 21) showed minimal discrepancies compared to the development cohort (n = 100), but external validation revealed reduced predictive accuracy. The "Level of Independence at Admission," "Level of Independence at Three Months Post-Admission" and "Change in Independence over Three Months" models had notable drops in R^2 from 65.7% to 42.8%, 59.3% to 29.7%, and 76.3% to 35.9%, respectively. All models fell outside the 95% Confidence Interval (CI) of R^2 for the development cohort and showed increased RMSE and MAE values. Calibration showed overestimation of lower BI scores and underestimation of higher scores, with a substantial proportion of predictions falling outside the 95% PI of the development cohort. Correlation analysis indicated that longer hospital stays and Post-Traumatic Amnesia (PTA) were linked to higher prediction errors, while higher BI scores at admission and focal injuries in Traumatic Brain Injury (TBI) were associated with lower errors. The web-based tool included a page for applying the models, one for visualising recovery trajectories via an interactive flow diagram, and another for accessing detailed model information.

Discussion
The implementation of prediction models involves several key phases, beginning with structured data collection, model development, and evaluation. Once the model demonstrates sufficient performance, it can be implemented into clinical practice. To maintain its relevance, continuous monitoring and updating are essential. In this study, we focused on two main components—external validation of the prediction models and the practical implementation through the development of a user-friendly tool. By addressing these, we have taken significant steps towards implementing prediction models into clinical practice. These components underscore the importance of structured data collection, rigorous validation, and practical application to ensure the models' effectiveness in real-world settings. By implementing prediction models, we aim to employ a data-driven approach that brings us closer to precision neurorehabilitation.

Keywords: prediction modelling, multivariable prediction model, external validation, acquired brain injury, tool development, clinical practice

List of Abbreviations

ABI	Acquired Brain Injury
AIC	Akaike Information Criterion
BH	Benjamini-Hochberg
BI	Barthel Index
CFIR	Consolidated Framework of Implementation Research
CI	Confidence Interval
CVA	Cerebrovascular Accident
DTC	Daan Theeuwes Centre
GCP	Good Clinical Practice
GDPR	General Data Protection Regulation
MAE	Mean Absolute Error
MCID	Minimal Clinically Important Difference
MFS	Measurement Feedback System
OOD	Out-of-Distribution
PI	Prediction Interval
PII	Personally Identifiable Information
PMM	Predictive Mean Matching
PALOC	Post-Acute Level of Consciousness
PTA	Post-Traumatic Amnesia
R^2	Coefficient of Determination
RFE	Recursive Feature Elimination
RMSE	Root Mean Square
SAH	Subarachnoid Haemorrhage
TBI	Traumatic Brain Injury
TRIPOD	Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis

Contents

1. Introduction

Acquired Brain Injury (ABI) includes a broad spectrum of mechanisms of brain injury, establishing it as a vastly important area in medicine and public health (1). Traumatic Brain Injury (TBI) and stroke are the two most prevalent forms of ABI (2), affecting approximately 85 million individuals annually (3; 4). ABI can result in serious consequences, including cognitive impairment, physical disability, and emotional distress, which profoundly impact quality of life (5; 6; 7; 8). Neurorehabilitation focuses on regaining motor function, daily activities, cognition, communication and psychosocial factors to increase the ability to participate in society (9).

ABI patients exhibit significant heterogeneity in terms of the impairment profile, progression during neurorehabilitation, and outcomes, driven by factors such as demographics, neuropathology aetiology, and environmental influences (1). This heterogeneity makes it challenging for healthcare professionals to assess rehabilitation potential and establish reliable prognoses (10). Consequently, the heterogeneity of ABI complicates the application of standardised treatment protocols, underscoring the need for the development of personalised treatment strategies (11).

Predictive modelling in healthcare involves developing and applying mathematical algorithms and statistical techniques to predict various outcomes relevant to patient care, such as prognosis and risk of disease (12). Prediction models in healthcare serve multiple purposes: they help refine our understanding of rehabilitation mechanisms, provide a quantitative framework for setting expectations for both patients and clinicians, and aid in optimising treatment plans for individual patients (13). While stroke rehabilitation models offer meaningful insights (14), prognostic models for post-acute rehabilitation in patients with ABI often lack the systematic robustness required for generalisation beyond the original development setting. This limitation hinders their translation to other environments, particularly highly specialised settings, highlighting the need for models that not only perform reliably across diverse healthcare settings but are also adaptable to the specific needs of specialised centres (15).

Recent literature indicates that while prediction models show potential in the prediction of functional rehabilitation outcomes for post-stroke patients, they require further development, validation, and standardisation to become reliably applicable in

clinical practice. Researchers stress the importance of rigorous validation methods, as many of these models still lack sufficient external validation, limiting their real-world applicability (16). Recent research across various fields and settings shows that only 12.8% of developed models underwent external validation. Even when external validation is sufficient, successful implementation in clinical settings does not automatically follow, which suggests that the proportion of models actually implemented in clinical practice is even lower (17). This highlights a substantial gap between model development and implementation, underscoring the need for prediction models that are externally validated and applicable clinical settings.

Next to sufficient performance in external validation, many factors influence the success of implementing prediction models in clinical practice. Identifying and understanding these factors is crucial, as limited focus on how clinicians receive and use these models often hinders their practicality (18). In a previously conducted literature review on the barriers and facilitators to implementing clinical prediction models (see Appendix A), guided by the Consolidated Framework of Implementation Research (CFIR) (19), several important considerations were highlighted. In the initial stages, a thorough needs assessment and strong stakeholder engagement are crucial to laying the groundwork for a smoother development and implementation process. Co-creation with end-users and continuous feedback loops are essential to maintaining stakeholder consensus and ensuring the model's practicality. Proper structure and design, emphasising clinically relevant and statistically significant variables, along with transparency, comprehensive information provision, clear data representation, and compliance with regulatory pathways, significantly enhance user acceptance and usability. Furthermore, continuous evaluation and regular updates, and adequate resource allocation are necessary to ensure the model's ongoing relevance and effectiveness. Altogether, these factors contribute to more effective approaches for implementing prediction models into clinical workflows (20), helping to close the implementation gap.

The overarching goal of this study is to facilitate the implementation of previously developed prediction models for functional independence in young adults with ABI (21). To achieve this, we focused on two key steps. First, we evaluated the robustness of these models in predicting functional outcomes by externally validating them using a consecutive cohort of newly admitted patients after the development cohort. Second, to facilitate the practical implementation of these models, we developed a user-friendly interface. Informed by insights from our literature review and feedback from clinicians, this interface was designed to assist healthcare professionals in interpreting and applying the model's predictions in real-world settings. The findings of this study are expected to contribute to the field by advancing the iterative development of prediction models and taking essential steps toward creating an effective tool that paves the way for more personalised and effective treatments. Altogether, these efforts contribute to the advancement of a data-driven approach in the transition towards precision neurorehabilitation.

2. Design and methods

2.1. Study design

This study aims to externally validate and take the first steps towards implementation of multivariable prediction models for the prognosis in ABI, utilising prospectively and retrospectively collected clinical data from patients admitted to the Daan Theeuwes Centre (DTC). The DTC is an Intensive Neurorehabilitation Centre for young adults aged 16 to 35 years with severe ABI, undergoing both inpatient and outpatient neurorehabilitation treatments. The previously developed models will be evaluated on the subsequent consecutive cohort of patients. To facilitate the implementation of the prediction models in clinical practice, this study will also include the development of a proof-of-concept tool for clinicians. As highlighted in the previously conducted literature review the emphasis is on statistically significant and clinically relevant variables, co-creation, transparency, comprehensive information provision, clear data depiction, and compliance with regulatory pathways. Together, these efforts represent key steps toward the overarching goal: advancing a data-driven approach in the transition toward precision neurorehabilitation.

This study is structured according to the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) guidelines (22), providing a framework for transparent reporting of prediction model development and validation. However, it is important to note that while this study describes parts of the model development for context, the focus is on the external validation and the tool development. The prediction models were developed prior to this study (21).

2.2. Participants

The participants for this study were young adults admitted to the DTC for Intensive Neurorehabilitation, a level three facility specialising in the neurorehabilitation of individuals with ABI. Eligibility criteria for admission to the neurorehabilitation program included: (1) young adults between the ages of 16 and 35; (2) severe ABI with complex multifactorial impairment; (3) sufficient consciousness to understand and follow the program, indicated by a Post-Acute Level of Consciousness (PALOC) score of 8 (23); (4) medical stability, with no requirement for oxygen or a tracheal cannula, although a replaceable bone flap or PEG tube was not considered a barrier to admission; and (5) active involvement of the patient's support system. Additionally, patients were preferably admitted directly from the acute phase to maximise functional recovery through intensive rehabilitation. Exclusion criteria included (1) patients with neurodegenerative conditions and a poor prognosis, (2) strict isolation requirements, (3) significant psychiatric dysregulation or aggressive behaviour interfering with effective rehabilitation, as well as current addiction problems. All patients and/or their legal guardians provided written informed consent for the reuse of clinical data. The acquisition and storage of the data complied with established guidelines, including the General Data Protection Regulation (GDPR) for data protection and Good Clinical Practice (GCP) standards for clinical research. This study was approved by the medical ethical committee of Amsterdam University Medical Centres (ref: W22_122 # 22.162).

The model development and evaluation did not influence the treatment course.

2.3. Data

The clinical data for this study were derived from the Measurement Feedback System (MFS). The MFS systematically captures structured clinical data at predetermined time points during both inpatient and outpatient neurorehabilitation treatments (21). This systematically collected and structured data provides robust outcome data for prediction models in neurorehabilitation. The development and validation datasets were collected from the same healthcare setting, with identical eligibility criteria, outcomes, and predictors. The data used for the development group includes patients who initiated neurorehabilitation between April 2021 and June 2023. The external validation group consists of a temporally distinct cohort of newly admitted patients who began neurorehabilitation between June 2023 and December 2023.

Comprehensive demographic and medical background information was obtained through custom-made questionnaires distributed to the patient's primary caregiver via the MFS. The collected demographic data included personal information, such as age, sex, socio-cultural background, education, participation in sports, and physical characteristics like dexterity. The medical background encompassed pre-injury learning disorders, neurological disorders, psychiatric disorders, and other medical conditions. Additionally, it included information on prior head injuries and lifestyle factors. Demographic and medical background information was prospectively collected through the MFS, with predefined parameters determined in advance.

The aetiology of ABI and trauma centre data were determined retrospectively through a detailed review of medical referral letters and subsequently standardised using predefined structured registration forms. For all patients, the trauma centres where they were initially admitted were contacted to obtain all relevant medical correspondence and to minimise missing data for analysis. The data extraction process involved identifying the cause of injury (TBI, Cerebrovascular Accident (CVA), infection, or other) and documenting key injury characteristics for each type. For TBI, this included details such as the presence of skull fractures, nerve injuries, and intracranial injuries. For CVA, information was collected on the type and location of the CVA, such as Subarachnoid Haemorrhage (SAH), intracerebral bleeding, or occlusions. Additionally, for infections and other diagnoses, specific details like the type of infection and other related conditions were recorded.

Subsequently, an extensive dataset regarding hospital admission and discharge was collected from the trauma centre where the patient was initially admitted. This included the duration of hospital stay, period of unconsciousness, mechanical ventilation, and the presence of Post-Traumatic Amnesia (PTA). Furthermore, physiological parameters recorded at trauma centre admission were documented, along with details of neurosurgical or radiological interventions. Brain injury severity was assessed using the Glasgow Coma Scale (24), and initial radiology reports regarding CT scans were evaluated based on the Marshall and Rotterdam CT criteria. Functional impairments observed at discharge were also recorded. A detailed overview of the collected data is provided in Table B.1 in Appendix B. Lastly, the Barthel Index (BI) (25) of all patients was extracted from the MFS, which is every 6 weeks. The BI is a measure for the level of independence in daily life activities and is scored between 0 and 20 (see Table 1). The BI is a highly effective tool in rehabilitation, including stroke patients, due to its strong reliability (25; 26) and inter-rater consistency (27). Its sensitivity to functional improvements in patient independence makes it particularly useful to track patients over time (28). The BI's Minimal Clinically Important Difference (MCID) of 1.85 points allows for the detection of significant functional changes (29).

Table 1: **Barthel Index (BI) Scoring Categories**

BI Score	Characteristics
20	Fully independent in basic activities of daily living and mobility
15 – 19	Moderately to highly independent on assistance
10 – 14	Requires assistance but is able to complete some tasks autonomously
5 – 9	Severely dependent on assistance
0 – 4	Fully dependent on assistance

2.4. Data-preprocessing

To ensure data quality and consistency, several pre-processing steps were performed before model development and evaluation, ensuring the datasets were both complete and suitable for further analysis. Predictors with more than 20% missing data were removed to avoid biases, as excessive missing data can lead to unreliable results. Missing values in the remaining data were imputed using the MICE (30) package with Predictive Mean Matching (PMM), a method that preserves relationships within the data while minimising the potential loss of information. Imputation was performed to avoid the exclusion of valuable predictors and to reduce the impact of missing data on model performance.

2.5. Outcomes

The outcome of the developed prediction models is functional independence, as measured by the BI. Three models are developed to predict: (1) "Level of Independence at Admission", (2) "Level of Independence at Three Months Post-Admission" and (3) "Change in Level of Independence over Three Months". Patients without a baseline BI score were excluded from the study. For those who discontinued treatment before the three-month mark, the most recent available BI score was used as the outcome measurement. If a patient reached the maximum BI score before three months, that score was carried forward, assuming their condition remained stable. For patients missing the three-month outcome but with available measurements at earlier and later time points, spline interpolation was used to estimate the missing value.

2.6. Pre-selection of predictors

The candidate predictors for model development included demographic data, medical background, ABI aetiology, and trauma centre information. The large number of predictors compared to a relatively small number of patients led to high dimensionality in the dataset. To avoid overfitting and refine the predictor set, a multi-step pre-selection process was performed. First, manual selection of predictors was performed by experts in the field. Missing data were handled as described in the data preparation. Further refinements were made by excluding binary variables with fewer than 10 occurrences, those with near-zero variance, and variables showing multicollinearity ($r > 0.7$) or linear combinations. These steps were taken to prevent the model from being disproportionately influenced by predictors with limited variability and to enhance its overall stability and generalisability.

2.7. Sample Size

The final development cohort included 100 patients for the "Level of Independence at Admission" and "Level of Independence at Three Months Post-Admission" models, while the model for "Change in Level of Independence Over Three Months" included 70 patients. Further details on the development cohort can be found in the article describing the model development (21). The maximum number of patients assessed for the external validation cohort was equal to the number of newly admitted patients at the centre who met the eligibility criteria.

2.8. Prediction Model Development

The prediction models were developed in a previous study using the CARET package in R. The preprocessing and pre-selection process identified 31 predictors, including BI at Admission, which is only used for the "Change in Level of Independence over Three Months" model. Subsequently, the number of predictors was limited using Recursive Feature Elimination (RFE). The prediction models were trained using a Generalised Linear Model with stepwise Akaike Information Criterion (AIC). A four-fold cross-validation with five repetitions was employed to mitigate overfitting. Bootstrapping with 100 resamples provided robust estimates of predictive accuracy and confidence intervals. Performance was evaluated using the Coefficient of Determination (R^2), Root Mean Square (RMSE), and Mean Absolute Error (MAE). The best-performing sets of predictors were saved. Model performance and predictor sets are presented in Table 2. Full model-building details are provided elsewhere (21).

2.9. External Validation

In this study, the previously developed models were validated. As previously stated, the development and validation datasets were collected from the same healthcare setting, with identical eligibility criteria, outcomes, and predictors. The same data preparation steps regarding missing data and imputation were applied to the external validation cohort as in the development phase. Heterogeneity between the development and external validation datasets was assessed; however, no significant heterogeneity was expected, given that both datasets were derived from the same centre and population. There were no subgroups or variations that required special handling in the analysis. The final set of predictors identified during development was selected from the complete external dataset. The saved models were employed to make predictions, and observed and predicted values were compared. No recalibration or updates to the model were performed during the external validation phase. The model was applied to the validation cohort as originally developed, and its performance was assessed without modification.

2.10. Statistical Analysis

All statistical analyses were performed using RStudio version 2022.07.0+548. Descriptive statistics summarised the validation cohort's characteristics, which were compared with the development cohort to assess baseline similarities across all predictors. For the continuous predictors, the Shapiro-Wilk test was first used to assess normality. Normally distributed variables were compared using an independent two-sample t-test, while non-normally distributed variables were assessed using the Wilcoxon rank-sum test (Mann-Whitney U test). For binary variables, contingency tables were constructed, and the Chi-square test was used for predictors with expected counts greater than five observations per category. When expected counts were less than five per category, Fisher's exact test was applied. To control for false discovery rates, Benjamini-Hochberg (BH) multiple testing correction was applied. A significance level of $p < 0.05$ was used to identify statistically significant differences. Missing data and imputation were evaluated, and differences in data completeness between the development and validation cohorts were also assessed. A difference in completeness of greater than ten percent was deemed notable.

External validation was conducted using the CARET package in R, which supports prediction model training and evaluation. To evaluate the performance of the prediction models and compare it to the performance calculated in the development cohort, R^2, RMSE, and MAE were calculated for each model in the external validation. R^2 was used to assess the fit between the predicted and observed outcomes, where values closer to 1 indicated better model performance. RMSE measured the average magnitude of the prediction errors by calculating the square root of the average squared differences between the predicted and actual outcomes. This metric gives more weight to larger errors, making it particularly useful for identifying models with occasional significant prediction errors. MAE provided insight into the average absolute difference between predicted and actual outcomes, providing a more straightforward interpretation of the typical prediction error. For both RMSE and MAE, values closer to 0 indicated better model performance. Additionally, a 95% Prediction Interval (PI) was calculated for the development cohort to provide a range where 95% of future predictions are expected to fall. While the 95% PI is similar to a 95% Confidence Interval (CI), which estimates the range within which the true population mean lies, the prediction interval also accounts for the uncertainty in predicting individual future outcomes. A narrower interval reflects greater confidence in the model's predictions and is therefore more practical for use in

Table 2: **Performance and Predictor Sets for the Prediction Models for Functional Independence**

Level of Independence at Admission			**Level of Independence at Three Months Post-Admission**			**Change in Level of Independence after Three Months**		
Performance								
R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE
0.657	4.562	3.489	0.593	4.415	3.167	0.763	3.605	2.614
Predictors								
Hospital Length of Stay*			Hospital Length of Stay*			BI at Rehabilitation Admission*		
PTA at Rehabilitation Admission*			Discharge to Intermediate Care Facility*			Left-handedness*		
Discharge to Intermediate Care Facility*			Neurosurgery Performed*			PTA at Rehabilitation Admission*		
Epidural Bleeding on CT Scan*			PTA at Rehabilitation Admission*			TBI with Focal Intracranial Injury*		
Intracerebral Hemorrhage (CVA)*			Skull Fracture(s)			Subarachnoid Bleeding on CT Scan*		
Cranioplasty			Epidural Bleeding on CT Scan			Placement of ICP Monitor*		
Education Level of Parents			Age at Rehabilitation Admission			Participation in Sports*		
Has Children			Participation in Sports					
Sex (Female)								

Abbreviations: R^2, Coefficient of Determination; *RMSE*, Root Mean Square Error; *MAE*, Mean Absolute Error; *PTA*, Post Traumatic Amnesia; *BI*, Barthel Index; *ICP*, Intracranial Pressure.
Each model was trained on a sample size of 100 participants. All models were significant and significant predictors are marked with an *.

clinical settings. The predictions for the external validation cohort were evaluated to determine whether they fell within the 95% PI.

Additional analyses were performed in the external validation cohort to further evaluate model performance. Calibration plots were generated and analysed to assess the agreement between predicted and observed values, providing a visual assessment of how well the model predictions aligned with actual outcomes. These plots were compared across models to evaluate their relative calibration performance. Furthermore, Pearson's correlation was used to compare normally distributed continuous variables, Spearman's correlation was applied for non-normally distributed continuous variables, and rank-biserial correlation was used for binary variables. These correlation coefficients were calculated between each predictor and the absolute prediction error (i.e., the absolute difference between predicted and observed values) to identify the predictors that significantly influenced prediction errors. Given the exploratory nature of this analysis, multiple testing correction was not applied.

Class imbalance techniques were not deemed necessary, as the prediction models involve continuous outcomes. Given the relatively small sample size, stratified fairness testing was not performed. Additionally, no subgroups were expected to be disproportionately favoured, as the model is designed for a focused group. Due to the sensitive nature of patient data, data and code sharing for this study are not feasible. All patient information is protected under privacy regulations; therefore, access to the dataset is restricted. However, detailed information on the methodology and analysis is provided within this paper.

2.11. Model Implementation

A web-based tool was designed as a proof of concept to facilitate the implementation of previously developed prediction models for functional independence in young adults with ABI. The design and structure of the tool were informed by key insights from the previously conducted literature review, which emphasised the importance of statistically significant and clinically relevant variables, co-creation, transparency, comprehensive information provision, clear data representation, and compliance with regulatory pathways. The prediction models guided the inclusion of statistically significant variables, while co-creation was employed through multiple feedback sessions with clinicians to identify and integrate clinically relevant variables into the tool's design. Although no formal testing was conducted, basic adjustments and improvements were made iteratively based on clinician feedback to ensure the tool's functionality and ease of use. The tool was developed using RStudio version 2022.07.0+548 and the Shiny application builder.

3. Results

3.1. Participants

A total of 25 patients were admitted to the DTC since the development phase and considered for inclusion in the external validation cohort. Four patients were excluded due to missing data: two patients had not been admitted to a trauma centre but only attended an outpatient clinic before being referred to the DTC. Given that a substantial proportion of the predictors are related to hospital admission, stay, and discharge, this led to excessive missing data, and these patients were therefore not included in the analysis. Additionally, two patients lacked a measurement of BI at admission. For the remaining 21 patients, missing data at the three-month outcome ($n = 2$) were interpolated using earlier and later BI measurements. Three patients achieved the maximum BI score before the three-month time point, and these scores were carried forward, assuming that their level of independence remained stable over time. Lastly, for the "Change in Level of Independence Over Three Months" model, seven patients were excluded because no change in functional inde-

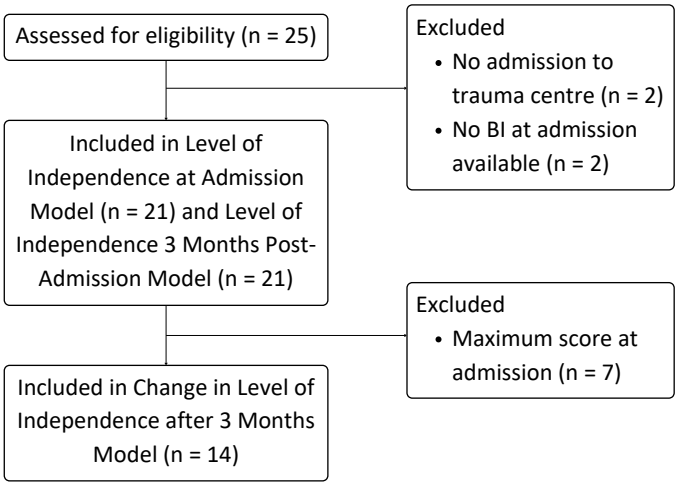


Figure 1: Flow chart for inclusion and exclusion of patients.
Abbreviations: *BI*, Barthel Index.

pendence could be predicted, as they had already reached the maximum BI score at admission.

The final sample sizes consisted of 21 patients for both the "Level of Independence at Admission" and "Level of Independence at Three Months Post-Admission" models. For the "Change in Level of Independence Over Three Months" model, the sample included 14 patients. Figure 1 provides a flow diagram detailing the patient selection process.

3.2. Characteristics

The external validation cohort consisted of patients aged between 15.3 and 34.8 years at injury onset, with 33.3% female sex. Most patients (71.4%) were diagnosed with TBI, all of whom had intracranial injuries (100%), 60% suffered a skull fracture, and 33.3% had cranial nerve injuries. Stroke accounted for 19.1% of cases, with 25% being intracerebral. The remaining 9.5% of patients had other ABI causes. Detailed patient characteristics regarding demographics, medical background, ABI aetiology, and trauma centre data can be found in Table B.1 in Appendix B.

In the validation cohort, overall missing data were relatively low. One important predictor, "PTA at admission to DTC", showed a slightly lower completeness rate of 81%. The variable with the lowest completeness was "EMV score at hospital discharge", with 71% completeness, though it was not a significant predictor in the model. All other variables either showed high completeness (greater than 90%) or were less relevant due to their specificity.

Six predictors demonstrated a difference of more than 10% in completeness between the development and validation cohorts, indicating potential variability in data quality across the two groups. Specifically, four demographic predictors ("Has Children", "Education Level of Parents", "Participation in Sports", and "Left-Handedness") exhibited lower completeness in the validation cohort, ranging from 15% to 21% less than in the development cohort. In contrast, two hospital-related predictors ("Epidural Bleeding on CT Scan" and "Intraventricular or Subarachnoid Bleeding on CT Scan") showed higher completeness in the validation cohort by 27% and 31%, respectively.

Comparison of the development cohort and validation cohort on demographic and clinical characteristics revealed that most group characteristics did not show statistically significant differences. There were no significant differences in demographic characteristics, medical background, ABI aetiology, or outcome measures between the cohorts. Among the trauma centre data, only "Discharge to a Rehabilitation Centre" showed a statistically significant difference between the two cohorts ($p = 0.033$). In the development cohort, 43.3% of patients were admitted directly to specialised medical rehabilitation after discharge from the trauma centre, compared to a much larger proportion of 85.7% in the validation cohort. These results suggest that, apart from one variable in the trauma centre data, the development and validation cohorts are largely comparable across the assessed variables. For a detailed overview of the development cohort, validation cohort and the comparison, see Table B.1 in Appendix B.

3.3. Recovery Level of Independence

A significant improvement in the level of independence, as measured by the BI, was observed over the three-month rehabilitation period in the external validation cohort. The Wilcoxon signed-rank test indicated a statistically significant increase from admission to three months post-admission ($Z = 4.088$, $p < 0.001$, sum of positive ranks = 104). The effect size, measured by Cohen's d, was 1.19 (95% CI: 0.56 – 1.82), indicating a large effect on the recovery of independence. A detailed illustration of the progression in the level of independence is provided in Figure 2. For a comparison of the progression between the development and the validation cohort, see Figure B.1 in Appendix B.

3.4. External validation of the Prediction Models

The performance of all three prediction models was assessed for the external validation cohort and compared to the development cohort. Across all models, a significant drop in performance was observed during external validation, as reflected by a noticeable decrease in R^2. Importantly, the R^2 values for all models in the external validation cohort fell outside the 95% CI of the development cohort, indicating a significant reduction in predictive accuracy compared to the model development phase.

More specifically, for the "Level of Independence at Admission" model, performance decreased significantly, with R^2 dropping from 0.657 (95% CI: 0.555 – 0.758) in the development cohort to 0.428 in the validation cohort. RMSE increased from 4.6 to 5.9, and MAE rose from 3.5 to 4.8. Similarly, the "Level of Independence at Three Months Post-Admission" model experienced a significant reduction in performance, with R^2 dropping from 0.593 (95% CI: 0.465 - 0.695) in the development cohort to 0.297 in the validation cohort. The RMSE increased from 4.4 to 5.6, and the MAE rose from 3.2 to 4.3. The "Change in Independence Over Three Months" model saw the greatest decline in performance. In the development cohort, the model had an R^2 of 0.763 (95% CI: 0.609 – 0.850), but this dropped significantly to 0.359 in the validation cohort. RMSE rose from 3.6 to 6.4, and MAE increased from 2.6 to 4.8.

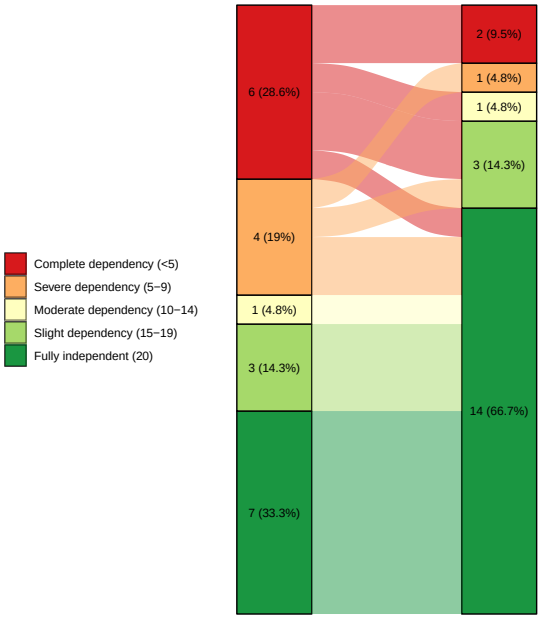


Figure 2: Flow diagram of the trajectories of all patients in the external validation cohort.

This drop in accuracy indicates increased variability and uncertainty in the external validation phase compared to the development phase. Moreover, none of the error measures approached the MCID of 1.85 for the BI in stroke patients (29). The performance metrics for both the development and external validation phases are summarised in Table 3.

3.5. Agreement Between Observed and Predicted Values

Calibration analysis was performed to assess the agreement between the predicted and observed BI scores, identifying any systematic deviations or biases in the model's predictions across different levels of independence. All three models showed a general positive correlation between predicted and observed BI scores, following the overall trend of increasing predicted scores with higher levels of observed independence.

The "Level of Independence at Admission" tends to underestimate independence for patients with lower observed BI scores ($0 - 10$) and to overestimate for higher scores ($11 - 20$). Additionally, only 57.1% ($n = 12$) of the predicted values in the validation cohort fell within the 95% PI of 5.9 from the development cohort (see Figure 3a).

In the "Level of Independence at Three Months Post-Admission" model, predicted scores were concentrated above 8. There was a also tendency for overestimation at the lower observed scores ($0 - 10$) and underestimation at the higher range ($11 - 20$), with a grouping of underestimated points at observed BI scores of 20. In this model, 71.4% ($n = 15$) of the predicted values fell within the 95% PI of 6.2 from the development cohort (see Figure 3b).

For the "Change in Independence Over Three Months" model, there was greater variability compared to the other models. Overestimation was common for lower observed changes in independence ($0 - 10$), while underestimation occurred for larger observed changes (above 15). Only 50% ($n = 7$) of the predicted values in the validation cohort fell within the 95% PI of 5.0 from the development cohort (see Figure 3c).

3.6. Predictor Influence on Prediction Error

To explore how individual predictors influence the magnitude of prediction error, a correlation analysis was performed. It revealed that "PTA at Rehabilitation Admission" is moderately to strongly associated with the magnitude of prediction error in the "Level of Independence at Admission" model. This indicates that the presence of PTA complicates accurate predictions of functional independence. Additionally, patients with longer hospital stays show larger discrepancies between predicted and actual outcomes in the "Level of Independence at Three Months Post-Admission" model. Conversely, in the "Change in Level of Independence over Three Months" model, a higher BI at admission and the presence of focal injury are associated with smaller prediction errors. Tables with all calculated correlations can be found in Appendix C.

3.7. Implementation

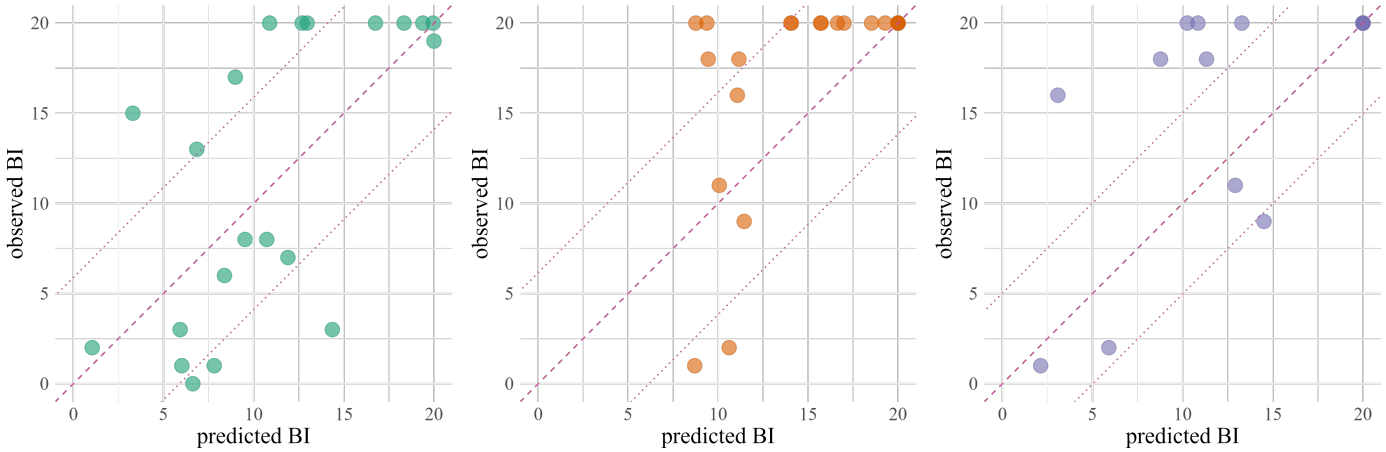
Although external validation has not yet confirmed strong model performance in temporally external data, the first step towards implementation has been taken by developing a proof-of-concept tool to identify potential obstacles and evaluate its use before full implementation. The development of the interface and incorporation of clinician feedback ensure a clear plan is in place once the model's performance reaches a level that supports full clinical implementation. The tool was structured based on key insights from the literature review, particularly the emphasis on statistically significant and clinically relevant variables, co-creation, transparency, comprehensive information provision, clear data depiction, and compliance with regulatory pathways. These principles guided the design of three pages: "Decision Support Tools" to generate predictions, "Patient-Like-Me" depicting an interactive flow diagram (see Figure 4b), and "About" for additional information provision. The full-page visualisations from the developed tool are provided in Appendix D, as substantial space is required to clearly depict all features and functionalities.

The first page of the tool (Figure D.1) serves as the primary interface for the use of the prediction models. For each model, users enter values for the final set of predictors identified during development. By linking the data server and user input, the tool generates real-time predictions using the prediction models. Additionally, predictors identified as statistically significant during the development phase (21) were incorporated as filters in the flow diagram using dropdown menus (see Subfigure 4b and 4c). The filters allow the flow diagram to adapt, displaying only patients matching the selected criteria (Figure 5). Making these significant predictors accessible as filters enhances clinical interpretability and transparency, allowing clinicians to explore the factors driving the model's predictions and better understand how each predictor influences functional independence.

Table 3: **Performance metrics of the development and external validation phases for the prediction models.**

	Level of Independence at Admission		**Level of Independence after 3 Months**		**Change in Independence after 3 Months**	
Cohort	Development	Validation	Development	Validation	Development	Validation
R^2	0.657	0.428	0.593	0.297	0.763	0.359
95% CI	0.555 – 0.758	-	0.465 – 0.695	-	0.609 – 0.850	-
RMSE	4.6	5.9	4.4	5.6	3.6	6.4
MAE	3.5	4.8	3.2	4.3	2.6	4.8
95% PI	5.9	-	6.2	-	5.0	-

Abbreviations: R^2, Coefficient of Determination; *RMSE*, Root Mean Square Error; *CI*, Confidence Interval; *MAE*, Mean Absolute Error; *PI*, Prediction Interval.



(a) Level of Independence at Admission (b) Level of Independence at Three Months Post-Admission (c) Change in Level of Independence over Three Months

Figure 3: Calibration plots for the external validation cohort, showing observed vs. predicted Barthel Index (BI) values. The dashed line represents perfect calibration (predicted = observed), and the dotted lines show the 95% Prediction Interval (PI) from the development cohort.

To further enhance transparency and provide clear information, The third tab, titled "About" (Figure D.8), provides comprehensive information about the prediction models, including instructions for use and additional information on the performance metrics. This ensures that clinicians have access to important information about the models' reliability, promoting trust in the tool and transparency regarding its development and function. While no formal feedback mechanism is integrated, contact information is available for suggestions or issues.

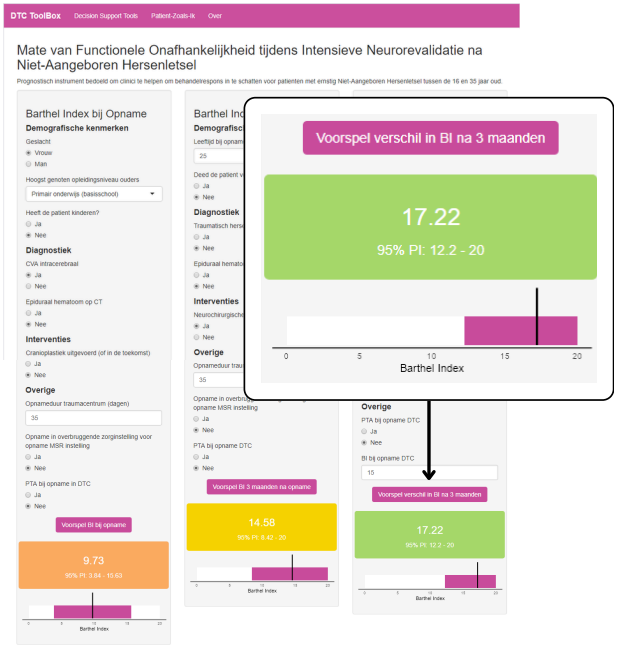
Clear data depiction was achieved through an intuitive layout. On the first page, input fields were provided for the significant predictors of each model, accompanied by a colour-coded button to calculate the prediction (Figure 4a). Predicted values, displayed alongside a 95% PI, are shown to two decimal places, allowing clinicians to precisely compare predictions with the MCID. Additionally, these values are visualised in a graph, ensuring that the data are presented in a clear and easily understandable format for clinicians. To comply with regulatory pathways regarding privacy, such as the GDPR and GCP, the tool does not store any Personally Identifiable Information (PII). Patient data are used temporarily for input and calculations but are not saved or stored at any point.

4. Discussion

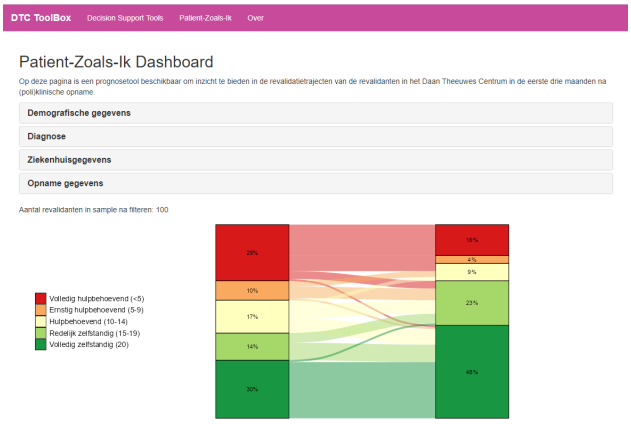
4.1. Interpretation of Results

The goal of this study is to contribute to precision neurorehabilitation, where patient-specific treatments are guided by data-driven insights to enhance functional recovery. Achieving this ambition requires the successful implementation of prediction models that help clinicians assess patient outcomes and tailor treatment to personalised needs. In this study, we aimed to externally validate prediction models for functional independence in young adults with severe ABI during the first three months of intensive neurorehabilitation and develop a proof-of-concept tool for the practical application of these models. These efforts represent an important step toward enabling the implementation of prediction models in clinical practice. The results of this study highlight the critical role of structured data collection as the basis for continuous model improvement and practical implementation. Additionally, they offered important insights into the functional recovery trajectories of young adults with severe ABI, stressing the need for further model refinement to improve accuracy. These prediction models must evolve alongside emerging clinical insights and the collection of patient data.

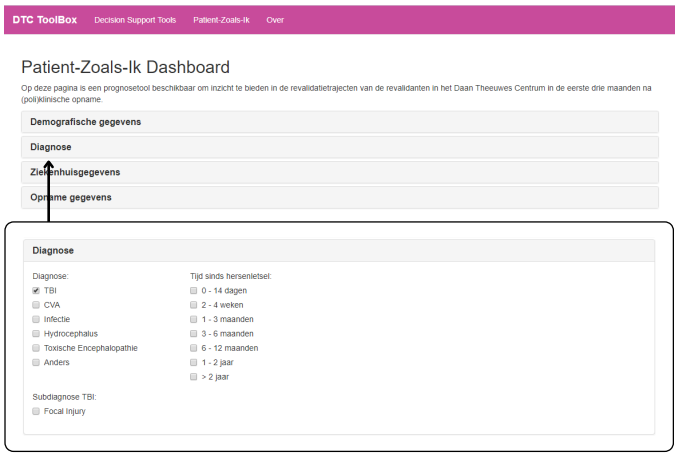
By applying the prediction models to a subsequent consecutive cohort, we have highlighted the challenges of generalising models to new patient cohorts and the importance of continu-



(a) The "Decision Support Tools" page allows clinicians to input predictors. The output inlcudes the Barthel Index (BI) (black line) with a 95% prediction interval (PI) (pink bar).



(b) The "Patient-Like-Me" page depicts the trajectories of 100 included patients in develop- ment between admission and three months Post-Admission.



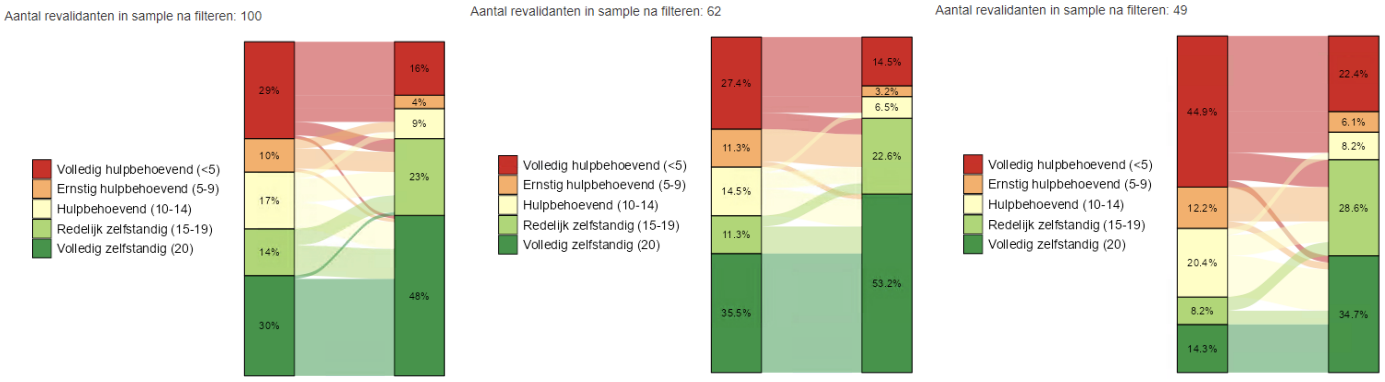
(c) On the "Patient-Like-Me" page, dropdown menus allow data filtering by categories and subcategories, such as "Diagnosis" and "Focal Injury", providing a focused representation.

Figure 4: Some of the Tool's functionalities.

ous model refinement as previously highlighed by Pommerich et al. (15). The external validation offered us insights into both the strengths and limitations of the prediction model. While the model maintained its predictive capabilities to some extent, it displayed lower accuracy and generalisability in the external cohort compared to the development phase . Although a ceiling effect was anticipated due to the capping of predictions at the maximum score of 20, the actual results revealed a different trend. Instead of predictions clustering around the maximum predicted score, high observed BI scores were often underestimated. Similarly, lower observed BI scores were frequently overestimated, following the same trend in reverse. This suggests that the predictors currently used in the model may not fully capture the factors driving both high and low levels of functional independence. It underscores the complexity of predicting functional outcomes in this population and highlights areas where the model can be improved for more accurate predictions. Further refinement in future development cycles could address this issue by utilising larger datasets or more advanced models to better capture patients' recovery potential across the full BI spectrum.

Additionally, the correlation analysis revealed that predictors associated with more challenging rehabilitation trajectories—such as PTA and long hospital stays—were associated with a higher prediction error. This aligns with clinical expectations, as these factors are known to indicate more complex recovery paths, which are inherently more difficult to predict. Incorporating more nuanced information about these variables could help improve model accuracy for these difficult cases. Conversely, a higher BI at admission was linked to smaller prediction errors in the "Change in Level of Independence over Three Months" model. Although this was also found by Meyer et al. (14) there is also less room for error with higher initial independence. Focal injury was also associated with smaller errors, which may be due to most patients with focal injury having higher BI scores at admission. Given the small sample size (two high BI scores in three patients with focal injury), this finding should be explored further with a larger dataset to determine if the observed pattern is consistent. Additionally, implementing Out-of-Distribution (OOD) detection could enhance the model's robustness by identifying patients whose characteristics fall outside the development data, alerting clinicians to underrepresented profiles. This would prevent over-reliance on the model in cases where accuracy may be compromised, improving reliability and clinical utility in complex rehabilitation cases.

In an effort to bridge the implementation gap, steps were taken to optimise adoption of the prediction models once they reach a level suitable for full clinical use. The development of the web-based tool is grounded in insights from both the literature review and feedback sessions with clinicians in neurorehabilitation settings. One of the tool's most valuable features for transparency is the interactive flow diagram. By allowing filtering based on key demographic, medical, and injury-related factors, it offers clinicians insights into the patient data underpinning the prediction model and helps them better understand recovery trajectories. Although the flow diagram is based on



(a) Flow diagram of the full sample (n = 100).

(b) Flow diagram depicting patients in the sample with TBI. (n = 62).

(c) Flow diagrams depicting patients in the sample with PTA at Rehabilitation Admission (n = 49).

Figure 5: The interactive flow diagram allows clinicians to filter on patient data to visualise patients specific recovery trajectories depicted in Barthel Index (BI) and to offer insights into the data used to derive the prediction models.
Abbreviations: *TBI*, Traumatic Brain Injury; *PTA*, Post-Traumatic Amnesia.

the development cohort (n = 100), filtering narrows the dataset significantly. However, this more focused subset of patients still provides uniquely valuable insights into the rehabilitation trajectories of patients with unprecedented representativeness when compared to data from the literature or other rehabilitation centres.

An essential part of implementing the tool is establishing clear performance targets for clinical use, such as using the MCID (29) as a benchmark for individual predictions. While such targets ensure reliability, the tool can still provide valuable insights during development by aligning its goals with the current level of accuracy. Each iteration offers an opportunity to reassess its applications, and even if not yet precise enough for individual outcome predictions, the tool's predictions—along with their uncertainty range—remain useful for outlining potential prognostic scenarios. This allows for both refining the model and delivering practical value in clinical settings as the tool evolves.

While the current tool focuses on predicting functional independence at admission and after three months of rehabilitation, there is potential for expanding its scope. Additional outcome measures, such as cognitive or physical recovery, could be integrated to provide a more comprehensive view of patient progress. Moreover, extending the timeline beyond three months to include more follow-up points would offer a clearer picture of a patient's long-term recovery trajectory. By incorporating these features, the tool could become an even more valuable asset in supporting a data-driven approach to optimising neurorehabilitation pathways.

The results of this study highlight the critical role of structured data collection as the basis for continuous model improvement and practical implementation. As Campagnini et al. (16) found, the implementation of digital infrastructures can effectively support a data-driven clinical environment. In this study, the MFS provided a solid foundation for both the development and validation cohorts. By leveraging this system, we achieved minimal discrepancies between datasets. However, maintaining this consistency over time is vital. As medical correspondence varies depending on the source of the data, the terminology in clinical data can become complex. Inconsistent use of terms or variables can introduce deviations within a dataset or even within individual variables. Such small variations can have a significant impact on model performance, especially when they affect strong predictors.

The ongoing data collection through the MFS supports the continuous refinement of the prediction models by expanding the dataset. As the dataset grows, or more complex models are introduced, an iterative process of alternating between retraining and validation will continue until the models reach a satisfactory level of predictive accuracy. A first step in this can be found in Appendix E, where an update of the model including patients from both cohorts (n = 121) is presented. Once the models demonstrate sufficient reliability, they will be implemented in clinical practice and continuously monitored through ongoing external validation cycles, which will track how shifts in patient populations may affect predictive accuracy. Model updates or retraining will only occur when significant changes in the data or clinical context require it, ensuring that the models remain relevant and accurate over time. This approach allows for the dynamic adaptation of the models to reflect evolving clinical needs and supports their long-term clinical applicability.

4.2. Contribution to the Field

Similar models to those employed in this study have been developed in fields like orthopaedics (31; 32). While they are applied in different clinical contexts, these models also highlight the importance of comprehensive datasets. More closely related to our field, studies by Meyer et al. (14) and Pommerich et al. (15), which focused on stroke patients, emphasised the difficulty of generalising prediction models due to variability in predictor sets across studies. In their review, Pommerich (15) et al. examined regression-based prognostic models for functional independence following post-acute brain injury rehabilitation. The literature search yielded only six studies with internally validated multivariable prognostic models, and just two included external validation—both of which lacked clear proce-

dures. They emphasised that many current prognostic models lack the methodological rigour needed for widespread clinical use. Additionally, Kwakkel et al. (33) underscored the importance of balancing practicality and precision when developing clinically relevant and pragmatic models for stroke rehabilitation. While there are similarities between our work and other models, our prediction model stands out for its specific focus on neurorehabilitation for young adults with ABI during the crucial early recovery phase. This fills a significant gap in the literature and offers a tool that could improve patient-specific care in this population.

4.3. Limitations and Strengths

An important limitation of this study is the variability in data quality across different sources. As patients were referred from various trauma centres, the communication of injury and treatment details was not always consistent. This variability impacted the completeness of data in both the development and validation cohorts, highlighting the need for more standardised documentation practices. However, implementing such changes may pose challenges due to increased clinician workloads and the need for behavioural shifts. Despite not achieving full completeness, the MFS provided a solid framework for capturing all available data in a structured and usable format and ensured minimal discrepancies between cohorts.

Additionally, the relatively small sample size limits the generalisability and statistical power of the prediction models, potentially constraining the model's applicability to new patient cohorts. However, the continuous nature of data collection through the MFS at the Daan Theeuwes Centre ensures that larger, more comprehensive datasets will be accumulated over time. This ongoing data collection will support further development and evaluation cycles, potentially enhancing model performance and allowing for the exploration of more advanced methods, such as clustering and neural networks, to improve predictive accuracy and power.

A key strength of this study is the inclusion of an interactive flow diagram, which enhances transparency and offers insights into the patient data underlying the prediction models. This transparency builds clinician trust and supports more informed, data-driven decision-making regarding patient recovery trajectories. In addition, the parallel development of both the prediction model and the user-friendly tool streamlines the implementation process, taking important steps towards practical clinical use as the model continues to evolve and is refined.

4.4. Conclusions

We have taken steps toward implementing prediction models for functional independence in young adults with ABI through two main components—external validation of prediction models and the practical implementation via the development of a user-friendly tool. Together, these components emphasise the importance of structured data collection, robust validation, and practical application to ensure the models' effectiveness in clinical practice. As the dataset grows, there is potential to explore more sophisticated modelling techniques, which could enhance the model's predictive accuracy and practical utility. Moving forward, future iterations could expand on the model by incorporating outcomes such as cognitive and physical recovery, while also extending the timeline beyond the current three-month focus. These ongoing advancements bring the field closer to a data-driven approach, where neurorehabilitation becomes increasingly personalised, tailored to the unique needs and recovery trajectories of young adults with ABI.

References

[1] Goldman L, Siddiqui EM, Khan A, Jahan S, Rehman MU, Mehan S, et al. Understanding Acquired Brain Injury: A Review. Biomedicines. 2022 9;10(9). Available from: /pmc/articles/PMC9496189/ /pmc/articles/PMC9496189/?report=abstracthttps: //www.ncbi.nlm.nih.gov/pmc/articles/PMC9496189/.

[2] Feigin VL, Barker-Collo S, Krishnamurthi R, Theadom A, Starkey N. Epidemiology of ischaemic stroke and traumatic brain injury. Best Practice & Research Clinical Anaesthesiology. 2010 12;24(4):485-94.

[3] Dewan MC, Rattani A, Gupta S, Baticulon RE, Hung YC, Punchak M, et al. Estimating the global incidence of traumatic brain injury. Journal of neurosurgery. 2018 4;130(4):1080-97. Available from: https: //pubmed.ncbi.nlm.nih.gov/29701556/.

[4] Mukherjee D, Patil CG. Epidemiology and the Global Burden of Stroke. World Neurosurgery. 2011 12;76(6):S85-90.

[5] Stocchetti N, Zanier ER. Chronic impact of traumatic brain injury on outcome and quality of life: A narrative review. Critical Care. 2016 6;20(1):1-10. Available from: https://link. springer.com/articles/10.1186/s13054-016-1318-1https: //link.springer.com/article/10.1186/s13054-016-1318-1.

[6] Dijkers MP. Quality of life after traumatic brain injury: a review of research approaches and findings. Archives of Physical Medicine and Rehabilitation. 2004 4;85(SUPPL. 2):21-35.

[7] Haghgoo HA, Pazuki ES, Hosseini AS, Rassafiani M. Depression, activities of daily living and quality of life in patients with stroke. Journal of the Neurological Sciences. 2013 5;328(1-2):87-91.

[8] Moeller D, Carpenter C. Factors affecting quality of life for people who have experienced a stroke. https://doiorg/1012968/ijtr2013204207. 2014 6;20(4):207-16. Available from: https://www.magonlinelibrary. com/doi/10.12968/ijtr.2013.20.4.207.

[9] Jolliffe L, Lannin NA, Cadilhac DA, Hoffmann T. Systematic review of clinical practice guidelines to identify recommendations for rehabilitation after stroke and other acquired brain injuries. BMJ Open. 2018 2;8(2):e018791. Available from: https://bmjopen.bmj.com/content/8/2/e018791https: //bmjopen.bmj.com/content/8/2/e018791.abstract.

[10] Cullen N, Chundamala J, Bayley M, Jutai J. The efficacy of acquired brain injury rehabilitation. Brain Injury. 2007;21(2):113-32. Available from: https://www.tandfonline.com/doi/abs/10. 1080/02699050701201540.

[11] Covington NV, Duff MC. Heterogeneity Is a Hallmark of Traumatic Brain Injury, Not a Limitation: A New Perspective on Study Design in Rehabilitation Research. American Journal of Speech-Language Pathology. 2021 2;30(2S):974-85. Available from: https://pubs.asha.org/ doi/abs/10.1044/2020_AJSLP-20-00081.

[12] Lee YH, Bang H, Kim DJ. How to Establish Clinical Prediction Models. Endocrinology and Metabolism. 2016 3;31(1):38. Available from: /pmc/articles/PMC4803559//pmc/articles/ PMC4803559/?report=abstracthttps://www.ncbi.nlm.nih. gov/pmc/articles/PMC4803559/.

[13] Reinkensmeyer DJ, Burdet E, Casadio M, Krakauer JW, Kwakkel G, Lang CE, et al. Computational neurorehabilitation: modeling plasticity and learning to predict recovery. Journal of NeuroEngineering and Rehabilitation 2016 13:1. 2016 4;13(1):1-25. Available from: https://link.springer.com/articles/10.1186/ s12984-016-0148-3https://link.springer.com/article/10. 1186/s12984-016-0148-3.

[14] Meyer MJ, Pereira S, McClure A, Teasell R, Thind A, Koval J, et al. A systematic review of studies reporting multivariable models to predict functional outcomes after post-stroke inpatient rehabilitation. Disability and rehabilitation. 2015;37(15):1316-23. Available from: https: //pubmed.ncbi.nlm.nih.gov/25250807/.

[15] Pommerich UM, Stubbs PW, Eggertsen PP, Fabricius J, Nielsen JF.

Regression-based prognostic models for functional independence after postacute brain injury rehabilitation are not transportable: a systematic review. Journal of clinical epidemiology. 2023 4;156:53-65. Available from: https://pubmed.ncbi.nlm.nih.gov/36764467/.

[16] Campagnini S, Arienti C, Patrini M, Liuzzi P, Mannini A, Carrozza MC. Machine learning methods for functional recovery prediction and prognosis in post-stroke rehabilitation: a systematic review. Journal of NeuroEngineering and Rehabilitation. 2022 12;19(1):1-22. Available from: https://jneuroengrehab.biomedcentral.com/articles/10.1186/s12984-022-01032-4.

[17] Arshi B, Smits LJ, Wynants L, Cowley LE, Reeve K, Rijnhart E. Number of publications on new clinical prediction models: a systematic literature search. 2023 7. Available from: https://osf.io/4txc6.

[18] Kappen TH, Van Loon K, Kappen MAM, Van Wolfswinkel L, Vergouwe Y, Van Klei WA, et al. Barriers and facilitators perceived by physicians when using prediction models in practice. Journal of Clinical Epidemiology. 2016 2;70:136-45. Available from: http://www.jclinepi.com/article/S0895435615004278/fulltexthttp://www.jclinepi.com/article/S0895435615004278/abstracthttps://www.jclinepi.com/article/S0895-4356(15)00427-8/abstract.

[19] Damschroder LJ, Reardon CM, Widerquist MAO, Lowery J. The updated Consolidated Framework for Implementation Research based on user feedback. Implementation Science 2022 17:1. 2022 10;17(1):1-16. Available from: https://implementationscience.biomedcentral.com/articles/10.1186/s13012-022-01245-0.

[20] Proctor EK, Powell BJ, McMillen JC. Implementation strategies: Recommendations for specifying and reporting. Implementation Science. 2013 12;8(1):1-11. Available from: https://implementationscience.biomedcentral.com/articles/10.1186/1748-5908-8-139.

[21] Van der Veen R, Oosterlaan J, Klein Kranenbarg E, Bos M, Welsink-Karssies M, van Westrhenen A, et al. Towards Precision Rehabilitation Medicine after Acquired Brain Injury: Exploring the Prediction of Patient Independence using Structured Clinical Data. Manuscript submitted for publication. 2024.

[22] Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) the TRIPOD statement. Circulation. 2015;131(2):211-9. Available from: https://www.ahajournals.org/doi/10.1161/CIRCULATIONAHA.114.014508.

[23] Van De Wiel M, Eilander H. De PALOC-s: de postacute bewustzijnsschaal voor het vaststellen van het bewustzijnsniveau van mensen met niet-aangeboren hersenletsel.

[24] Jain S, Iverson LM. Glasgow Coma Scale. StatPearls. 2023 6. Available from: https://www.ncbi.nlm.nih.gov/books/NBK513298/http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1032822.

[25] Collin C, Wade DT, Davies S, Horne V. The Barthel ADL Index: a reliability study. International disability studies. 1988;10(2):61-3. Available from: https://pubmed.ncbi.nlm.nih.gov/3403500/.

[26] Evaluation of the psychometric properties of the Barthel Index in an Italian ischemic stroke population in the acute phase: a cross-sectional study - PubMed;. Available from: https://pubmed.ncbi.nlm.nih.gov/31172937/.

[27] Duffy L, Gajree S, Langhorne P, Stott DJ, Quinn TJ. Reliability (inter-rater agreement) of the Barthel Index for assessment of stroke survivors: systematic review and meta-analysis. Stroke. 2013 2;44(2):462-8. Available from: https://pubmed.ncbi.nlm.nih.gov/23299497/.

[28] Castiglia SF, Galeoto G, Lauta A, Palumbo A, Tirinelli F, Viselli F, et al. The culturally adapted Italian version of the Barthel Index (IcaBI): assessment of structural validity, inter-rater reliability and responsiveness to clinically relevant improvements in patients admitted to inpatient rehabilitation centers. Functional neurology. 2017 10;22(4):221-8. Available from: https://pubmed.ncbi.nlm.nih.gov/29306359/.

[29] Hsieh YW, Wang CH, Wu SC, Chen PC, Sheu CF, Hsieh CL. Establishing the minimal clinically important difference of the Barthel Index in stroke patients. Neurorehabilitation and neural repair. 2007 9;21(3):233-8. Available from: https://pubmed.ncbi.nlm.nih.gov/17351082/.

[30] Package 'mice' Title Multivariate Imputation by Chained Equations. 2023.

[31] Tsehaie J, Spekreijse KR, Wouters RM, Feitz R, Hovius SER, Slijper HP, et al. Predicting Outcome After Hand Orthosis and Hand Therapy for Thumb Carpometacarpal Osteoarthritis: A Prospective Study. Archives of Physical Medicine and Rehabilitation. 2019 5;100(5):844-50.

[32] Wouters RM, Porsius JT, Van Der Oest MJW, Slijper HP, Souer JS, Selles RW, et al. Psychological Characteristics, Female Sex, and Opioid Use Predict Acute Postoperative Pain in Patients Surgically Treated for Thumb Base Osteoarthritis: A Cohort Study. Plastic and Reconstructive Surgery. 2020 12;146(6):1307-16. Available from: https://journals.lww.com/plasreconsurg/fulltext/2020/12000/psychological_characteristics,_female_sex,_and.17.aspx.

[33] Kwakkel G, Kollen BJ. Predicting activities after stroke: what is clinically relevant? International journal of stroke : official journal of the International Stroke Society. 2013 1;8(1):25-32. Available from: https://pubmed.ncbi.nlm.nih.gov/23280266/.

Appendix A. Literature Review

”Evaluating the Barriers and Facilitators of the Implementation of Prediction Models in Clinical Practice”

E.H.M. Lie[a,d], M. Königs[c,d], W.C. Peul[b]

[a]*Delft University of Technology, Erasmus Medical Centre, Leiden University Medical Centre*
[b]*Department of Neurosurgery, LUMC, Leiden University, Leiden, Netherlands*
[c]*Emma Neuroscience Group, Emma Children's Hospital, Amsterdam UMC, University of Amsterdam, Netherlands*
[d]*Daan Theeuwes Centre for Intensive Neurorehabilitation, Woerden, Netherlands*

Abstract

The objective of this study was to systematically review the barriers and facilitators to implementing clinical prediction models (CPMs) in clinical practice. Despite the extensive development of CPMs, their transition into clinical settings remains limited. This study aimed to identify key factors influencing adoption across various healthcare settings, synthesise current evidence, and propose a structured framework for enhancing the implementation of CPMs to advance neurorehabilitation care of acquired brain injury (ABI). A comprehensive literature review was conducted using Medline and Embase databases. Articles published between 2008 and the present were screened based on inclusion criteria focused on studies examining the implementation of multivariate predictive models in clinical settings. A total of 18 studies were included after a systematic selection process. Data were extracted and categorised using the Consolidated Framework for Implementation Research (CFIR), with key barriers and facilitators identified and translated into actionable steps within a proposed implementation framework. The review identified 20 barriers and 23 facilitators related to CPM implementation, organised across five CFIR domains: Innovation, Outer Setting, Inner Setting, Characteristics of Individuals, and Implementation Process. Notably, factors such as model transparency, adaptability, user interface design, and integration into existing workflows emerged as critical facilitators. Complexity, lack of perceived need, and poor resource availability were significant barriers. This study provides a comprehensive overview of the barriers and facilitators to implementing CPMs in clinical practice and proposes a practical framework to guide their integration. Addressing these factors early in the implementation process is essential for enhancing the adoption of precision medicine approaches, ultimately optimising clinical workflows and improving patient outcomes in ABI rehabilitation and beyond.

Keywords: prediction modelling, implementation, acquired brain injury, adoption, clinical practice, clinical decision making, barriers, facilitators

Contents

1. Introduction

Traumatic brain injury (TBI) and stroke are the two most prevalent forms (1) of acquired brain injury (ABI), affecting approximately 85 million individuals annually (2; 3). ABI can result in serious consequences, including cognitive impairment, physical disability, and emotional distress, which can profoundly impact the quality of life for affected individuals (4; 5; 6; 7). Neurorehabilitation focuses on regaining

motor function, daily activities, cognition, communication and psychosocial factors to increase the ability to participate in society (8).

A major challenge in successful neurorehabilitation after ABI is caused by distinct differences in the presentation of ABI between individuals (9). The complex interaction between a range of relevant factors, including demographics, aetiology of the neuropathology and environmental factors, results in significant heterogeneity in terms of the impairment profile, progression during neurorehabilitation, and outcomes among individuals with ABI (9). This heterogeneity limits healthcare professionals in establishing a prognosis and communicating reasonable expectations (10). Consequently, this complicates the development of effective treatment strategies for ABI, as the diverse presentations require personalised approaches and can hinder the ability to develop and apply standardised treatment protocols (11).

The emerging field of precision medicine tailors treatment to individual characteristics, potentially overcoming challenges in ABI therapy (12; 13). Detailed information about patient differences and their influence on outcomes is essential to apply precision medicine effectively. However, in current clinical practice, a significant amount of information is assessed subjectively, and personal preferences in the choice of assessment tools further contribute to variability. As a result, challenges arise in ensuring the quality and completeness of the data collected (14; 15). This lack of standardised data collection obstructs research and the development of effective treatment strategies. Systematic collection and analysis of clinical data can improve our understanding of the complex interactions and enable more precise, individualised treatments for ABI patients (16). Furthermore, extensive databases enable the application of predictive analytics techniques to gain valuable insights into rehabilitation potential and to improve treatment and outcomes (17; 18).

Prediction modelling in healthcare involves developing and applying mathematical algorithms and statistical techniques to predict various outcomes relevant to patient care, such as prognosis and risk of disease (19). For instance, models have been developed to predict the occurrence of postoperative nausea and vomiting (PONV) (20) and sepsis (21). In a notable study, Boussina et al. (22) demonstrated the impact of a clinical prediction model (CPM) on sepsis management outcomes. They developed a real-time deep-learning model implemented in two emergency departments (EDs) and found that its use was associated with a reduction in in-hospital sepsis-related mortality.

Unlike the aforementioned sepsis model, few models have been successfully integrated into clinical practice. While there has been a surge in the development of CPMs, a significant implementation gap persists (23; 24; 25; 26). Although a substantial number of models have been published, their transition into routine clinical practice remains remarkably low. External validation is one of the last steps before actual implementation, which assesses the model's performance in a new, independent dataset. A systematic review of CPM development articles reveals that while thousands of models have been published, only a small fraction undergo external validation (27). Specifically, only 12.8% of these models underwent external validation, implying that the proportion of models implemented in clinical settings is likely to be much lower. This small percentage highlights a substantial gap between model development and practical application. Additionally, there is often a need for more focus on how clinicians receive and use these models, which further limits their implementation (20).

Implementing prediction models in clinical practice presents a significant obstacle. It is essential to identify and understand the factors that influence clinicians' adoption of prediction modelling, such as organisational dynamics and individual preferences of end-users, in addition to the characteristics of the tool itself (28). Incorporating these factors into developing implementation strategies leads to more effective strategies (29). Advanced strategies have the potential to increase the actual implementation of CPMs. These developments drive the transition towards precision medicine, ultimately improving patient care and outcomes.

Extensive research has focused on implementing CPMs in specific settings, such as the emergency room and psychiatry (30; 31). While these settings may differ in context, they often encounter similar challenges, illustrating that the lessons learned and strategies developed are broadly applicable across healthcare settings. Analysing experiences from all these settings can provide comprehensive insights into developing an implementation strategy for ABI patients.

This study aims to conduct a thorough literature review to identify barriers and facilitators of implementing CPMs in clinical practice. The results will be used to develop a framework that provides clinicians with a comprehensive overview of barriers and facilitators of these models. Additionally, they will be translated into actionable steps for guidance during and adequate focus on certain factors before and during the implementation of CPMs, ensuring that they are effectively adopted and used to improve patient outcomes, streamline workflows, and enhance the precision of clinical decision-making.

2. Design and methods

2.1. Search strategy

A comprehensive search was conducted across two electronic databases, Medline and Embase, to identify relevant literature. The search used a combination of keywords and Boolean operators such as "prediction model," "implement," and "clinical practice." The search strategy targeted studies published between 2008 and the present, written in English, and focusing on implementing multivariable predictive models in clinical practice. Additionally, a citation search was performed to identify any relevant articles missed in the initial database search. The complete search string can be found in Appendix A.

2.2. Inclusion and Exclusion Criteria

Studies were included if they met the following criteria: (1) addressed the implementation of prediction modelling in clinical practice; (2) were published in peer-reviewed journals; (3) were written in English; (4) were published between 2008 and

the present as this timeframe aligns with major advancements in multivariable models that occurred towards the end of the 2000s (32). Studies were excluded if they: (1) focused on model development, performance, or clinical utility measures; (2) did not systematically outline barriers and facilitators for implementation; (3) did not involve the implementation of a multivariable prediction model.

2.3. Study Selection Process

The study selection process consisted of two phases: screening and data extraction. Initially, titles and abstracts of retrieved articles were screened against the inclusion and exclusion criteria. Full-text articles were then retrieved and assessed for eligibility.

2.4. Data Synthesis and Analysis

Relevant data were extracted from each included study using a standardised data extraction form. Key information extracted included study characteristics such as author, publication year, location, study design, medical field, setting, type of model, and users. Moreover, factors related to implementing prediction models in clinical practice were extracted using the Consolidated Framework for Implementation Research (CFIR) (28). The CFIR is a theoretical framework used to understand and evaluate the implementation of interventions across various settings, including healthcare. It comprises five major domains: Intervention Characteristics, Outer Setting, Inner Setting, Characteristics of Individuals, and Process. Each domain is further divided into multiple constructs, as outlined in Table 1.

From each study, the barriers and facilitators identified by the authors were extracted and categorised into domains and their constructs outlined by the CFIR. Within constructs, similarities and differences between articles were presented to offer a complete overview of the implementation factors observed across studies. All constructs that were initially not directly associated with barriers and facilitators found in the included articles were analysed to detect any overlooked or misclassified data to ensure thorough coverage. If necessary, constructs were merged, renamed or added to fit the study context. The barriers and facilitators were translated into actionable steps and divided into four key phases: Problem Definition and Planning, Model Development and Evaluation, Training and Deployment and Evaluation and Updating. This structured approach allows for the planning of specific focus required at each stage, ensuring that all potential issues are addressed timely and systematically.

3. Results

3.1. Literature Search

A search of electronic databases yielded 979 articles. After removing duplicates, 538 articles remained for title and abstract screening. Following this initial screening, 34 articles underwent full-text review. Ultimately, ten studies met the inclusion and exclusion criteria. Additionally, eight studies were identified and included through citation search, resulting in 18 studies in the qualitative synthesis. Figure 1 presents a PRISMA diagram outlining the study selection process.

3.2. Overview of Included Studies

The included studies were published between 2014 and 2023 and represented global research from sites across various countries such as the United States (n=10, 55,6%), the Netherlands (n=3, 16.7%), Canada (n=2, 11.1%), Japan (n=1, 5.6%), Singapore (n=1, 5.6%) and Sweden (n=1, 5.6%). Of the 18 included studies, nine (50,0%) included data from multiple centres, while nine (50,0%) included data from a single centre. Among the studies involving multiple centres, four (44,4%) conducted their research in non-hospital care facilities, three (33,3%) in a combination of care facilities, one (11,1%) exclusively in academic hospitals, and one (11,1%) in non-academic hospitals. Regarding the monocentre studies, six (66,7%) were conducted in academic hospitals, two (22,2%) in non-academic and one (11,1%) in a non-hospital care facility.

Different types of prediction models were implemented. Seven (38,9%) articles discussed barriers and facilitators regarding multiple models, six (33,3%) implemented a type of multivariate regression model, and five (27,8%) implemented machine learning (ML) models, of which one (5,6%) implemented an XGBoost model, one (5,6%) implemented a deep learning (DL) model and three (16,7%) machine learning models were unspecified.

The included studies encompassed a diverse range of methodologies to investigate what factors lead to success of implementation, including a mixed methods approach (n=5, 27,8%), combining interviews, surveys, focus groups and questionnaires, (semi-structured) interviews (n=5, 27,8%), opinion articles (n=3, 16,7%), surveys (n=2, 11,1%), a focus group (n=1, 5.6%), a review (n=1, 5.6%), and a case example (n=1, 5.6%). The outcomes from these methodologies, such as interview responses, survey data, and writer comments, were systematically analysed and translated into barriers and facilitators. These findings were then structured into a framework that can be used as a guide during implementation. An overview of the included studies can be found in Appendix B.

3.3. Factors Influencing the Implementation of Prediction Models in Clinical Practice

Our search found 20 barriers and 23 facilitators organised by 22 CFIR constructs. Figure 2 summarises the CFIR and corresponding constructs, barriers and facilitators. Figure 3 depicts the number of the included studies reporting each construct, and Table 2 presents an overview of all barriers and facilitators identified within their respective constructs.

3.3.1. Innovation Domain

Of the included studies, 77,8% (n=14) reported barriers and/or facilitators related to six constructs within the Innovation domain: Transparency, Relative Advantage, Design, Complexity, Evidence-Based and Adaptability.

Transparency

Mentioned at the highest frequency (n=9, 50,0%) within the innovation domain is the Transparency construct (20; 33; 34; 35; 36; 37; 38; 39; 40). Healthcare professionals expressed the desire to be able to view and understand the inner workings of the prediction models in order to trust them. They want to be able to assess the performance metrics of the CPM, such as false positives and negatives. Furthermore, they want to know each contributing parameter and its significance in the model to inter-

Table 1: Overview of domains of the Consolidated Framework of Implementation Research (CFIR).

Domain	Contents
Innovation	Constructs specific to the implemented prediction models, such as their complexity, adaptability, and evidence base.
Outer Setting	External factors that may influence implementation, such as regulatory requirements and local attitudes.
Inner Setting	Organisational context in which the prediction models are being implemented, including culture, infrastructure, and availability of resources and training.
Individuals	Characteristics of healthcare providers and other stakeholders involved in the implementation process, including attitudes, beliefs, and comfort with technology.
Implementation Process	Planning, assessing needs, and evaluating the implementation effort.

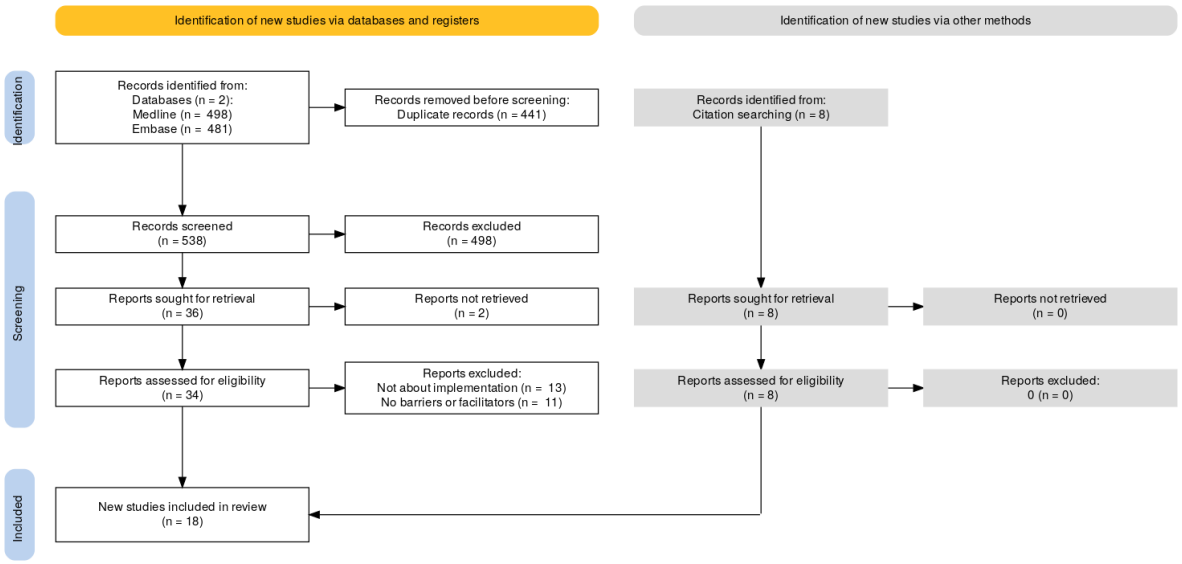


Figure 1: PRISMA Flow Diagram

pret results effectively (33; 37; 39; 40; 41). This transparency is essential for confidently using the model in clinical practice and communicating results with colleagues and patients (33; 39).

Relative Advantage

The other most reported barriers and facilitators were related to the relative advantage of the model. Nine (50,0%) articles emphasised the model's actionability in aiding clinical decision-making. This actionability involved the model's ability to propose actionable options (20; 34; 35; 37; 38; 39; 42; 43; 44), such as intervention suggestions (34; 43; 38), additional questions to ask the patient, and additional diagnostic tests to perform (34; 37). However, in one (5,6%) study, it was slightly more often argued that it would not be desirable for a CPM to suggest interventions (37). Another study (5,6%) added that a CPM could add value by helping focus on important aspects, serving as a reminder, ensuring continuity, providing a second opinion, or making judgement visible to others (34). Additionally, clinicians highlighted that the model is expected to serve as a complementary tool that enhances, rather than replaces, clinical judgement (34; 37).

Design

Seven (38.9%) articles discussed the design of a CPM. Frequently discussed topics included clear user interface (34; 37; 38; 40) and understandable and interpretable data depiction using graphs, timelines and colours (34; 37; 38) as they significantly contribute to model understanding and interpretation. Moreover, accessibility emerged as an important topic, with preferences depending on the application. While one (5,6%) study noted a positive experience with web apps and mobile device apps due to the mobility of the app (33)], other studies (n=5, 27,8%) stressed the need for Electronic Health Record (EHR) integration for the centralisation of information and accessibility reasons (20; 34; 35; 39; 44). One (5,6%) study suggested that model adoption could be improved by presenting

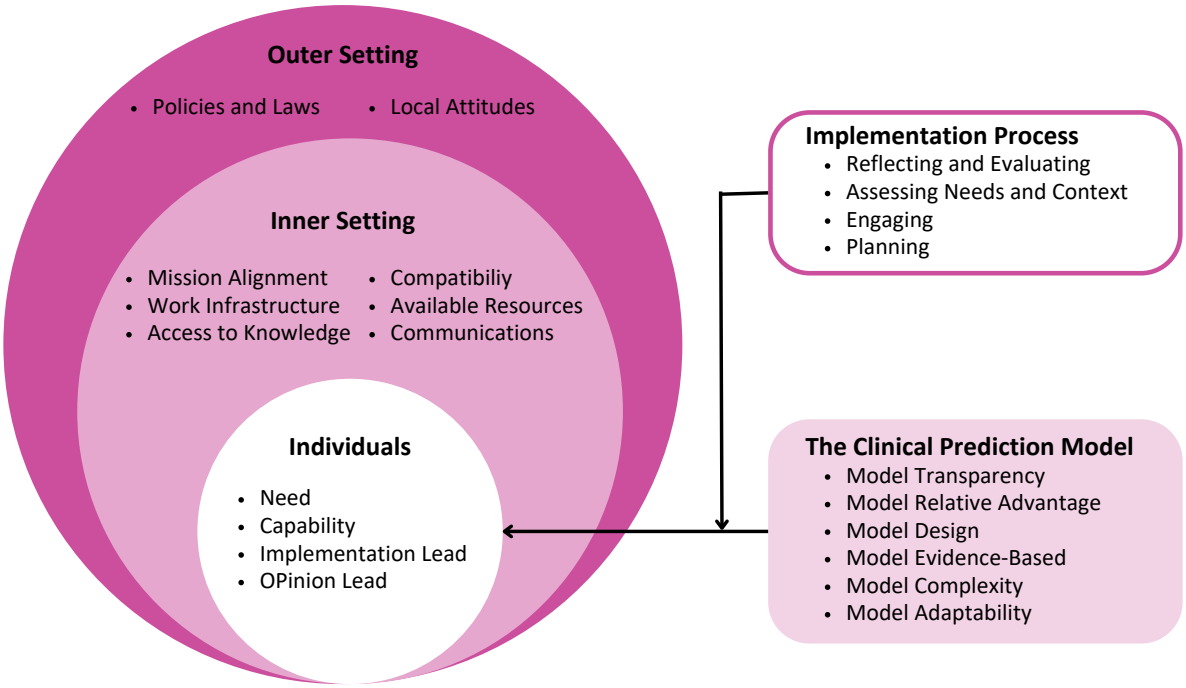


Figure 2: A visual overview of the constructs identified per domain in the included literature.

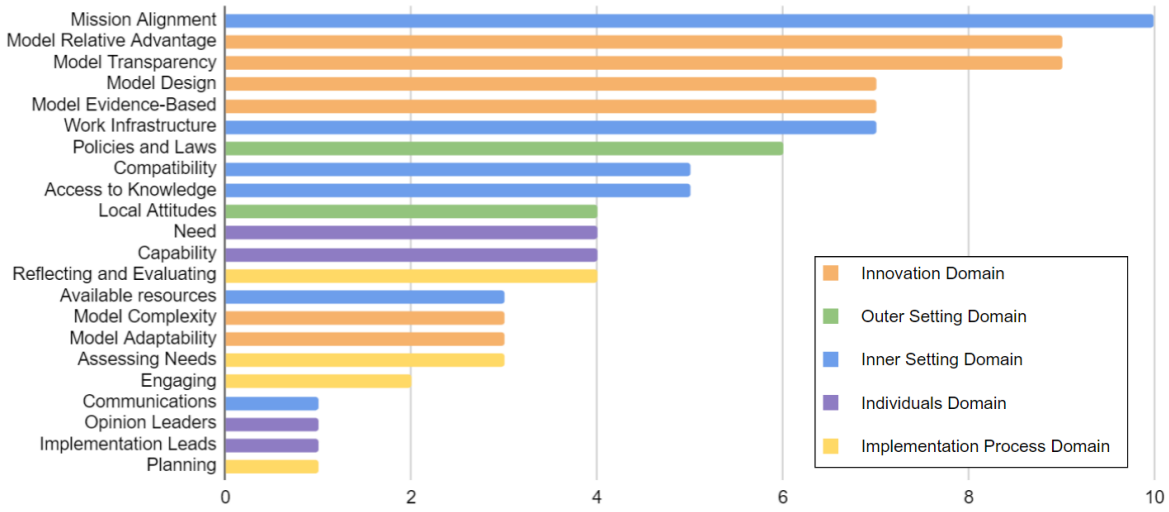


Figure 3: Number of studies discussing each construct coloured by domain; Innovation (orange), Outer Setting (green), Inner Setting (blue), Individuals (purple), Implementation Process (yellow).

Table 2: Overview of barriers and facilitators of the implementation of prediction models in clinical practice.

	Barriers	**Facilitators**
I. Innovation Domain		
Model Transparency	Poor transparency	
Model Relative Advantage		Model outperforms clinical judgement Assist clinical decision-making Not a substitute for clinical judgement
Model Design	Poor accessibility Not a multimodal package	Clear user-interface Understandable and interpretable data depiction Web-app and/or mobile app
Model Evidence-Based	Poor performance	Clinically and statistically important factors
Model Complexity	High complexity	
Model Adaptability		Fit local context and needs
II. Outer setting		
Policies & Laws	Liability	Regulatory compliance
Local Attitudes	Ethics	Availablility to clinicians and patients
III. Inner setting		
Mission Alignment	Potential (psychological) harm	Model fits purpose Improves care, comfort, or costs
Work Infrastructure	Complex and time-consuming factors Alert fatigue and timing	Quick, easy and user friendly Standardised workflow and documentation
Access to Knowledge		Systematic training before use Information available about the model
Compatibility	Not integrated in EHR	
Available resources	Lack of input resources Lack of output resources Low data quality in EHR Personnel Funding	
Communications		Expectation management
IV. Individuals		
Characteristics		
Need	Lack of perceived need	
Capability	Intuitive vs. analytical approach	Only use of familiar and capable
Roles		
Opinion Leaders		Recommended by expert in the field
Implementation Leads		Key stakeholders oversee and guide use
V. Implementation process		
Reflecting & Evaluating		Evaluation, updating and maintenance
Assessing Needs	Unclear consensus among stakeholders	Co-creation: developers and clinicians
Engaging		Outreach local clinicians and prompts
Planning	Identifiable leads and endpoints	

EHR, Electronic Health Record

the model as a multimodal package (45), meaning that various tools can be accessed through the same application.

Evidence-Based
The main topic in this construct, which was discussed in seven (38.9%) articles, was the barrier of poor performance (33; 36; 38; 42; 43), with one (5,6%) study mentioning the importance of adequate sample size and sufficient performance in external validation to enhance the reliability and acceptability of prediction models (33). Doubts about the model's validity were found to be a strong barrier to clinical implementation (38). Additionally, one (5,6%) study highlighted the significance of utility studies in measuring the impact on clinical care and patient outcomes, noting that a CPM is not valuable if it does not change behaviour or is too generalised (33). Another study (5.6%) underscored that demonstrating the safety of CPMs serves as a strong justification for their use (42). Furthermore, incorporating clinically and statistically important parameters ensures a clinically valuable outcome (33; 38; 40; 45).

Complexity
Three (16.7%) articles expressed concerns about the complexity of prediction models in clinical practice (35; 36; 40). One (5.6%) article highlighted that while transparency in CPMs is beneficial, it can also add complexity. Even when all parameters and performance metrics are available, understanding the inner workings and results of the model can remain difficult due to the numerous variables involved. This complexity can be challenging for clinicians, underscoring the need for clear and concise explanations to aid their interpretation and application of the model's outputs (35).

Adaptability
Three (16.7%) studies noted that a model needs to be adaptable to a specific setting regarding input parameters used, necessary output parameters, actionable options available in the department, and the threshold at which action is required by the treating clinician to be effective across multiple settings (36; 37; 40).

3.3.2. Outer Setting Domain
Identified within the Outer Setting domain, 38,9% (n=7) of the included studies reported barriers and/or facilitators related to two constructs: Policies and laws and Local Attitudes.

Policies and Laws
Six (33,3%) articles, of which five (27,8%) in the US, expressed concerns about compliance with the law and regulatory pathways and the liability of using prediction modelling in clinical practice. Compliance with laws and regulatory pathways relating to the innovation is crucial to mitigate legal risks and enhance the acceptance of prediction models in clinical practice (39; 41; 42). Concerns were raised about the potential for malpractice risk and liability exposure due to the use, misuse, or non-use of prediction models in clinical practice (42). Clear guidelines and legal protections regarding liability were deemed essential to implement prediction models in clinical settings safely (39; 41).

Local Attitudes
In addition to regulatory responsibility, there is a strong ethical responsibility for using CPMs. Four (22.2%) articles identified ethics as a significant factor in implementing CPMs. Ethical considerations include the ethics of risk communication (36), ensuring that patients have the autonomy to consent to or opt out of data use (41), and clarifying who will have access to risk information (37; 41), how it will be displayed or stored in the EHR (41), and who will be expected to respond (36). There is a risk of inappropriate treatment, especially if prediction models are used without adequate training or not for the intended purpose (41). Institutional leadership, procedural infrastructure, and regulatory oversight are essential to ensure that prediction models are used safely and effectively, considering all ethical guidelines and responsibilities (42).

3.3.3. Inner Setting Domain
Reported by 72,2% (n=13) of the included studies, barriers and/or facilitators related to seven constructs within the Inner Setting domain included Mission Alignment, Work Infrastructure, Access to Knowledge, Compatibility, Available resources, Culture and Communications.

Mission Alignment
The most discussed topic (n=10, 55.6%) highlighted the importance of mission alignment in implementing CPMs. Most articles suggest that a CPM is valuable if it leads to one of the following outcomes: improved patient outcomes, enhanced patient understanding, increased physician satisfaction, or reduced health system costs (35; 37; 43; 44). However, one article (5.6%) emphasised that merely improving process metrics without impacting clinical outcomes is insufficient, arguing that true added value lies only in improved patient outcomes (35).

Work Infrastructure
Seven (38.9%) articles addressed the need for CPMs to fit the work infrastructure to facilitate implementation. To fit high-pressure environments with limited resources and time, CPMs should be quick, easy, and user-friendly (33; 38; 40; 43). Models that include too many complex parameters, especially those that are time-consuming to obtain or require additional resources, can hinder their applicability in clinical settings. Despite their significant contribution to prediction, including less easily measured variables restricts a model's practical use in these environments (33; 43; 45). Alert fatigue (35; 37; 39; 40) was also highlighted multiple times, with one (5,6%) study underscoring the relevance of delivering the message in a structured manner to the right person at a relevant and convenient time. Standardised workflows for using CPMs can facilitate their integration into clinical practice and prevent confusion regarding their application (37; 39; 40). Proper documentation is essential to report the role of CPMs within the workflow and explain decision-making processes (37).

Access to Knowledge
Five (27.8%) articles highlighted the importance of access to knowledge in understanding and effectively using CPMs and their workflows. Whether mandatory or not, systematic training before the clinical use of CPMs is crucial for ensuring that healthcare professionals are adequately prepared (34; 37; 39; 40). Furthermore, detailed information about the algorithm, key

test characteristics such as accuracy and sensitivity, and how to integrate the tool into clinical care plans and practices should always be available for reference (34; 37; 39). Additionally, clinicians must be well-trained in their field to understand when and when not to use CPMs and effectively communicate results with patients (39).

Compatibility
The inconvenience of the lack of integration into existing EHR systems was discussed by five (27.8%) studies (20; 34; 35; 39; 44). Relevant information should be available in a consolidated place within the EHR to ensure accessibility for those who need it and that all patient data is saved in one central location (39).

Available resources
Three (16.7%) articles stressed the necessity of adequate available resources to enable the use of CPMs. The accuracy of CPMs based on EHR data heavily depends on the quality of the EHR input data. Inaccurate, noisy, biased, unsystematically recorded, or incomplete EHR data leads to insufficient performance of the CPM (35; 37; 41). Additionally, the healthcare facility must be able to provide the model's output, such as increased hospitalisation rates in resource-limited settings. This can lead to over-hospitalisation and diminished care quality (37). Furthermore, ongoing development and maintenance of CPMs require dedicated personnel and expertise to ensure their continued effectiveness (37). This requirement for ongoing support is further challenged by a lack of funding, particularly for personnel and the technical specifications necessary for integrating and expanding models in clinical settings (35).

Communications
One (5.6%) study noted managing expectations as a barrier to the clinical implementation of these models, as enthusiasm often led to unrealistic expectations among clinicians (35).

3.3.4. Individuals Domain
Barriers and/or facilitators related to four constructs were reported within the Individuals Domain by 44,4% (n=8) of the included studies, including Need, Capability, Implementation Leaders and Opinion Leaders.

Need
A lack of perceived need for CPMs was discussed in four (22.2%) articles (34; 35; 38; 44). Clinicians felt they did not require additional tools like CPMs alongside their clinical judgement (34). Furthermore, they were uncertain about the necessity of new tools in general and expressed concerns about alert and change fatigue (39). Even in one study (5.6%), where research suggested that clinical judgement often overestimated the score, a lack of perceived need for the CPM was observed (44).

Capability
Four (22.2%) articles highlighted factors related to the capability of end-users of CPMs (20; 35; 39; 40). Clinicians, who often rely on an intuitive approach to care and decision-making, may find it challenging to adapt to CPMs based on an analytical approach (20; 40). The values used for evaluating prediction models are less familiar to most clinicians, leading to a lack of knowledge and understanding to assess machine learning models' validity (35; 40). Clinicians emphasised the importance of having the necessary capabilities and skills to use CPMs effectively and feeling familiar with the system to ensure successful implementation (39; 40).

Implementation Leaders
One (5.6%) study highlighted the need for organisational leadership. The setup should involve key stakeholders who can guide the appropriate clinical application of prediction models. These leaders are crucial for promoting the effective and safe implementation of CPMs across various settings (42).

Opinion Leaders
One (5.6%) study highlighted the significant influence of recommendations and clinical practice guidelines from experts in the field. This support is an important facilitator as it enables clinicians to effectively integrate prediction models into their practice, ensure proper application, and communicate results to patients with confidence (33).

3.3.5. Implementation Process Domain
For five constructs within the implementation process domain, 33,3% (n=6) of the included studies reported barriers and/or facilitators, encompassing Assessing Needs, Reflecting and Evaluating, Engaging and Planning.

Reflecting and Evaluating
Four (22,2%) studies emphasised the necessity of continuous evaluation, updating, and maintenance to keep CPMs relevant and effective (34; 35; 41; 42). Regular updates ensure the model remains applicable and accurate with the latest data (42).

Assessing Needs
Three (16.7%) studies emphasised the importance of thoroughly assessing the needs that a CPM should address (35; 39; 42). Co-creation between developers, clinicians, and other stakeholders was identified as a strong facilitator (35; 39; 42). Clinicians expressed the need to be involved in discussions, planning, training, and making system requests throughout the process, from identifying clinical needs and informing development to determining optimal intervention points and defining questions and interventions. This involvement enhances acceptance and minimises workflow disruption (39). However, one (5.6%) study noted the challenge of building consensus among stakeholders, which was a barrier (35).

Engaging
Two (11.1%) studies discussed the importance of outreach and prompts as facilitators for engaging clinicians with CPMs (37; 41). These studies noted that clinicians preferred receiving notifications over opening an extra tool on their own initiative (37; 41). However, they also acknowledged this approach's risk of alert fatigue (37).

Planning
One (5.6%) study emphasised the importance of having identifiable leads and clear endpoints in the planning process for the implementation of CPMs (35). This approach is particularly

beneficial for smaller institutions and those with less experience in the clinical use of models, as it brings clarity and can help reduce costs (35).

3.4. Implementation strategy

The identified factors are translated into actionable steps and divided into four specific phases: Problem Definition and Planning, Model Development and Evaluation, Training and Deployment, and Evaluation and Updating. This phased approach clarifies which barriers and facilitators must be addressed at each process stage. Many barriers and facilitators are defined early in the process, well before the actual implementation phase. By organising these factors according to the implementation phases, we can ensure that the appropriate focus is given to each aspect of the process at the right time, thereby enhancing the overall effectiveness and success of implementing CPMs. The phases and corresponding actions are provided in Appendix C.

4. Discussion

The primary aim of this study was to evaluate the barriers and facilitators of implementing CPMs in clinical practice, focusing on their application in ABI patients. From 18 articles, 20 barriers and 23 facilitators were identified across five domains containing 22 constructs, providing comprehensive insights into the factors influencing CPM implementation. The main contribution of this study is a review of the current literature on implementing CPMs in clinical practice across various healthcare settings. Based on this evidence, a framework has been developed to identify and structure barriers and facilitators using the CFIR. Furthermore, this framework has been translated into an actionable guide for clinicians to effectively integrate CPMs, thereby advancing the adoption of precision medicine and ultimately improving patient outcomes.

4.1. Interpretation of Results

Barriers and facilitators were observed in all five domains of the CFIR, with a significant concentration in the Innovation Domain and the Inner Setting Domain. This concentration is likely due to the early development stage of CPMs. Implementation difficulties are often caused by challenges earlier in the process, which should be addressed at that stage. To provide practical guidance, the adapted CFIR framework was reorganised into phased, actionable steps, supporting the process from initial development to implementation. This section presents findings according to these phases, highlighting key actions and their corresponding impact on implementing a CPM.

The success of a CPM in clinical practice heavily depends on the first phase: Project Definition and Planning. Research underscores that co-creation is crucial in healthcare innovation (46). Clinicians should express their specific needs, and model developers and institutional bodies should clearly define the range of possibilities and restrictions throughout the full process (47).

A thorough needs assessment involves identifying the clinical problem and tailoring the CPM to the local context, which may reduce the lack of perceived need during implementation and ensure a smooth integration of the CPM into the local workflow. As supported by Davis et al. (48), ensuring the system is perceived as useful and easy to use can significantly enhance user acceptance and integration into clinical practice. Furthermore, Engelhardt et al. (49) showed that perceived need can be a strong facilitator, as clinicians used a CPM in consultations, not for its original intended purpose, but because it conveniently aided in patient explanations and understanding.

Key characteristics of the model should be determined. Input parameters should be limited in number, clinically relevant, statistically significant, and easy and quick to retrieve. The most convenient source of input data is directly from EHRs, but this data must be systematically collected and evaluated to ensure it is suitable for model incorporation (50). Depending on the application of a CPM and end-users, output parameters should be actionable or not, based on preference, and their consequences should be considered within the given context. As Kline-Simon et al. (51) found, CPMs can cause a significant increase in workload. Furthermore, careful evaluation of who should and should not have access is necessary for security and ethical reasons, ensuring safe data use while maximising potential care improvements. Additionally, human resources are crucial for successful implementation. A committee should oversee the process, plan, secure milestones, and ensure consensus among stakeholders. Involving an expert can drive progress and manage expectations by clearly communicating goals. Identifying and allocating personnel with the right expertise and securing sufficient funding is essential to support all phases, including technological and personnel costs.

During Model Development and Evaluation, the second phase, co-creation should continue through a feedback loop to maintain stakeholder consensus. The most important precondition for successful implementation, which can be facilitated during development, is achieving sufficient performance metrics and the ability to outperform clinical judgement (52). Compliance with medical device regulations is crucial to ensure the CPM meets legal and safety standards. Furthermore, developers must consider the environment in which the model will be used and ensure it operates efficiently under the time pressure users face. Next to the technical requirements of a CPM, the tool's design is crucial to address barriers to implementation. In the CPM, the included parameters and their significance should be available for comparison with one's judgement for making well-considered decisions. Moreover, comprehensive information and instructions about the model should be readily available to ensure proper usage. The balance between necessary notifications must be carefully managed to avoid alert fatigue. Studies have shown that while clinicians appreciate reminders, excessive alerts can lead to desensitisation and reduced effectiveness. For instance, Ancker et al. (53) found that high workloads and complex work environments significantly contribute to alert fatigue, diminishing the overall effectiveness of clinical decision support systems. Therefore, optimising the alert system to support clinical workflow without overwhelming clinicians is crucial.

The design should reflect decisions from the planning phase, such as whether the model will be web-based, a mobile application, or integrated with the EHR. As Mohammed et al. (54) showed, it is important to tailor the application to the environ-

ment and resources available. User-friendliness, including intuitive interfaces and clear data depiction, is essential to facilitate ease of use and quick interpretation. A standardised workflow for model usage and proper documentation ensures ethical use and compliance with regulatory standards. The TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis) guidelines were developed to ensure that studies on prediction models are reported transparently and comprehensively, improving their reproducibility and reliability (55).

The third phase, Training and Deployment, focuses on sufficient education and standardised workflows. Before implementing the CPM in the clinical environment, end-users must undergo thorough training to ensure they are fully capable and confident in using the model effectively. The complexity of a model may not always be reducible, but proper education can mitigate confusion and help clinicians understand the logic behind the model. This training allows clinicians to compare their intuitive approaches with the model's analytical output, enhancing acceptance and perceived need. Training should cover the standardised workflow and emphasise the importance of proper documentation.

The last phase, Evaluation and Updating, focuses on keeping the model relevant and effective. Binuya et al. (56) provide methodological guidance for evaluating and updating clinical prediction models, emphasising the importance of systematic reviews and structured approaches to maintain model accuracy and applicability over time (56). Continuous impact measurement is essential to assess whether the CPM improves clinical care and process metrics. Additionally, out-of-distribution analysis can identify when the model encounters data that differ significantly from the training data, helping to detect when predictions might be less reliable. Co-creation remains vital throughout all phases to ensure feedback is incorporated, aligning the model with clinical practice and increasing its adoption.

Overall, the key findings of this study can be structured to provide a framework and guidance in all stages of implementation of a CPM. Implementation research reveals that the proportion of implemented models is extremely low relative to developed models. Barriers and facilitators should be addressed in the early stages to ensure that steps can be taken towards implementation. Addressing barriers and leveraging facilitators in the initial stages, particularly through needs assessment, stakeholder engagement, and resource allocation, lays the groundwork for a smoother development and implementation process. Co-creation and continuous feedback loops are essential throughout the development phase to maintain stakeholder consensus and ensure the model's practicality and compliance with regulatory standards. Proper design and user training are vital to mitigate potential barriers such as alert fatigue and to enhance user acceptance and usability. Finally, continuous evaluation, updating, and resource allocation are necessary to ensure the model's ongoing relevance, effectiveness, and integration into clinical practice, ultimately improving patient outcomes and streamlining workflows.

4.2. Limitations and strengths

The terminology describing prediction modelling in the literature can be quite varied, often leading to the identification of many articles when setting up a search. Due to time constraints, concessions were made regarding the search terms to obtain a manageable number of articles for review. Expanding the search terms might have resulted in a more extensive selection of articles. Despite this, the identified factors are likely representative and reflect the same ratio found in a broader search. Furthermore, a citation search was applied to identify additional studies. Another limitation is the reliance on self-reported data from clinicians, which is subject to subjectivity and can introduce bias. Nonetheless, the various approaches, such as the mixed methods approach, incorporating interviews, focus groups, and surveys, provide multiple ways to capture subjective data, reducing this issue. Additionally, the included studies varied in design and quality, which may affect the consistency and comparability of the findings. Lastly, one limitation of this study is the potential bias due to the number of studies conducted in the United States, which may restrict the generalisability of the findings to other healthcare settings. However, the other studies represent various countries, enhancing the robustness and applicability of the conclusions across different healthcare systems.

A major strength of this study is the systematic approach to identifying and categorising barriers and facilitators using the CFIR framework. This structured method ensures that our findings are organised within a well-established theoretical model, providing a comprehensive approach that minimises the potential of overlooking important factors. Developing a practical and actionable overview based on these findings also offers a valuable tool for clinicians and healthcare administrators. Lastly, incorporating all healthcare settings enabled the collection of more valuable data across various settings.

4.3. Contribution to the field

Our findings significantly contribute to advancing precision medicine for ABI patients and other settings by providing a structured pathway for implementing CPMs. This pathway can be used to develop and integrate more personalised treatment strategies. Furthermore, the developed framework and guide can be applied to research in broader healthcare contexts, supporting overall healthcare improvements and stimulating innovation.

4.4. Future directions

Future research should focus on the impact of data quality on CPMs. Studies focusing on enhancing data collection are necessary, emphasising fully leveraging EHR integration as it remains central to patient data management. In addition to addressing existing challenges, user design should prioritise transparency in data and models, with experiments exploring how data depiction influences clinical reasoning and decision-making. Moreover, medical students should be familiarised with CPMs early to understand the advantages and pitfalls. Finally, a continuous impact measurement framework should be established to ensure CPMs remain aligned with their primary

goal of improving healthcare as they become more widely implemented.

5. Conclusion

This study systematically identified and categorised barriers and facilitators to implementing CPMs in clinical practice using the CFIR framework. The primary finding is that the successful implementation of CPMs is dependent on addressing key factors across all five CFIR domains: Innovation, Outer Setting, Inner Setting, Individuals, and Implementation Process. The significant concentration of barriers and facilitators within the Innovation Domain and Inner Setting Domain highlights the need for mitigating obstacles early in the process. By translating barriers and facilitators into an actionable guide for clinicians, this study facilitates the effective adoption of CPMs in rehabilitation for ABI patients and contributes to the broader field of precision medicine. Ultimately, this approach aims to enhance patient outcomes, streamline clinical workflows, and support the development of more accurate and personalised treatment strategies and other healthcare innovations.

References

[1] Feigin VL, Barker-Collo S, Krishnamurthi R, Theadom A, Starkey N. Epidemiology of ischaemic stroke and traumatic brain injury. Best Practice & Research Clinical Anaesthesiology. 2010 12;24(4):485-94.

[2] Dewan MC, Rattani A, Gupta S, Baticulon RE, Hung YC, Punchak M, et al. Estimating the global incidence of traumatic brain injury. Journal of neurosurgery. 2018 4;130(4):1080-97. Available from: https://pubmed.ncbi.nlm.nih.gov/29701556/.

[3] Mukherjee D, Patil CG. Epidemiology and the Global Burden of Stroke. World Neurosurgery. 2011 12;76(6):S85-90.

[4] Stocchetti N, Zanier ER. Chronic impact of traumatic brain injury on outcome and quality of life: A narrative review. Critical Care. 2016 6;20(1):1-10. Available from: https://link.springer.com/articles/10.1186/s13054-016-1318-1https://link.springer.com/article/10.1186/s13054-016-1318-1.

[5] Dijkers MP. Quality of life after traumatic brain injury: a review of research approaches and findings. Archives of Physical Medicine and Rehabilitation. 2004 4;85(SUPPL. 2):21-35.

[6] Haghgoo HA, Pazuki ES, Hosseini AS, Rassafiani M. Depression, activities of daily living and quality of life in patients with stroke. Journal of the Neurological Sciences. 2013 5;328(1-2):87-91.

[7] Moeller D, Carpenter C. Factors affecting quality of life for people who have experienced a stroke. https://doiorg/1012968/ijtr2013204207. 2014 6;20(4):207-16. Available from: https://www.magonlinelibrary.com/doi/10.12968/ijtr.2013.20.4.207.

[8] Jolliffe L, Lannin NA, Cadilhac DA, Hoffmann T. Systematic review of clinical practice guidelines to identify recommendations for rehabilitation after stroke and other acquired brain injuries. BMJ Open. 2018 2;8(2):e018791. Available from: https://bmjopen.bmj.com/content/8/2/e018791https://bmjopen.bmj.com/content/8/2/e018791.abstract.

[9] Goldman L, Siddiqui EM, Khan A, Jahan S, Rehman MU, Mehan S, et al. Understanding Acquired Brain Injury: A Review. Biomedicines. 2022 9;10(9). Available from: /pmc/articles/PMC9496189//pmc/articles/PMC9496189/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC9496189/.

[10] Cullen N, Chundamala J, Bayley M, Jutai J. The efficacy of acquired brain injury rehabilitation. Brain Injury. 2007;21(2):113-32. Available from: https://www.tandfonline.com/doi/abs/10.1080/02699050701201540.

[11] Covington NV, Duff MC. Heterogeneity Is a Hallmark of Traumatic Brain Injury, Not a Limitation: A New Perspective on Study Design in Rehabilitation Research. American Journal of Speech-Language Pathology. 2021 2;30(2S):974-85. Available from: https://pubs.asha.org/doi/abs/10.1044/2020_AJSLP-20-00081.

[12] Collins FS, Varmus H. A New Initiative on Precision Medicine. New England Journal of Medicine. 2015 2;372(9):793-5. Available from: https://www.nejm.org/doi/full/10.1056/NEJMp1500523.

[13] König IR, Fuchs O, Hansen G, von Mutius E, Kopp MV. What is precision medicine? European Respiratory Journal. 2017 10;50(4).

[14] Schiele F. Quality of data in observational studies: separating the wheat from the chaff. European Heart Journal-Quality of Care and Clinical Outcomes. 2017;3:99-100. Available from: https://academic.oup.com/ehjqcco/article/3/2/99/3039238.

[15] Kilkenny MF, Robinson KM. Data quality: "Garbage in – garbage out". Health Information Management Journal. 2018 9;47(3):103-5. Available from: http://www.who.int/healthinfo/.

[16] Maceachern SJ, Forkert ND. Machine learning for precision medicine. Genome. 2021;64(4):416-25. Available from: https://cdnsciencepub.com/doi/10.1139/gen-2020-0131.

[17] Pencina MJ, Peterson ED. Moving From Clinical Trials to Precision Medicine: The Role for Predictive Modeling. JAMA. 2016 4;315(16):1713-4. Available from: https://jamanetwork.com/journals/jama/fullarticle/2516691.

[18] Marcano-Cedeño A, Chausa P, García A, Cáceres C, Tormos JM, Gómez EJ. Data mining applied to the cognitive rehabilitation of patients with acquired brain injury. Expert Systems with Applications. 2013 3;40(4):1054-60.

[19] Lee YH, Bang H, Kim DJ. How to Establish Clinical Prediction Models. Endocrinology and Metabolism. 2016 3;31(1):38. Available from: /pmc/articles/PMC4803559//pmc/articles/PMC4803559/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC4803559/.

[20] Kappen TH, Van Loon K, Kappen MAM, Van Wolfswinkel L, Vergouwe Y, Van Klei WA, et al. Barriers and facilitators perceived by physicians when using prediction models in practice. Journal of Clinical Epidemiology. 2016 2;70:136-45. Available from: http://www.jclinepi.com/article/S0895435615004278/fulltexthttp://www.jclinepi.com/article/S0895435615004278/abstracthttps://www.jclinepi.com/article/S0895-4356(15)00427-8/abstract.

[21] Yang HS. Machine Learning for Sepsis Prediction: Prospects and Challenges. Clinical Chemistry. 2024 3;70(3):465-7. Available from: https://dx.doi.org/10.1093/clinchem/hvae006.

[22] Boussina A, Shashikumar SP, Malhotra A, Owens RL, El-Kareh R, Longhurst CA, et al. Impact of a deep learning sepsis prediction model on quality of care and survival. npj Digital Medicine 2024 7:1. 2024 1;7(1):1-9. Available from: https://www.nature.com/articles/s41746-023-00986-6.

[23] Bauer MS, Kirchner JA. Implementation science: What is it and why should I care? Psychiatry research. 2020 1;283. Available from: https://pubmed.ncbi.nlm.nih.gov/31036287/.

[24] Chen JH, Asch SM. Machine Learning and Prediction in Medicine — Beyond the Peak of Inflated Expectations. New England Journal of Medicine. 2017 6;376(26):2507-9. Available from: https://www.nejm.org/doi/abs/10.1056/NEJMp1702071.

[25] Seneviratne MG, Shah NH, Chu L. Bridging the implementation gap of machine learning in healthcare. BMJ Innovations. 2020 4;6(2):45-7. Available from: https://innovations.bmj.com/content/6/2/45https://innovations.bmj.com/content/6/2/45.abstract.

[26] Sendak M, Gao M, Nichols M, Lin A, Balu S. Machine Learning in Health Care: A Critical Appraisal of Challenges and Opportunities. eGEMs. 2019 1;7(1):1. Available from: /pmc/articles/PMC6354017//pmc/articles/PMC6354017/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC6354017/.

[27] Arshi B, Smits LJ, Wynants L, Cowley LE, Reeve K, Rijnhart E. Number of publications on new clinical prediction models: a systematic literature search. 2023 7. Available from: https://osf.io/4txc6.

[28] Damschroder LJ, Reardon CM, Widerquist MAO, Lowery J. The updated Consolidated Framework for Implementation Research based on user feedback. Implementation Science 2022 17:1. 2022 10;17(1):1-16. Available from: https://implementationscience.biomedcentral.com/articles/10.1186/s13012-022-01245-0.

[29] Proctor EK, Powell BJ, McMillen JC. Implementation strategies: Recommendations for specifying and reporting. Implementation Science. 2013 12;8(1):1-11. Available from: https://implementationscience.biomedcentral.com/articles/10.1186/1748-5908-8-139.

[30] Chan SL, Lee JW, Ong MEH, Siddiqui FJ, Graves N, Ho AFW, et al. Implementation of Prediction Models in the Emergency Department from an Implementation Science Perspective-Determinants, Outcomes, and Real-World Impact: A Scoping Review. Ann Emerg Med. 2023 3;82(1):22-36. Available from: http://dx.doi.org/10.1016/j.annemergmed.2023.02.001http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?holding=inleurlib_fft&cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=36925394https://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med23&AN=36925394.

[31] Baldwin H, Loebel-Davidsohn L, Oliver D, Salazar de Pablo G, Stahl D, Riper H, et al. Real-World Implementation of Precision Psychiatry: A Systematic Review of Barriers and Facilitators. Brain sciences. 2022 7;12(7). Available from: https://pubmed.ncbi.nlm.nih.gov/35884740/.

[32] Kaul V, Enslin S, Gross SA. History of artificial intelligence in medicine. Gastrointestinal endoscopy. 2020 10;92(4):807-12. Available from: https://pubmed.ncbi.nlm.nih.gov/32565184/.

[33] Chowdhury MZI, Turin TC. Precision health through prediction modelling: factors to consider before implementing a prediction model in clinical practice. J Prim Health Care. 2020 3;12(1):3-9. Available from: http://dx.doi.org/10.1071/hc19087http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?holding=inleurlib_fft&cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=32223844https://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med17&AN=32223844.

[34] Dowding D, Russell D, McDonald MV, Trifilio M, Song J, Brickner C, et al. "A catalyst for action": Factors for implementing clinical risk prediction models of infection in home care settings. J Am Med Inform Assoc. 2021 3;28(2):334-41. Available from: http://dx.doi.org/10.1093/jamia/ocaa267http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?holding=inleurlib_fft&cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=33260204https://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med19&AN=33260204.

[35] Watson J, Hutyra CA, Clancy SM, Chandiramani A, Bedoya A, Ilangovan K, et al. Overcoming barriers to the adoption and implementation of predictive modeling and machine learning in clinical care: what can we learn from US academic medical centers? JAMIA Open. 2020 7;3(2):167-72. Available from: https://dx.doi.org/10.1093/jamiaopen/ooz046.

[36] Wachtler C, Coe A, Davidson S, Fletcher S, Mendoza A, Sterling L, et al. Development of a Mobile Clinical Prediction Tool to Estimate Future Depression Severity and Guide Treatment in Primary Care: User-Centered Design. JMIR mHealth and uHealth. 2018 4;6(4). Available from: https://pubmed.ncbi.nlm.nih.gov/29685864/.

[37] Bentley KH, Zuromski KL, Fortgang RG, Madsen EM, Kessler D, Lee H, et al. Implementing Machine Learning Models for Suicide Risk Prediction in Clinical Practice: Focus Group Study With Hospital Providers. JMIR formative research. 2022 3;6(3). Available from: https://pubmed.ncbi.nlm.nih.gov/35275075/.

[38] van Oort L, Verhagen A, Koes B, de Vet R, Anema H, Heymans M. Evaluation of the usefulness of 2 prediction models of clinical prediction models in physical therapy: a qualitative process evaluation. J Manipulative Physiol Ther. 2014 3;37(5):334-41. Available from: http://dx.doi.org/10.1016/j.jmpt.2013.09.008http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?holding=inleurlib_fft&cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=24928642https://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med11&AN=24928642.

[39] Yarborough BJH, Stumbo SP, Schneider J, Richards JE, Hooker SA, Rossom R. Clinical implementation of suicide risk prediction models in healthcare: a qualitative study. BMC Psychiatry. 2022 3;22(1):789. Available from: http://dx.doi.org/10.1186/s12888-022-04400-5http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?holding=inleurlib_fft&cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=36517785https://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med22&AN=36517785.

[40] Sandhu S, Lin AL, Brajer N, Sperling J, Ratliff W, Bedoya AD, et al. Integrating a Machine Learning System Into Clinical Workflows: Qualitative Study. Journal of medical Internet research. 2020 11;22(11). Available from: https://pubmed.ncbi.nlm.nih.gov/33211015/.

[41] Yarborough BJH, Stumbo SP. A Stakeholder-Informed Ethical Framework to Guide Implementation of Suicide Risk Prediction Models Derived from Electronic Health Records. ARCH SUICIDE RES. 2023 3;27(2):704-17. Available from: http://dx.doi.org/10.1080/13811118.2022.2064255http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?holding=inleurlib_fft&cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=35446244https://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med23&AN=35446244.

[42] Dorajoo SR, Chan A. Implementing Clinical Prediction Models: Pushing the Needle Towards Precision Pharmacotherapy. Clin Pharmacol Ther. 2018 3;103(2):180-3. Available from: http://dx.doi.org/10.1002/cpt.752http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?holding=inleurlib_fft&cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=28722146https://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med15&AN=28722146.

[43] Baker T, Gerdin M. The clinical usefulness of prognostic prediction models in critical illness. EUR J INTERN MED. 2017 3;45:37-40. Available from: http://dx.doi.org/10.1016/j.ejim.2017.09.012http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?holding=inleurlib_fft&cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=28935477https://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med14&AN=28935477.

[44] Cao J, Singh K. Integrating risk prediction models into chronic kidney disease care. Curr Opin Nephrol Hypertens. 2020 3;29(3):339-45. Available from: http://dx.doi.org/10.1097/mnh.0000000000000603http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?holding=inleurlib_fft&cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=32205582https://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med17&AN=32205582.

[45] Park MS, Weissman SM, Postula KJV, Williams CS, Mauer CB, O'Neill SM. Utilization of breast cancer risk prediction models by cancer genetic counselors in clinical practice predominantly in the United States. J Genet Couns. 2021 3;30(6):1737-47. Available from: http://dx.doi.org/10.1002/jgc4.1442http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?holding=inleurlib_fft&cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=34076301https://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med20&AN=34076301.

[46] Grindell C, Coates E, Croot L, O'Cathain A. The use of co-production, co-design and co-creation to mobilise knowledge in the management of health conditions: a systematic review. BMC Health Services Research. 2022 12;22(1):1-26. Available from: https://bmchealthservres.biomedcentral.com/articles/10.1186/s12913-022-08079-y.

[47] Bird M, McGillion M, Chambers EM, Dix J, Fajardo CJ, Gilmour M, et al. A generative co-design framework for healthcare innovation: development and application of an end-user engagement framework. Research Involvement and Engagement. 2021 12;7(1):1-12. Available from: https://researchinvolvement.biomedcentral.com/articles/10.1186/s40900-021-00252-7http://creativecommons.org/publicdomain/zero/1.0/.

[48] Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly: Management Information Systems. 1989;13(3):319-39.

[49] Engelhardt EG, Pieterse AH, Van Duijn-Bakker N, Kroep JR, De Haes HCJM, Smets EMA, et al. Breast cancer specialists' views on and use of risk prediction models in clinical practice: A mixed methods approach. Acta Oncologica. 2015 3;54(3):361-7. Available from: https://www.tandfonline.com/action/journalInformation?journalCode=ionc20.

[50] Goldstein BA, Navar AM, Pencina MJ, Ioannidis JPA. Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. Journal of the American Medical Informatics Association : JAMIA. 2017 1;24(1):198. Available from: /pmc/articles/PMC5201180//pmc/articles/PMC5201180/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC5201180/.

[51] Kline-Simon AH, Sterling S, Young-Wolff K, Simon G, Lu Y, Does M, et al. Estimates of Workload Associated With Suicide Risk Alerts After Implementation of Risk-Prediction Model. JAMA Network Open. 2020 10;3(10):e2021189-9. Available from: https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2771929.

[52] Steyerberg EW, Uno H, Ioannidis JPA, van Calster B, Ukaegbu C, Dhingra T, et al. Poor performance of clinical prediction models: the harm of commonly applied methods. Journal of Clinical Epidemiology. 2018 6;98:133-43.

[53] Ancker JS, Edwards A, Nosal S, Hauser D, Mauer E, Kaushal R. Effects of workload, work complexity, and repeated alerts on alert fatigue in a clinical decision support system. BMC medical informatics and decision making. 2017 4;17(1). Available from: https://pubmed.ncbi.nlm.nih.gov/28395667/.

[54] Mohammed NI, Jarde A, Mackenzie G, D'Alessandro U, Jeffries D. Deploying Machine Learning Models Using Progressive Web Applications: Implementation Using a Neural Network Prediction Model for Pneumonia Related Child Mortality in The Gambia. Frontiers in public health. 2022 2;9. Available from: https://pubmed.ncbi.nlm.nih.gov/35252109/.

[55] Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) the TRIPOD statement. Circulation. 2015;131(2):211-9. Available from: https://www.ahajournals.org/doi/10.1161/CIRCULATIONAHA.114.014508.

[56] Binuya MAE, Engelhardt EG, Schats W, Schmidt MK, Steyerberg EW. Methodological guidance for the evaluation and updating of clinical prediction models: a systematic review. BMC Medical Research Methodology 2022 22:1. 2022 12;22(1):1-14. Available from: https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-022-01801-8.

Appendix A. Literature Search

Appendix A.1. Search strings

Medline

((prediction*-model*).ti.) AND ((implement* OR clinical-pract* OR integrat* OR adopt* OR usefulness* OR evaluat*).ti.) NOT (news OR congres* OR abstract* OR book* OR chapter* OR dissertation abstract*).pt. NOT (exp animals/ NOT humans/) AND 2008:2030.(sa_year)

Embase

('prediction model'/exp/mj OR (prediction*-model*):ti) AND ('implementation'/de/mj OR 'clinical practice'/exp/mj OR (implement* OR clinical-pract* OR integrat* OR adopt* OR usefulness* OR evaluat*):ti) NOT ([Conference Abstract]/lim OR [preprint]/lim) NOT ((animal/exp OR animal*:de OR nonhuman/de) NOT ('human'/exp)) AND ([2008-2030]/py)

Appendix A.2. Results literature search

Table A.3: Literature search

Database	**Platform**	**Years of coverage**	**Records**
Medline	Ovid	1946 - Present	498
Embase	Embase.com	1971 - Present	481
Total			979
Total after duplicates removed			538

Appendix B. Included Articles

Author	Year	Country	Study Design	Setting	Care facilty	Model Type
Baker et al.	2017	Sweden	Opinion article	Multicentre	Combination	Multiple models
Bentley et al.	2022	United States	Focus Group Study	Monocentre	Non-academic hospital	ML model (unspecified)
Cao et al.	2020	United States	Review	Monocentre	Academic hospital	Multiple models
Chowdhury et al.	2020	Canada	Opinion articles	Monocentre	Academic hospital	MVR model
Dorajoo et al.	2018	Singapore	Opinion article	Monocentre	Academic hospital	Multiple models
Dowding et al.	2021	United Kingdom	(Semi-)structured interviews	Monocentre	Non-hospital care	MVR model
Engelhardt et al.	2015	the Netherlands	Mixed methods	Multicentre	Combination	MVR model
Fujimori et al.	2022	Japan	Mixed methods	Multicentre	Non-academic hospital	XGBoost model
Ho et al.	2023	United States	(Semi-)structured interviews	Monocentre	Academic hospital	ML model (unspecified)
Kappen et al.	2016	the Netherlands	Mixed methods	Monocentre	Academic hospital	MVR model
Park et al.	2021	United States	Survey study	Multicentre	Combination	Multiple models
Reger et al.	2019	United States	Case example	Monocentre	Non-academic hospital	MVR model
Sandhu et al.	2020	United States	(Semi-)structured interviews	Monocentre	Academic hospital	DL model
van Oort et al.	2014	the Netherlands	Mixed methods	Multicentre	Non-hospital care	MVR model
Wachtler et al.	2018	Australia	Survey	Multicentre	Non-hospital care	Multiple models
Watson et al.	2019	United States	(Semi-)structured interviews	Multicentre	Academic hospital	ML model (unspecified)
Yarborough et al.	2023	United States	Mixed methods	Multicentre	Non-hospital care	Multiple models
Yarborough et al.	2022	United States	Interviews	Multicentre	Non-hospital care	Multiple models

ML, Machine Learning; *MVR*, Multivariable regression; *DL*, Deep Learning

Table B.4: Characteristics of the included studies.

Appendix C. Checklist

Implementation Clinical Prediction Model - Checklist

1. Project Definition and Planning

1.1 Needs Assessment

- ○ 1.1.1 Conduct a thorough needs assessment to identify the clinical problem and ensure the CPM fits the local context and workflow.
- ○ 1.1.2 Define specific needs, possibilities, and restrictions collaboratively with clinicians, model developers, and institutional bodies throughout the entire implementation process.
- ○ 1.1.3 Determine if the system should be in the EHR or available as web-app
- ○ 1.1.4 Evaluate access carefully for security and ethical reasons to ensure safe data use while maximizing potential care improvements.
- ○ 1.1.5 Limit input parameters to those that are clinically relevant, statistically significant, and easy and quick to retrieve.
- ○ 1.1.6 Ensure output parameters are actionable, and consider their consequences within the given context.

1.2 Team

- ○ 1.2.1 Establish a committee to oversee the process, plan, secure milestones, and ensure stakeholder consensus. Assign identifiable leads and establish clear endpoints in the planning process.
- ○ 1.2.2 Involve an expert to drive progress and manage expectations by clearly communicating goals.

1.3 Resources

- ○ 1.3.1 Systematically collect and evaluate EHR data to ensure suitability for model incorporation.
- ○ 1.3.2 Identify and allocate personnel with the right expertise and secure sufficient funding to support all phases, including technological and personnel costs.

2. Model Development

2.1 Model

- ○ 2.1.1 Ensure the model achieves sufficient performance metrics and outperforms clinical judgment.
- ○ 2.1.2 Comply with medical device regulations to meet legal and safety standards.
- ○ 2.1.3 Develop the model to operate efficiently under the time pressures faced by users.

2.2 Design

- ○ 2.2.1 Make included parameters and their significance available for comparison with clinical judgment.
- ○ 2.2.2 Ensure accessibility by establishing compatibility with current systems and convenient access to the model in the clinical context.
- ○ 2.2.3 Manage the balance between necessary notifications and alert fatigue to maintain effectiveness.
- ○ 2.2.4 Ensure user-friendliness with intuitive interfaces and clear data depiction for quick interpretation and transparency.

3. Training and Deployment

3.1 Education

- ○ 3.1.1 Educate clinicians to understand the model's logic, enhancing acceptance and perceived need.
- ○ 3.1.2 Ensure end-users are well-trained before and during implementation, building necessary capabilities and confidence.

3.2 Workflow

- ○ 3.2.1 Establish a standardized workflow and proper documentation to ensure ethical use and compliance with regulatory standards.
- ○ 3.2.2 Guarantee CPMs are used for their intended purpose and in an ethical manner.

4. Evaluation and updating

4.1 Performance and Impact

- ○ 4.1.1 Conduct continuous evaluation, updating, and maintenance to ensure the model remains relevant and effective.
- ○ 4.1.2 Perform continuous impact measurement to assess whether the CPM improves clinical care and process metrics.

4.2 Information

- ○ 4.2.1 Ensure information on the model is always available for reference.
- ○ 4.2.2 Provide refresher courses to maintain user competency and confidence.

daan theeuwes centrum
voor intensieve neuroreanimatie

Appendix B. Data Characteristics

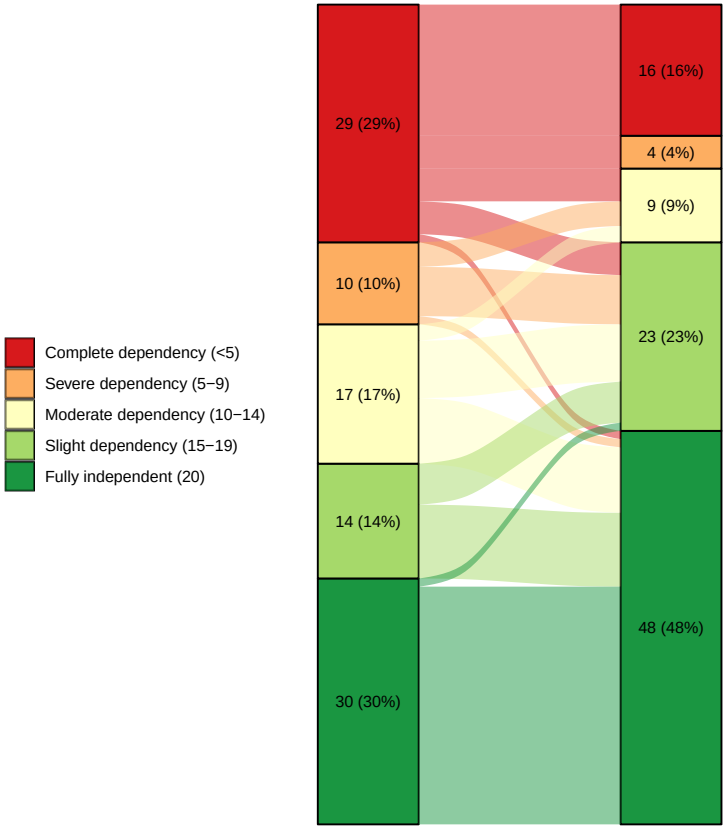
Table B.1: Characteristics of the development and validation cohorts including p-values and effect sizes.

	Development Cohort			Validation Cohort			Comparison		
	Frequency (%), Mean (SD)[1] or Median (IQR)[2]	**Range**	**Data Completeness (%)**	**Frequency (%), Mean (SD)[1] or Median (IQR)[2]**	**Range**	**Data Completeness (%)**	**p-values (t-test[1] or Wilcoxon[2] or Fisher's Exact [3] or Chi square [4])**	**Adjusted p-values (Benjamini-Hochberg)**	**Effect Size (Cohen's d[1] or Rank Biserial[2])**
Demographic Characteristics and Medical Background									
Sex (Female)	38.0%	-	100%	33.3%	-	100%	0.878[4]	1.0	-
Age at Injury Onset (Years)	24.4 (5.6)[1]	15.5 - 35.2	100%	24.6 (5.9)[1]	15.3 - 34.8	100%	0.512[1]	1.0	**−0.16[1]**
Age at Rehabilitation Admission (Years)	24.7 (4.5)[1]	16.0 - 35.4	100%	25.0 (5.8)[1]	16.5 - 35.0	100%	0.499[1]	1.0	**−0.16[1]**
Time since Injury (Days)	71 (24 – 206.5)[2]	6 - 2069	100%	59 (25 – 165)[2]	8 - 1114	100%	0.774[2]	1.0	0.04[2]
Left-handedness	12.1%	-	83%	14.3%	-	100%	0.723[4]	1.0	-
Immigration Background	22.9%	-	83%	38.1%	-	100%	0.254[4]	1.0	-
Has Children	7.2%	-	83%	9.5%	-	100%	0.661[3]	1.0	-
Participation in Sports	67.1%	-	82%	66.7%	-	100%	1.0[4]	1.0	-
Presence of Learning Disorders	21.5%	-	79%	15.0%	-	95%		1.0	-
Presence of Neurological Disorders	7.8%	-	77%	0.0%	-	95%	0.340[3]	1.0	-
Presence of Psychiatric Disorders	22.1%	-	77%	15.0%	-	95%	0.757[3]	1.0	-
Presence of Other Disorders	3.9%	-	77%	5.0%	-	95%	1.0[3]	1.0	-
Current Educational Enrollment	54.2%	-	83%	52.4%	-	100%		1.0	-
ABI Etiology									
Traumatic Brain Injury (TBI)	62.0%	-	100%	71.4%	-	100%	0.571[4]	1.0	-
- Skull Fractures	69.2%	-	100%	60.0%	-	100%	1.0[4]	1.0	-
- Cranial Nerve Injuries	15.4%	-	100%	33.3%	-	100%	0.120[4]	1.0	-
- Intracranial Injuries	100%	-	100%	100%	-	100%	0.571[4]	1.0	-
Cerebrovascular Accident (CVA)	24.0%	-	100%	19.1%	-	100%	0.780[3]	1.0	-
- Subarachnoid Hemorrhage	11.1%	-	100%	0.0%	-	100%	1.0[3]	1.0	-
- Intracerebral Hemorrhage	38.9%	-	100%	24.9%	-	100%	0.689[3]	1.0	-
- Subdural Hemorrhage	5.5%	-	100%	0.0%	-	100%	1.0[3]	1.0	-
- Extra-/Epidural Hemorrhage	0.0%	-	100%	0.0%	-	100%	1.0[3]	1.0	-
- Cerebral Infarction	33.3%	-	100%	74.8%	-	100%	0.375[3]	1.0	-
- Occlusion of Pre-cerebral Arteries	5.5%	-	100%	24.9%	-	100%	0.318[3]	1.0	-
- Other Cerebrovascular Diseases	17.6%	-	100%	24.9%	-	100%	1.0[3]	1.0	-
Infectious Causes	6.0%	-	100%	0.0%	-	100%	0.604[3]	1.0	-

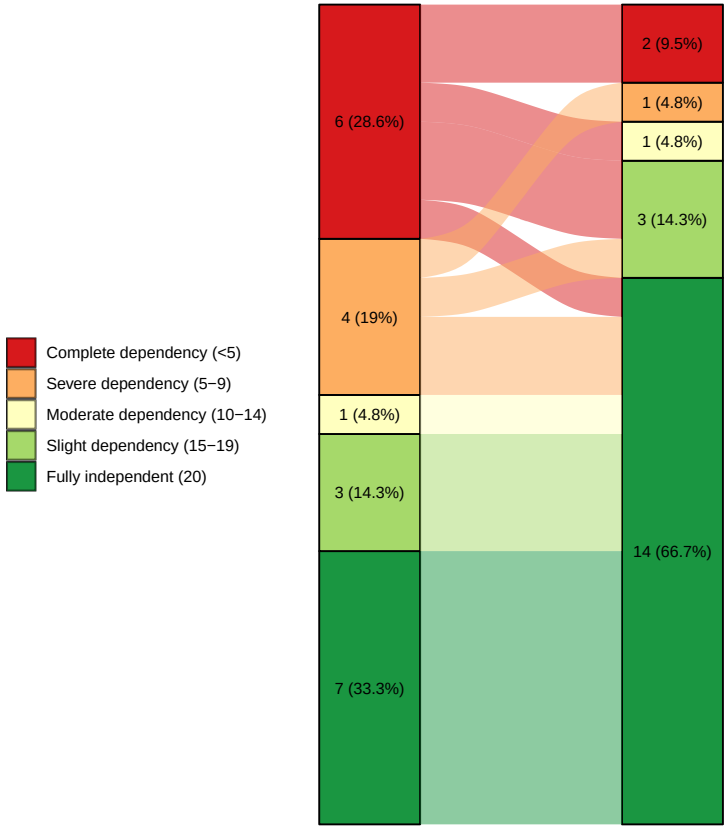
Continued on next page

	Development Cohort			Validation Cohort			Comparison		
	Frequency (%), Mean (SD)[1] or Median (IQR)[2]	**Range**	**Data Completeness (%)**	**Frequency (%), Mean (SD)[1] or Median (IQR)[2]**	**Range**	**Data Completeness (%)**	**p-values (t-test[1] or Wilcoxon[2] or Fisher's Exact [3] or Chi square [4])**	**Adjusted p-values (Benjamini-Hochberg)**	**Effect Size (Cohen's d[1] or Rank Biserial[2])**
Hydrocephalus	0.0%	-	100%	4.8%	-	100%	0.174[3]	1.0	-
Toxic Encephalopathy	1.2%	-	100%	9.5%	-	100%	0.0775[3]	1.0	-
Other Brain Diseases	8.4%	-	100%	4.8%	-	100%	1.0[3]	1.0	-
Trauma Centre Data									
Hospital Length of Stay	35 (16-61)[2]	4 - 120	95%	35 (24–66)[2]	8 - 99	100%	-	1.0	**0.15[2]**
EMV score (total) at hospital admission	3 (3-5)[2]	3 - 15	68%	5 (3–9)[2]	3 - 15	90%	-	1.0	**0.36[2]**
EMV score (total) at hospital discharge	14 (11-15)[2]	6 - 15	85%	15 (14–15)[2]	11 - 15	71%	-	1.0	**0.61[2]**
Abnormal pupil size at admission	52.4%	-	59%	64.7%	-	81%	0.541[4]	1.0	-
CT Marshall category	2 (2-5)[2]	1 - 6	65%	2 (2–5)[2]	2 - 6	95%	-	1.0	**0.26[2]**
Neurosurgery	62.4%	-	93%	52.4%	-	100%	0.550[4]	1.0	-
- Hematoma Relieve	40.9%	-	93%	28.6%	-	100%	0.426[4]	1.0	-
- Craniectomy	34.4%	-	93%	28.6%	-	100%	0.798[4]	1.0	-
- Craniotomy	16.1%	-	93%	0.0%	-	100%	0.069[3]	1.0	-
- Cranioplasty	29.0%	-	93%	19.1%	-	100%	0.426[3]	1.0	-
- Placement ICP	-	-	-	38.1%	-	100%	0.724[4]	1.0	-
- Placement of External Drain	17.2%	-	93%	14.3%	-	100%	-	1.0	-
- Placement of VPD	8.6%	-	93%	0.0%	-	100%	-	1.0	-
- Placement of LPD	4.3%	-	93%	0.0%	-	100%	-	1.0	-
- Third Ventriculostomy	0.0%	-	93%	0.0%	-	100%	-	1.0	-
- Surgical treatment of CSF leakage	0.0%	-	93%	4.8%	-	100%	-	1.0	-
- Other Neurosurgery	9.68%	-	93%	4.8%	-	100%	0.686[3]	1.0	-
Discharge to an intermediate care facility	20.6%	-	97%	14.3%	-	100%	0.762[3]	1.0	-
Discharge to a rehabilitation centre	43.3%	-	97%	85.7%	-	100%	<0.001[3]	**0.033**	-
PTA at admission	63.0%	-	81%	76.5%	-	81%	0.278[3]	1.0	-
Outcome Data									
Barthel Index on admission	11.4 (7.4)[1]	0 - 20	100%	11.6 (8.0)[1]	0 - 20	100%	0.936[1]	1.0	−0.02[1]
Barthel Index at 3 months	15.9 (5.5)[1]	0 - 20	100%	16.8 (6.0)[1]	1 - 20	100%	0.301[1]	1.0	**−0.25[1]**

SD: Standard Deviation, *IQR*: Interquartile Range, *ABI*: Acquired Brain Injury, *TBI*: Traumatic Brain Injury, *CVA*: Cerebrovascular Accident, *EMV*: Eye, Motor, Verbal score, *PTA*: Post-Traumatic Amnesia, *CT*: Computed Tomography, *VPD*: Ventriculoperitoneal Drain, *LPD*: Lumboperitoneal Drain, *CSF*: Cerebrospinal Fluid, *BI*: Barthel Index.



(a) Development Cohort (n=100)



(b) External Validation Cohort (n=21)

Figure B.1: Level of Independence Trajectories in Barthel Index (BI).

Appendix C. Correlation

Table C.1: **Correlation coefficients and p-values between predictors and absolute error in the "Level of Independence at Admission" model.**

	Correlation Coefficient	**P-value**
PTA at Rehabilitation Admission	**0.539**	**0.012**
Education Level of Parents	0.313	0.167
Hospital Length of Stay	0.203	0.378
Sex (Female)	0.140	0.544
Intracerebral Hemorrhage (CVA)	0.088	0.706
Discharge to an Intermediate Care Facility	0.076	0.743
Cranioplasty	−0.019	0.935
Discharge to a Rehabilitation Centre	−0.076	0.743
Epidural bleeding on CT scan	−0.080	0.730
Has Children	−0.257	0.260

Abbreviations: *PTA*: Post Traumatic Amnesia, , *CVA*: cerebrovascular Accident.
Correlation is calculated between each predictor and the prediction error (absolute difference between observed and predicted Barthel Index scores). Correlation Coefficients >0.5 and with a p- value <0.05 are depicted in bold.

Table C.2: **Correlation coefficients and p-values between predictors and absolute error in the "Level of Independence at Three Months Post-Admission" Model.**

	Correlation Coefficient	**P-value**
Hospital Length of Stay	**0.753**	$<$ **0.0001**
PTA at Rehabilitation Admission	0.434	0.049
Age at Admission	0.284	0.212
Neurosurgery Performed	0.111	0.631
Epidural bleeding on CT scan	−0.075	0.745
Skull Fracture(s)	−0.181	0.433
Discharge to a Rehabilitation Centre	−0.197	0.391
Participation in Sports	−0.199	0.388

Abbreviations: *PTA*, Post Traumatic Amnesia.
Correlation is calculated between each predictor and the prediction error (absolute difference between observed and predicted Barthel Index scores). Correlation Coefficients >0.5 and with a p- value <0.05 are depicted in bold.

Table C.3: **Correlation coefficients and p-values between predictors and absolute error in the "Change in Level of Independence over Three Months" Model.**

	Correlation Coefficient	**P-value**
PTA at Rehabilitation Admission	0.363	0.203
Subarachnoid Bleeding (TBI)	−0.011	0.970
Placement of ICP monitor	−0.018	0.952
Intraventricular or Subarachnoid Bleeding on CT scan	−0.099	0.736
Left Handedness	−0.284	0.325
Participation in Sports	−0.468	0.091
Barthel Index at Rehabilitation Admission	**−0.557**	**0.039**
Focal Injury (TBI)	**−0.579**	**0.030**

Abbreviations: *PTA*, Post Traumatic Amnesia; *TBI*, Traumatic Brain Injury; *ICP*, Intracranial Pressure.
Correlation is calculated between each predictor and the prediction error (absolute difference between observed and predicted Barthel Index scores). Correlation Coefficients >0.5 and with a p- value <0.05 are depicted in bold.

Appendix D. Tool Visualisations

In this Appendix, the visualisations of the tool can be found. For each of the three pages, important parts of the tool are highlighted.

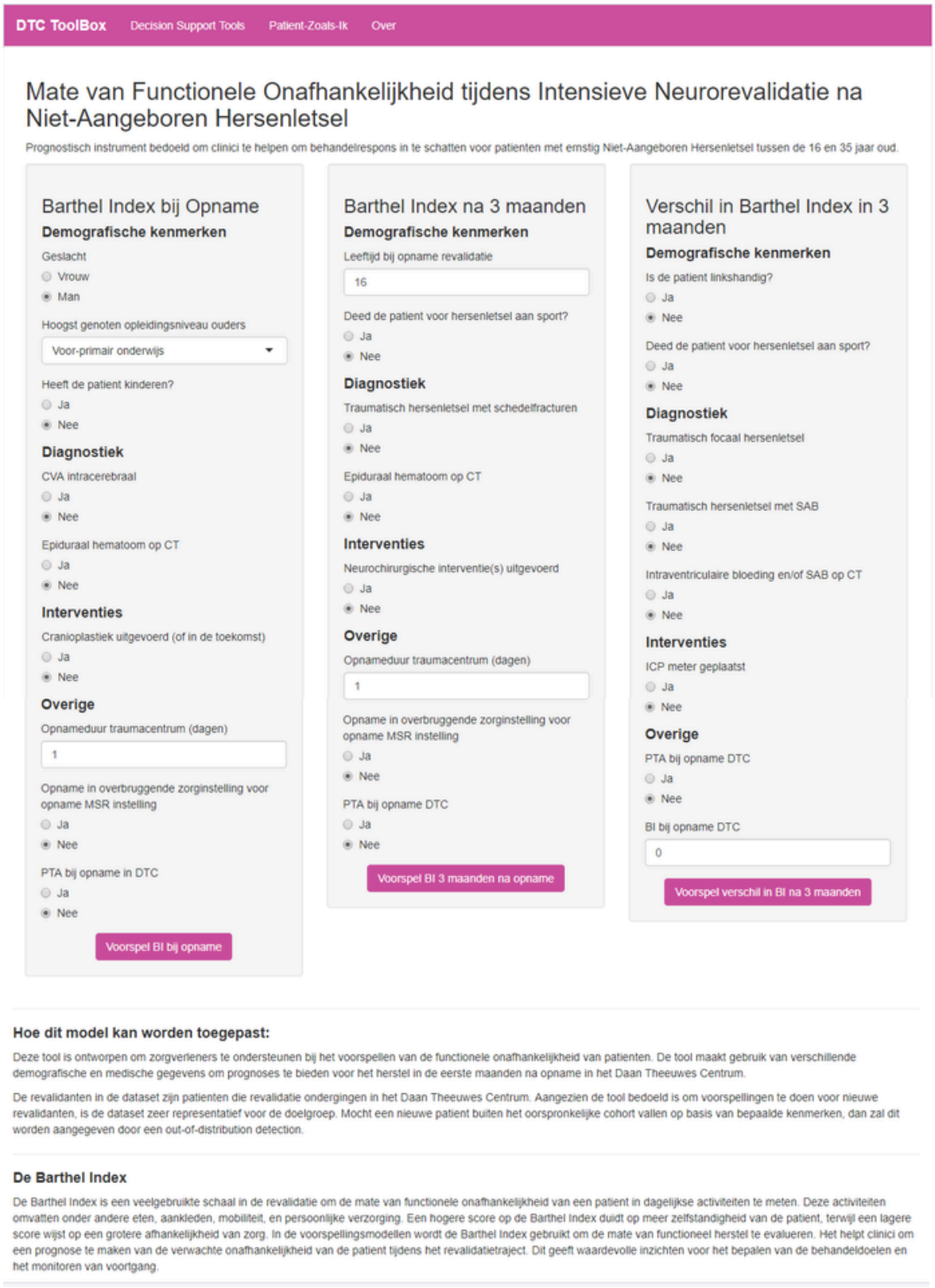


Figure D.1: The "Decision Support Tools" page requires clinician's input and uses prediction models to provide predictions of the Barthel Index (BI) for young adults with acquired brain injury (ABI).

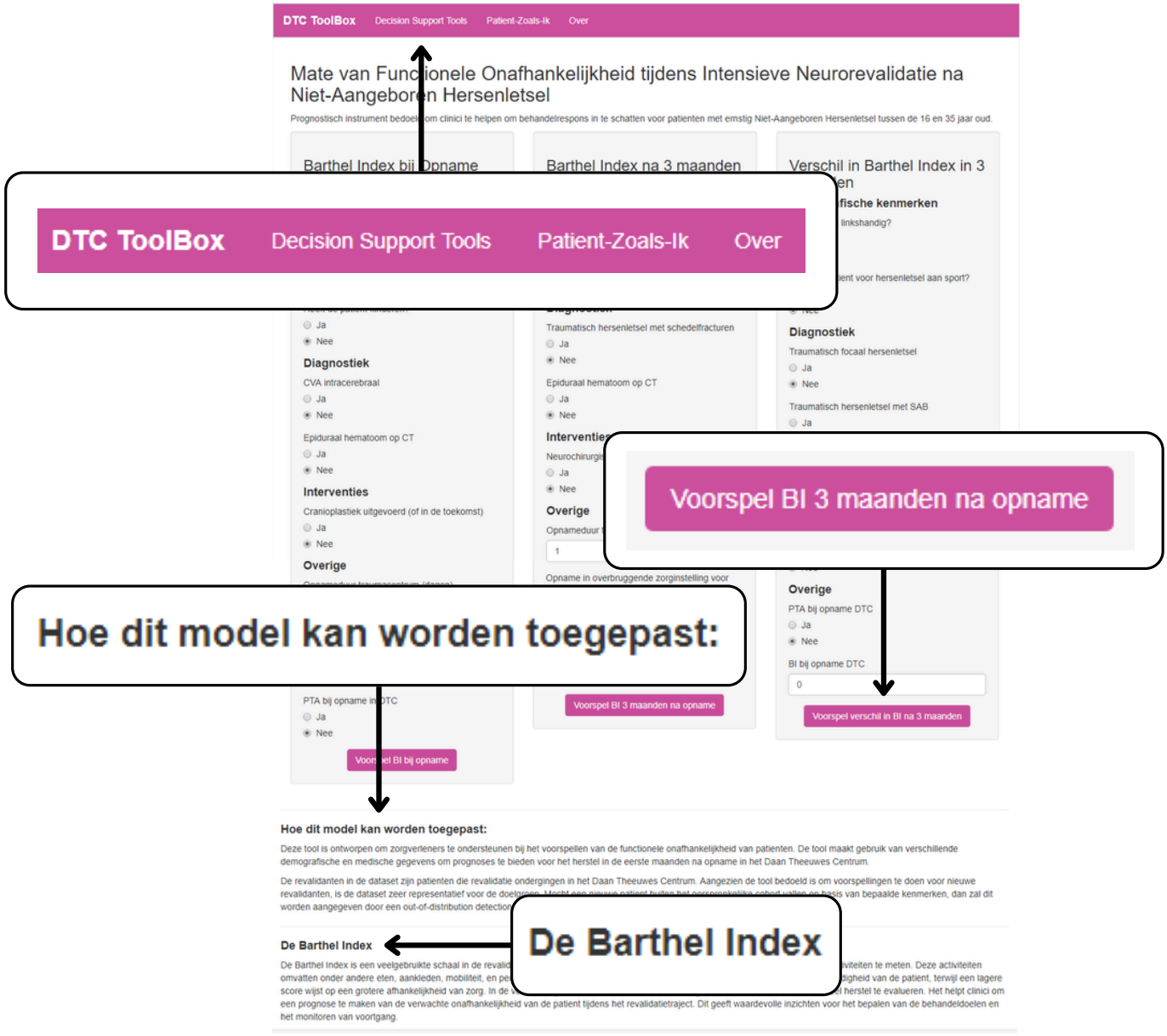


Figure D.2: The ToolBox features a navigation bar with all included pages. On the "Decision Support Tools" page, after input is provided for each model, users can click buttons to generate Barthel Index (BI) predictions. Below the prediction models, additional information is provided on how to use the models and details about the Barthel Index as the outcome measure.

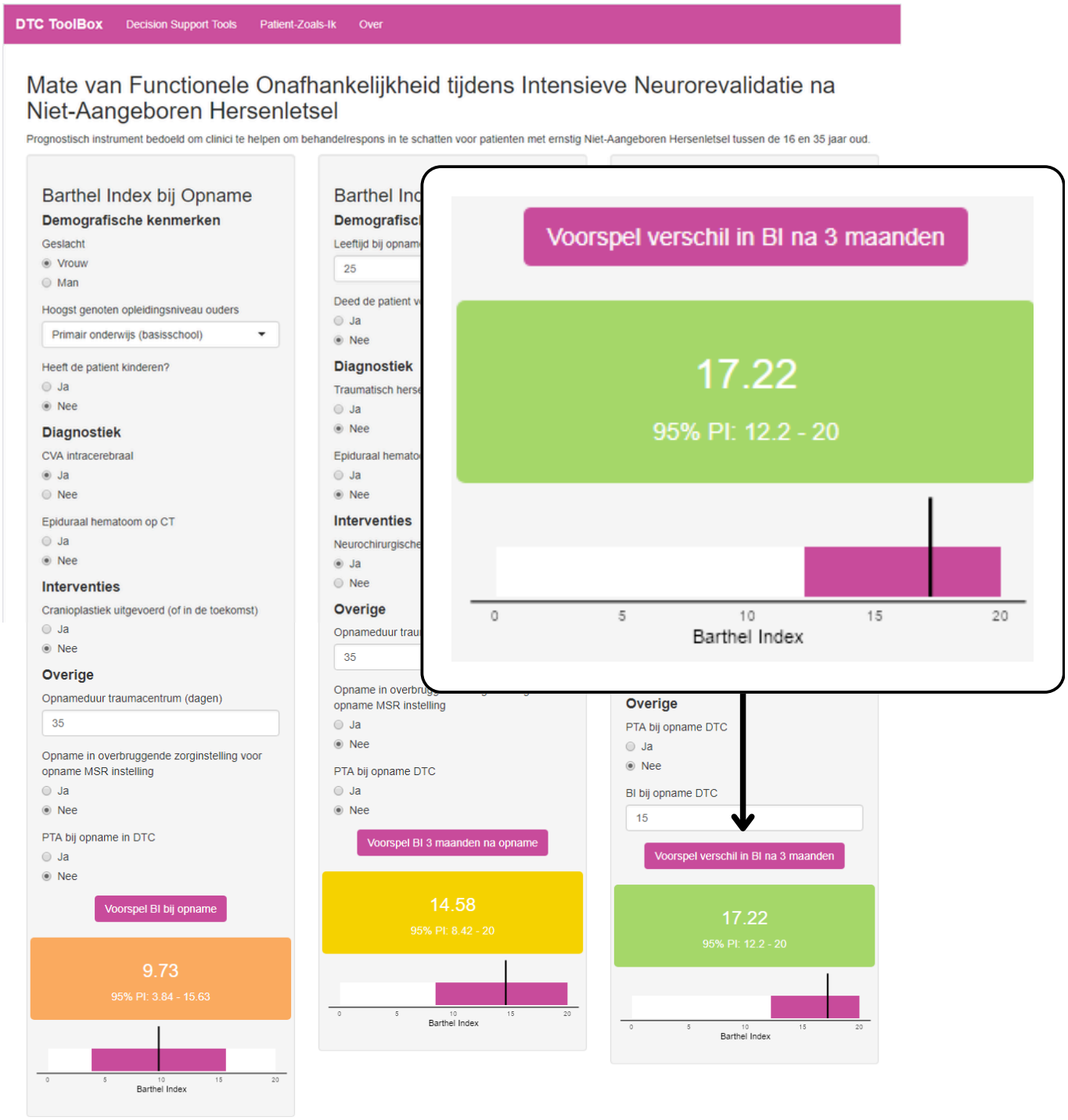


Figure D.3: On the "Decision Support Tools" page, after input is provided for each model and the predictions are made, the models provide the Barthel Index (BI), a 95% prediction interval (PI) and a graph depicting the outcomes.

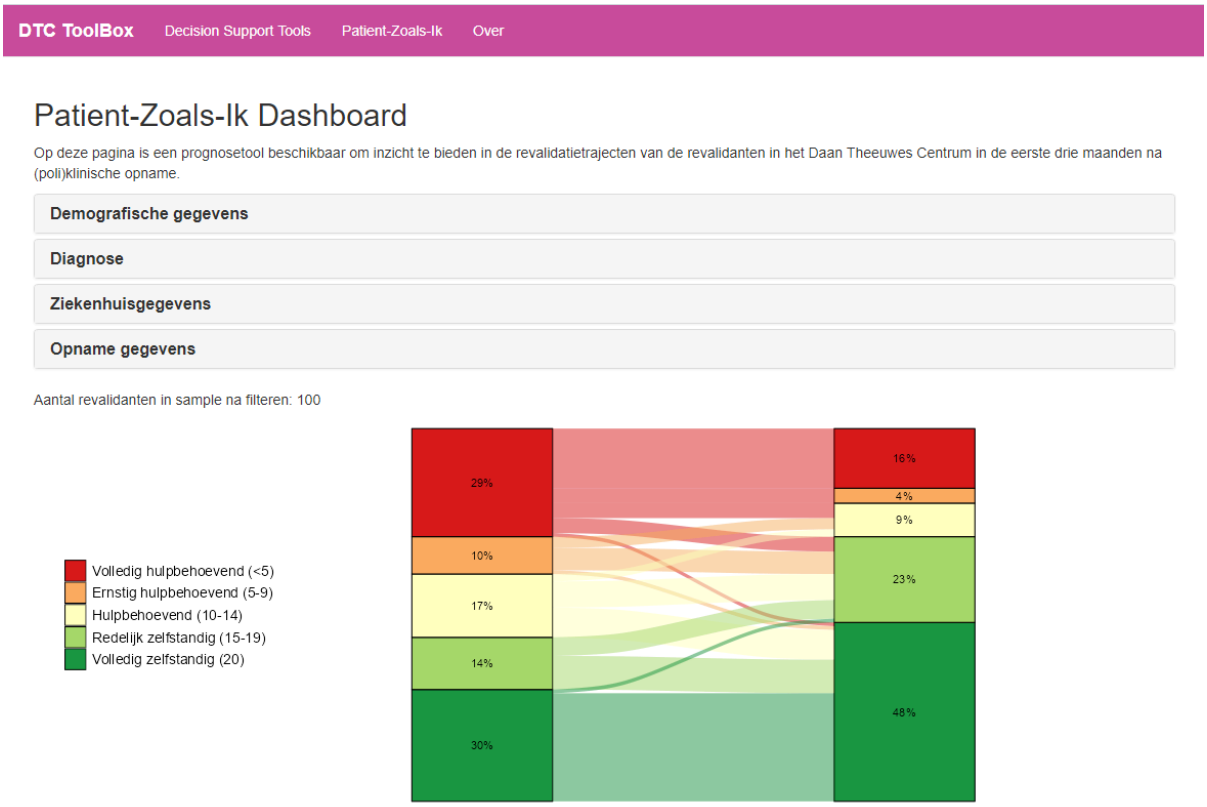


Figure D.4: The "Patient-Like-Me" page depicts the trajectories of 100 included patients in development between admission and three months Post-Admission.

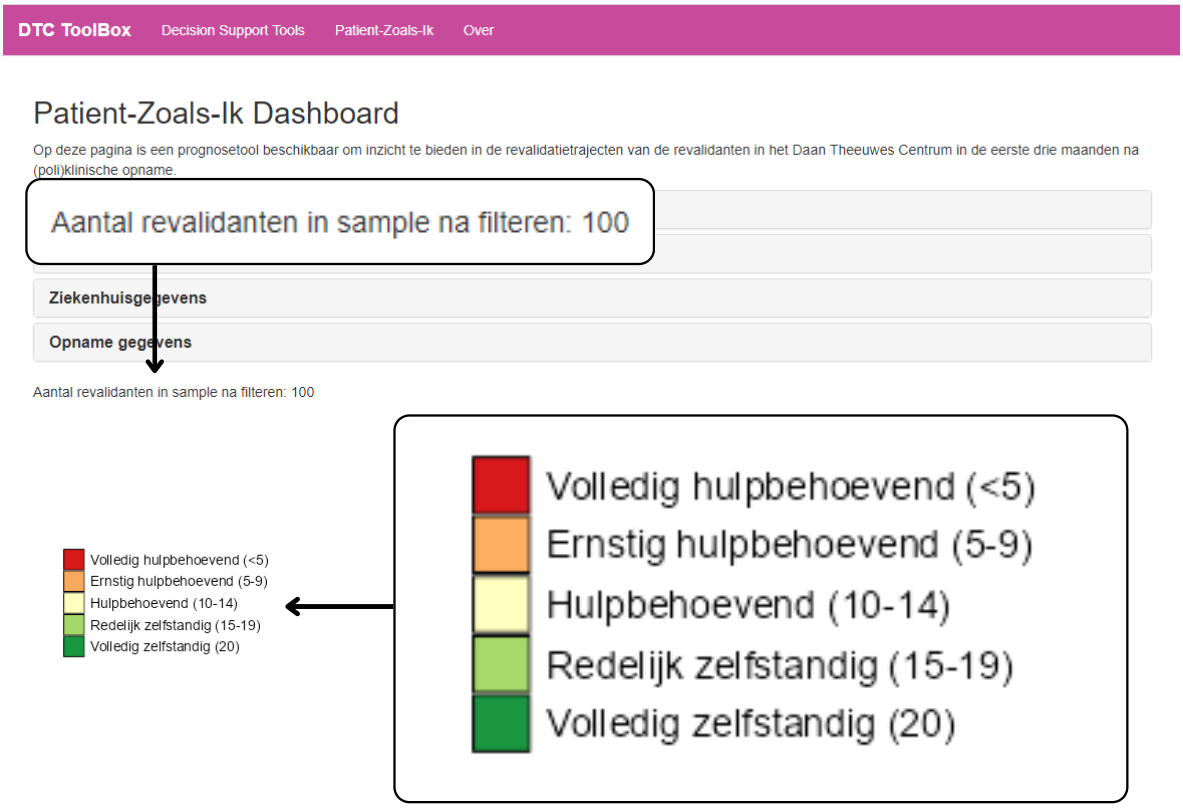


Figure D.5: In the "Patient-Like-Me" page the Barthel Index (BI) is divided in 5 categories: "Fully dependent", "Severely dependent", "Requires assistance", "Moderately to highly independent", "Fully independent". To provide additional information on the sample size, the "Patient-Like-Me" page provides information on how many patients are left in the sample after filtering.

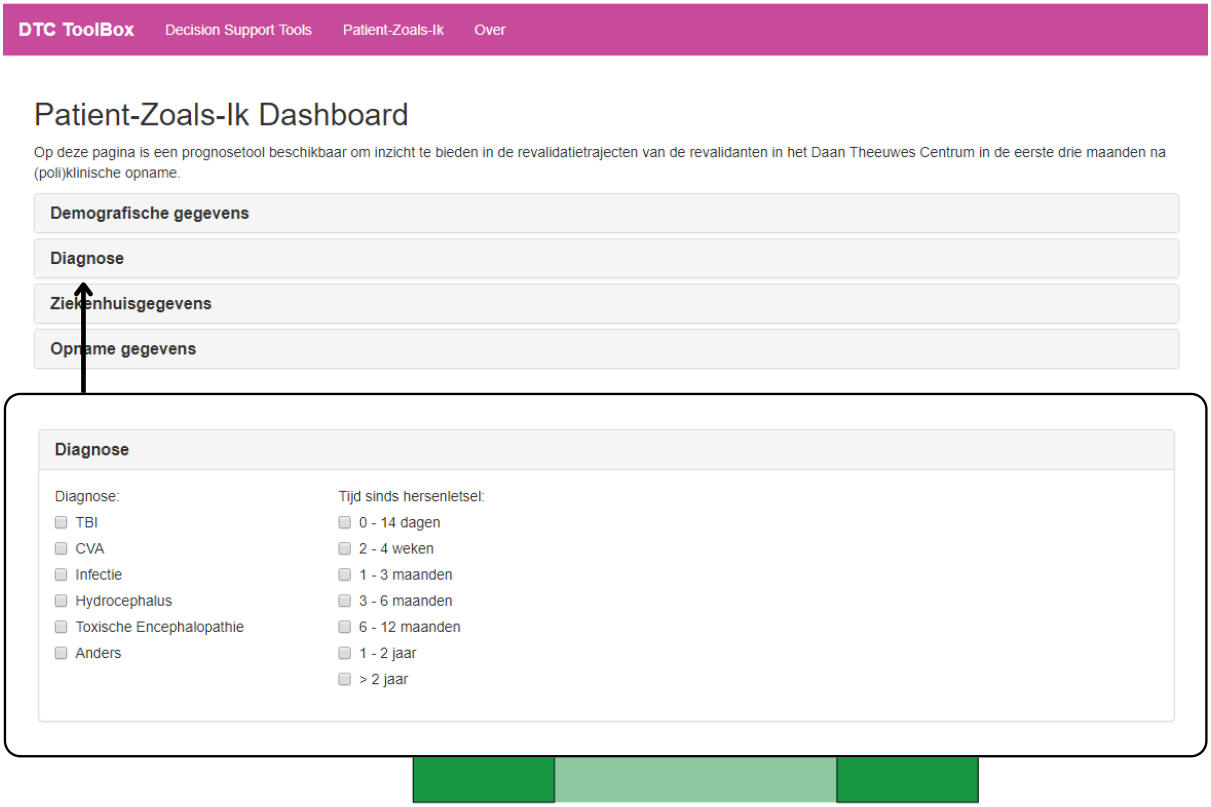


Figure D.6: The "Patient-Like-Me" page incorporates filters that allow users to adapt the graph to reflect patients with specific characteristics, such as demographic characteristics, diagnose, hospital-related characteristics and rehabilitation admission information.

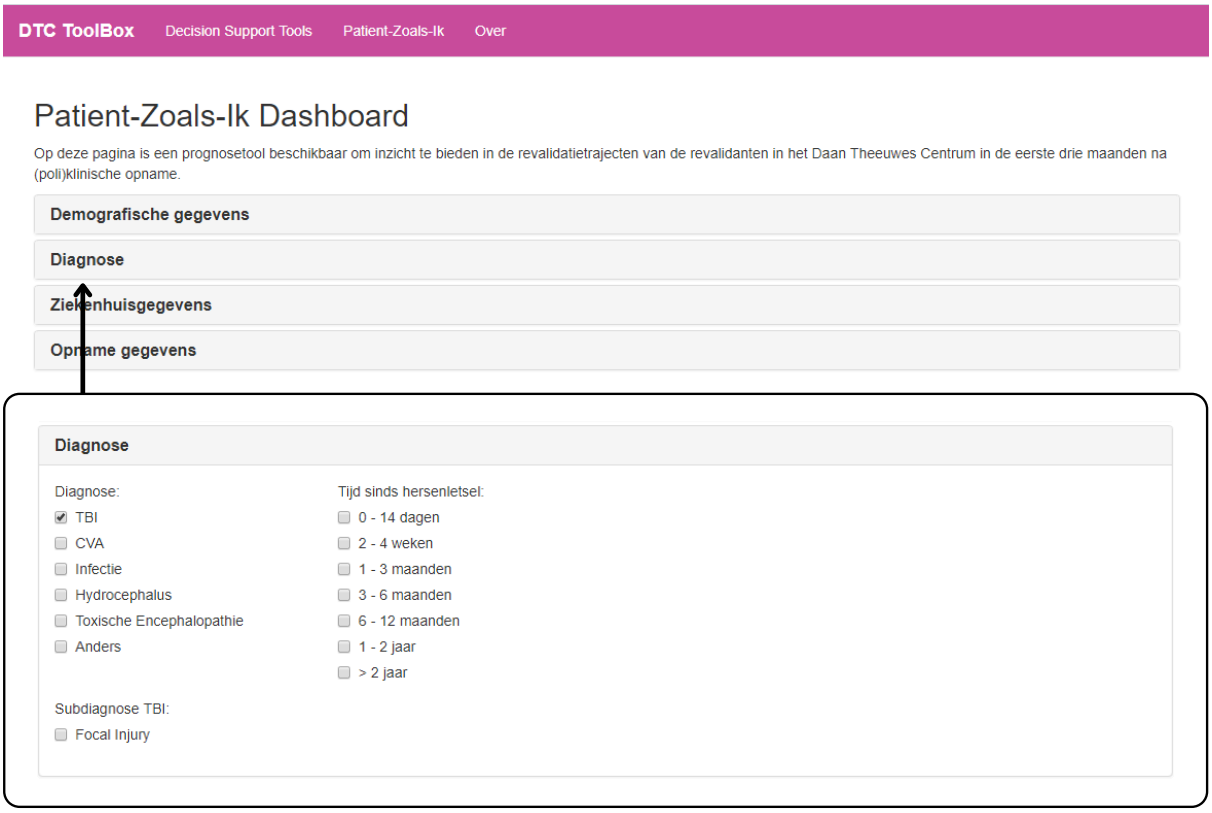


Figure D.7: Some filters in the "Patient-Like-Me" page incorporate subcategories, such as "Focal Injury" in "Diagnosis".

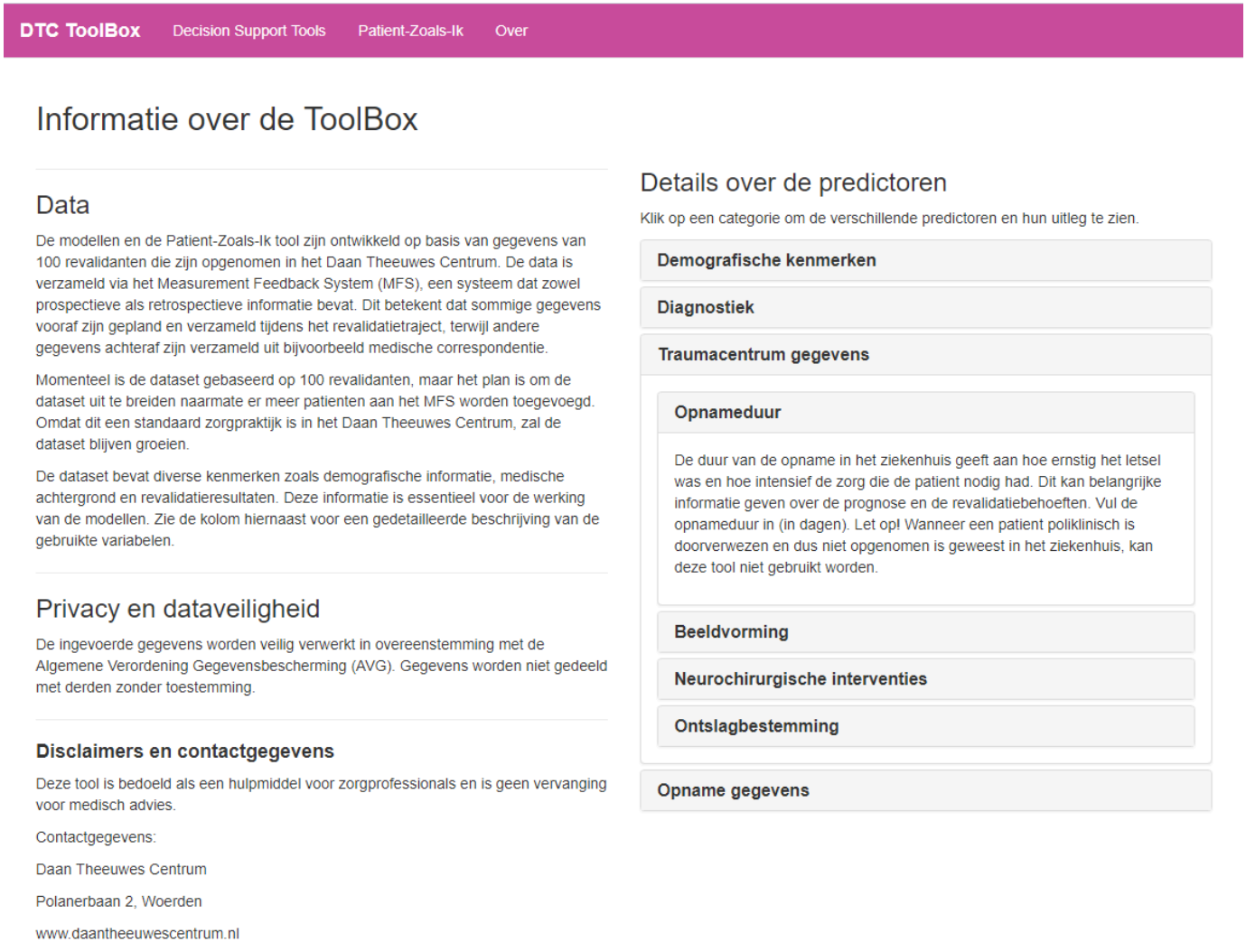


Figure D.8: The “About” page was added to provide more information on the models and the predictors. It provides detailed information on the predictors, including their definitions, how they relate to the outcome measure, and guidance on how to input the data correctly.

Appendix E. Model Updating

Table E.1: Updated prediction models for the level of independence for young adults with acquired brain injury.

	Predictors		Coefficient (β)	Standard Error	t-value	p-value
Level of Independence at Admission (n = 121)	PTA at Rehabilitation Admission		−8.1425	1.0135	−8.034	**<0.001**
	Discharge to Intermediate Care Facility		−5.8974	1.4182	−4.159	**<0.001**
	Nerve Injury		4.7550	1.6148	2.945	**0.004**
	Epidural bleeding on CT scan		3.9125	1.3907	2.813	**0.006**
	Neurosurgery Performed		−2.9793	1.1094	−2.686	**0.008**
	Education Level of Parents		0.8965	0.3688	2.431	**0.017**
	Left Handedness		−3.2003	1.5104	−2.119	**0.036**
	Presence of Psychological Disorder		−2.4484	1.2246	−1.999	**0.048**
	Intracerebral Bleeding (CVA)		−3.4146	1.7521	−1.949	0.054
	Sex (Female)		−1.6936	1.0646	−1.591	0.115
	R^2	0.562				
	RMSE	5.156				
	MAE	4.068				
Level of Independence at Three Months Post-Admission (n = 121)	Hospital Length of Stay		−0.10956	0.01855	−5.906	**<0.001**
	Discharge to Intermediate Care Facility		−4.46025	1.23171	−3.621	**<0.001**
	Epidural bleeding on CT scan		4.10369	1.41553	2.899	**0.005**
	PTA at Rehabilitation Admission		−2.24337	0.96780	−2.318	**0.022**
	Education Level of Parents		0.66931	0.32732	2.045	**0.043**
	Participation in Sports		1.90983	0.95119	2.008	**0.047**
	Skull Fracture(s)		1.62774	0.94716	1.719	0.089
	Neurosurgery Performed		−1.95323	1.19147	−1.639	0.104
	Hematoma Relieve		−2.11758	1.42427	−1.487	0.140
	Other Neurosurgery Performed		2.59607	1.83543	1.414	0.160
	R^2	0.567				
	RMSE	4.500				
	MAE	3.256				
Change in Level of Independence over Three Months (n = 121)	Barthel Index at Rehabilitation Admission		0.72746	0.08668	8.393	**<0.001**
	Hospital Length of Stay		−0.05024	0.02065	−2.433	**0.017**
	Discharge to Intermediate Care Facility		−3.01021	1.31403	−2.291	**0.025**
	Participation in Sports		2.19143	0.97484	2.248	**0.028**
	Neurosurgery Performed		−2.13257	1.09224	−1.952	0.055
	Presence of Psychological Disorder		1.93577	1.15651	1.674	0.098
	Axonal Intracranial Injury (TBI)		1.73275	1.04616	1.656	0.102
	Edema (TBI)		2.10809	1.37305	1.535	0.129
	R^2	0.727				
	RMSE	3.885				
	MAE	2.877				

Abbreviations: R^2, Coefficient of Determination; *RMSE*, Root Mean Square Error; *MAE*, Mean Absolute Error; *PTA*, Post-Traumatic Amnesia; *CVA*, Cerebrovascular Accident; *TBI*, Traumatic Brain Injury.
Significant predictors are depicted in bold.