A multicenter prospective study validated a nomogram to predict individual risk of dependence in ambulation after rehabilitation

(Article begins on next page)

16 July 2024

A multicenter prospective study validated a nomogram to predict individual risk of dependence in ambulation after rehabilitation

Bruno Bernardini 1, Luigi Baratto 2, Costanza Pizzi 3, Annibale Biggeri 4, Giovanna Cerina 5, Viviana Colantonio 5, Carla Corsini 5, Sara Ghirmai 5, Marco Pagani 5, Stefania Fracchia 6, Marisa Gardella 2, Dolores Catelan 7, Maria Luisa Malosio 8, Elisa Malagamba

A multicenter prospective study validated a nomogram to predict individual risk of dependence in ambulation after rehabilitation

Bruno Bernardini 1, Luigi Baratto 2, Costanza Pizzi 3, Annibale Biggeri 4, Giovanna Cerina 5, Viviana Colantonio 5, Carla Corsini 5, Sara Ghirmai 5, Marco Pagani 5, Stefania Fracchia 6, Marisa Gardella 2, Dolores Catelan 7, Maria Luisa Malosio 8, Elisa Malagamba 9

Affiliations
•IRCCS Humanitas Research Hospital, Neurocenter-Neurorehabilitation Unit, Rozzano, Milan, Italy. Electronic address: bruno.bernardini@humanitas.it.
•Department of Rehabilitation, La Colletta Hospital, Arenzano, Italy.
•Department of Medical Sciences, University of Turin, Turin, Italy.
•Department of Statistics, Computer Science, Applications "G. Parenti", University of Florence (FI), Florence, Italy; Unit of Biostatistics, Epidemiology and Public Health, Department of Cardiac, Thoracic, Vascular Sciences and Public Health, University of Padua (PD), Padua, Italy.
•IRCCS Humanitas Research Hospital, Neurocenter-Neurorehabilitation Unit, Rozzano, Milan, Italy.
•Geriatric Internal Medicine Unit, Garbagnate Hospital, Garbagnate Milanese, Milan, Italy.
•Unit of Biostatistics, Epidemiology and Public Health, Department of Cardiac, Thoracic, Vascular Sciences and Public Health, University of Padua (PD), Padua, Italy.
•Institute of Neuroscience, National Research Council (CNR), Milan, Italy; Laboratory of Pharmacology and Pathology of the Nervous System, IRCCS Humanitas Research Hospital, Rozzano, Milan, Italy.
•Department of Health Services of Liguria Region, Genoa, Italy.

Abstract

Objectives

To develop the Functional Risk Index for Dependence in Ambulation (FRIDA) score, a nomogram to predict individual risk of dependence in ambulation at discharge from postacute rehabilitation and validate its performance temporally and spatially.

Study Design and Setting

We analyzed the database of a multicenter prospective observational quality cohort study conducted from January 2012 to March 2016, including data from 8,796 consecutive inpatients who underwent rehabilitation after stroke, hip fracture, lower limb joint replacement, debility, and other neurologic, orthopedic, or miscellaneous conditions.

Results

A total of 3,026 patients (34.4%) were discharged dependent in ambulation. In the training set of 5,162 patients (58.7%), Lasso-regression selected advanced age, premorbid disability, and eight indicators of medical and functional adverse syndromes at baseline to establish the FRIDA score. At the temporal validation obtained on an external set of 3,234 patients (41.3%), meta-analyses showed that the FRIDA score had good and homogeneous discrimination (summary area under the curve 0.841, 95% confidence interval = 0.826–0.855, I_2 = 0.00%) combined with accurate calibration (summary Log O/E ratio 0.017, 95% confidence interval −0.155 to 0.190). These performances remained stable at spatial validation obtained on 3,626 patients, with substantial heterogeneity of estimates across nine facilities. Decision curve analyses showed that a FRIDA score–supported strategy far outperformed the usual "treat all" approach in each impairment categories.

Conclusion

The FRIDA score is a new clinically useful tool to predict an individual risk for dependence in ambulation at rehabilitation discharge in many different disabilities, and may also reflect well the case-mix composition of the rehabilitation facilities.

1. Introduction

Many of the patients admitted to rehabilitation are frail elderly with chronic multimorbidities and disabilities pre-existing the acute event. Depending on the impact of acute illness, complex clinical phenotypes are generated in the postacute period that can be difficult to disentangle and predict in their health trajectories.

Complexity in rehabilitation is a well-known issue but it remains essentially unresolved due to the heterogeneity of reference models [[1], [2], [3]] and paucity of validated measures. An instrument, known as the Rehabilitation Complexity Scale, was proposed by Turner-Stokes et al. in 2010 [4] and subsequently validated limited to psychometric properties [5,6]. Some concerns about the independence of the scale [7] and the lack of comparison on outcomes still do not allow its use as a prognostic tool. Several prognostic indices can be retrieved from the geriatric literature [[8], [9], [10]]. In general, these tools have focused on mortality or surrogate endpoints of little meaning to the patient, such as length of stay or institutionalization. When transferred to the clinical arena, they further lose value because of the lack of any link between patient assessment and treatment plan.

We believe that understanding and measuring complexity in rehabilitation is a major challenge for improving postacute care but we still need generalizable tools focused on meaningful patient outcomes and embedded in routine practice to achieve useful guidance for treatments. In this prognostic study, we chose dependence in ambulation as the outcome and handled indicators of the complexity of patients' history and medical and

functional adverse syndromes [11] at baseline to model the likelihood of its occurrence. The indicators come from the Indicators for Performance Evaluation in Rehabilitation, version 2.0 (IPER-2.0) system, a multidimensional core-set consistent with a comprehensive geriatric assessment [12,13] tailored for quality improvement in care processes and outcomes in rehabilitation [14].

Our aim is to develop and externally validate a new prognostic score that we named Functional Risk Index to predict Dependence in Ambulation (FRIDA), a nomogram to estimate individual risk of dependence in ambulation at discharge from rehabilitation. Because our sample consisted of subgroups of patients with different baseline risk and from different rehabilitation sites, we accounted for variability in its performance both across common impairments and across facilities in validating the FRIDA score.

2. Methods
2.1. Study design, setting, and participants
We used the IPER-2.0 Rehabilitation Quality Improvement Study dataset, which includes status-process-outcomes indicators and functional and quality-of-life measures routinely collected during a multicenter, prospective, observational cohort study. The IPER-2.0 study was led by the Health Agency of the Liguria Region with the endorsement of the Italian Society of Physical Medicine and Rehabilitation from January 2012 to March 2016, initially enrolling seven intensive rehabilitation facilities in Liguria, to which another four intensive and extensive rehabilitation centers in Northern and Central Italy were added on a voluntary basis. All facilities were accredited by the National Health System with the same structural standards by level of rehabilitative intensiveness. Special Rehabilitation Units, such as spinal, cardiorespiratory or severe brain injury units were not included in the IPER-2.0 study.

We considered all 8,796 consecutive patients aged 18 years and more admitted for rehabilitation included in the database, without exclusion criteria. We controlled for case-mix composition by classifying the main disabling condition as per rehabilitation impairment categories (RICs) system [15]. Grouping by RIC was useful for tracing multiple diagnoses back to unitary functional aggregates already known to have different prognoses, thus adjusting for baseline risk in our sample. For convenience of analysis, we collapsed the original 20 RIC categories to seven, namely stroke, hip fracture, lower limb joint replacement, debility, other neurological, orthopedic, or miscellaneous conditions (Supplementary Table 1).

2.2. Ethical approval
The IPER-2.0 study was conducted as per the National Code of Ethics and Good Practice (G.U. 72, March 26, 2012) that complies with the requirements of the EU General Data Protection Regulation 2016/679. As per practice, all patients signed consent for their data to be used for statistical analysis and research purposes. Because data from Ligurian centers were routinely due to the Regional Health Agency, an approval from local ethics committees to participate in the IPER-2.0 study was obtained only for non-Ligurian centers. The Humanitas Research Hospital Independent Ethics Committee approved this analysis (No. 1150, 2020), which was conducted on fully deidentified data.

2.3. Data collection
All baseline characteristics were collected within 24 hours of patient admission as per a multidisciplinary approach that always involved the physician, nurse, physiotherapist, and other rehabilitation professionals (e.g., speech therapist and neuropsychologist) as needed. To optimize data management, the Liguria Regional Health Agency developed a

web platform where the manager of each center uploaded all data in a deidentified mode at patient discharge. Each center could access the web platform limited to its own data.

2.4. Outcome

The outcome was dependence in ambulation at discharge (DAD) from rehabilitation, identified by dichotomizing the ambulation subcore of the modified Barthel Index, using the value of 8 as the threshold. The modified Barthel Index is a valid and reliable tool for measuring dependence in basic activities of daily living [16]. The ambulation subscore is a 5-level categorical scale ranging from 0 (unable to ambulate) to 15 (completely independent in ambulation for at least 50 m). Patients identified as "dependent" were unable to walk or needed the assistance of a person (subscore 0 to 8). "Independent" patients were able to walk independently for less (subscore 12) or more (subscore 15) than 50 m.

For all participating individuals, the outcome was standardly assessed face-to-face by the physiotherapist within 2 days of planned discharge. Both participants and assessors were unaware of the FRIDA score. Patients transferred to acute care hospital units or who died during their stay were counted as dependent in the ambulation.

2.5. Candidate predictors

We considered age, sex, and IPER-2.0 indicators profiling patient complexity for history (severe organ failure, dementia, chronic multimorbidity, cancer, social frailty, and premorbid disability) and the burden of care at baseline for adverse medical syndromes (reduced vigilance, delirium, medical instability, infection, depression, pain, dysphagia, malnutrition, pressure sores, urinary catheter or incontinence, and tracheostomy) and functional adverse syndromes (communicative disability, dependence in eating and in six key tasks of basic mobility) as potential predictors of DAD.

The IPER-2.0 indicators are all binary, focused on the presence of a target condition identified by clear clinical elements, or anchored by validated scales, or driven by laboratory parameters. The indicators are listed in Table 1 and the rationale and standards for their collection are provided in Appendix. Patient's premorbid disability has been classified by the modified Rankin Scale (mRS), a six-level score ranging from 0 (no symptoms) to 5 (severe disability) [17].

2.6. Statistical methods
2.6.1. General and descriptive statistics

To develop and validate our model, we used cross-validation and bootstrap methods to improve the generalizability of the estimates and meta-analysis to account for variability among subgroups [18,19]. In reporting the results, we followed the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis guidelines [20].

Because all items considered in this study were required to successfully complete the online data entry, no missing data were found. We divided the entire dataset into a training set of 5,162 patients (58.7%) discharged from January 2012 to May 2014 and a validation set consisting of 3,234 patients (41.3%) discharged from June 2014 to the end of the IPER-2.0 study in March 2016. Temporal partitioning is the preferred approach to achieve external validation of a prognostic model [21].

At baseline, categorical variables were presented as frequency and percentage (%) and compared with the chi-squared test or Fisher's exact test, whereas continuous variables were summarized by the median with interquartile range (IQR) and compared with the

Mann–Whitney U test. The bivariate association between candidate predictors was calculated using the Goodman–Kruskall gamma statistic. In the case of binary variables, the gamma test reduces to Yules' Q, which is a function of the odds ratio.

2.6.2. Coding of candidate predictors

- Patients' age was categorized into seven classes of years, namely 18–64, 65–69, 70–74, 75–79, 80–84, 85–89, and 90+.

- The premorbid mRS was rescaled into four categories by collapsing scores 0–1 and 4–5.

- Considering the subset of 12 indicators of medical complexity (IMCs) at baseline as a kind of "active multimorbidity," we obtained a scale from their count in two steps [22]. First, we performed a joint multiple correspondence analysis, removing the pain and depression indicators because of their low impact on the overall variance (Supplementary Figure 1). Second, assuming an equivalent prognostic value of the remaining IMCs, we summed and rescaled them to a maximum value of 5 based on the frequency distribution. The resulting scale ranged from 0 (no IMCs) to "5 or more," a value that includes five to nine possible IMCs. Pain and depression were introduced as individual covariates during modeling.

The indicators of adverse functional syndromes were included in the modeling as individual covariates. We did not group these indicators into a single scale to allow for greater degrees of freedom in profiling functional impairments of different nature and form. At the end of recoding, we obtained for modeling 19 binary indicators of the initial 28 and three categorical variables (i.e., age, mRS score, and IMCs count).

2.6.3. FRIDA score construction and internal validation
The construction of the FRIDA score and the assessment of its performance in the training set were accomplished in four steps:
1.
Variable selection of the prognostic model for DAD by fitting all potential predictors using Lasso (least absolute shrinkage and selection operator) logistic regression with 10-fold cross-validation [23].

2.
Internal model validation by cluster logistic bootstrapping (1,000 replications) on data clustered by RICs, rehabilitation centers, patient provenance (hospital wards vs. other provenance), and 4-month time periods. As per this procedure, bootstrap resampling was performed on jackknife estimates from each leave-one-cluster-out, thus generating more robust, bias-corrected confidence intervals.

3.
Checking for multicollinearity and statistical interactions and determining the FRIDA score as a nomogram using Stata's nomolog package [24]. A nomogram transfers the mathematical function of a model into a diagram, making a scoring system more accurate than the usual simplified metrics.

4.

Refitting the FRIDA score to assess the overall discrimination and calibration performance by the area under the receiver operating characteristic curve (AUC) and calibration plot [25], respectively.

2.6.4. External validation

External validation was achieved by transferring the estimated probability of DAD in the training set to the validation set, evaluating both RIC-specific (temporal validity) and facility-specific (spatial validity) performance. Discrimination was assessed by the c-statistic, reporting the AUC calculated by DeLong's method. Calibration was assessed by calibration plots and associated statistics.

To substantiate the stability and reproducibility of FRIDA score performance, we performed pooled meta-analyses of point estimates of AUCs and observed/expected (O/E) ratios with their respective 95% confidence intervals (CIs), extracted separately from each RIC and facility [26]. The O/E ratio is the ratio of total observed-to-expected DAD cases, with a value of 1 indicating perfect mean calibration and higher or lower values indicating underprediction or overprediction, respectively [27].

All meta-analyses were conducted as per a random effects model with Der Simonian-Laird inverse variance weighting. We reported the effect size with 95% CI and an approximate 95% prediction interval (PI). Heterogeneity was assessed by the I2 statistic and examined by sensitivity analysis for values more than 50%.

2.6.5. Clinical utility

The clinical utility of the FRIDA score was evaluated in the different RICs by decision curve analysis (DCA). The DCA quantifies in terms of net benefit the clinical impact of one or more diagnostic-therapeutic approaches compared with the default alternatives "treat all" and "treat none" [28]. Net benefit is calculated by subtracting for each risk threshold the proportion of false positives from the proportion of true positives weighted, when it matter, by the consequences of undertreatment or overtreatment [29]. In conducting the DCAs, we assumed that false-positive and false-negative decisions were of equal importance.

All analyses were conducted using Stata/SE, version 17.0 (StataCorp LLC, College Station, Texas). Reported P values are two-sided and statistical significance was set at P < 0.05.

3. Results

3.1. Descriptive statistic and bivariate analyses

We analyzed 8,796 adult patients, 37.9% men and 62.1% women, with a median age of 76 (IQR, 68–82) years, who underwent rehabilitation for orthopedic (61.4%), neurologic (31.5%), or other disabilities (7.1%). Lower extremity joint replacement (27.9%), hip fracture (26.1%), stroke (21.0%), other neurologic (10.5%), or orthopedic (7.5%) impairments were the most frequent RICs. A total of 7,790 patients (88.6%) were admitted by direct transfer from acute hospital wards, whereas 1,006 (11.4%) came from other facilities to continue rehabilitation or from home.

Baseline characteristics and rehabilitation outcomes are shown for the training and validation sets in Table 1. The two sets differed significantly in case-mix composition, patient provenance (P = 0.016), and characteristics such as age (P < 0.001), sex (P = 0.016), and premorbid disability (P < 0.001). The overall burden of medical care was higher among patients in the training dataset, as reflected by the higher prevalence of

ongoing infection (P = 0.006), depression (P = 0.002), pain (P < 0.001), malnutrition (P = 0.003), and urinary catheter (P = 0.001). Patients in the validation set showed a higher prevalence of urinary incontinence (P < 0.001), immobility-related indicators such as transferring from supine to sitting (P < 0.001), standing (P = 0.009), and walking at least 3 m (P = 0.045).

The overall median (IQR) length of stay in rehabilitation was 26 days (16–44), with significant differences related only to RICs (Supplementary Table 2).

A total of 8,342 patients (94.8%) were discharged in a planned manner, whereas 379 patients (4.3%) were transferred to acute hospital wards and 75 patients (0.8%) died during their rehabilitation stay. Inpatient mortality was significantly higher in the test set than in the training set (1.2% vs. 0.6%, P = 0.001).

DAD at discharge affected 3,026 patients (34.4%), significantly higher among patients in the test set than those in the training set (35.8% vs. 33.4%, P = 0.019). DAD occurrence across RICs showed high heterogeneity, ranging from 6.6% in the joint replacement category to 57.8% in the miscellaneous condition category (Supplementary Table 3).

Figure 1 shows the pattern of bivariate association between all candidate predictors for DAD in training set. They were attributed to two supercategories: history and complexity of care. Within history, severe chronic organ failure (cardiac, respiratory, hepatic, and renal) showed a strong positive reciprocal association, whereas dementia showed a strong positive association with age, premorbid mRS, many IMCs, and almost all indicators of functional dependence. From complexity of care, all medical and functional indicators had a strong reciprocal positive association, excluding pain and depression.

3.2. FRIDA score construction and internal validation
The predictive model selected from the Lasso estimator included older age groups, premorbid mRS score, and eight indicators of care complexity as predictors of DAD: IMCs count, communicative disability, dependence in eating, supine-to-sitting transfer, sitting balance, bed-to-chair transfer, sit-to-stand, and Standing (Supplementary Figure 2). Figure 2 shows the adjusted odds ratios of the predictors of DAD obtained from cluster bootstrapping (A) and the resulting FRIDA nomogram scoring system (B). The total FRIDA score ranged from 0 (no complexity) to 39 (extreme complexity) with the associated individual risk of DAD increasing from 2.4% to 99.1%. In the context of internal validation, the FRIDA score showed strong overall discrimination (AUC = 0.888; 95% CI, 0.879–0.897) and near perfect calibration (Supplementary Figure 3).

3.3. External validation
3.3.1. Temporal validation
The RIC-specific AUCs of discrimination ranged from 0.802 (95% CI, 0.706–0.898) in the "miscellaneous conditions" category to 0.866 (95% CI, 0.827–0.903) in the "other neurological conditions" category. The summary AUC was 0.841 (95% CI, 0.826–0.855; P < 0.001), with no heterogeneity among RICs in the discriminant effect (I2 = 0.00%). The 95% PI was 0.821–0.860 (Fig. 3A). For calibration, the log RIC-specific O/E ratios ranged from −0.738 (95% CI, −1.002; −0.474) in the joint replacement category to 0.341 (95% CI, 0.180–0.502), in the debility category. The summary log O/E ratio was 0.017 (95% CI, −0.155 to 0.190; P = 0.842), with evidence of substantial between-RICs heterogeneity (I2 = 88.43%, P < 0.001). The 95% PI was [−0.576, 0.611] (Fig. 3B).

Sensitivity analysis (Supplementary Figure 4) showed that removing the categories of joint replacement and debility was sufficient to drop the heterogeneity in calibration performance (I2 = 16.70%), producing an overall significant underestimation effect (summary log O/E ratio = 0.083; 95% CI, 0.015 to 0.151; P = 0.016). The 95% CI was [−0.068, 0.234].

3.3.2. Spatial validation
Spatial validation involved nine facilities with a total of 3,626 patients. Structure "A" which had no patients in the test set and structure "C" which had only eight patients in the test set were excluded (Supplementary Figure 5). Discrimination performance between facilities remained good with a range of AUCs from 0.759 (95% CI, 0.696–0.822) to 0.970 (95% CI, 0.952–0.988). The summary effect-size for discrimination was 0.861 (95% CI, 0.817–0.906; P < 0.001), with a 95% PI of [0.705, 1.017] (Fig. 3C). The summary calibration was 0.016 (95% CI, −0.135 to 0.167; P = 0.835), with a 95% PI of −0.487 to 0.519 (Fig. 3D). At sensitivity analysis, both discrimination and calibration showed substantial and unmodifiable heterogeneity.

The Supplementary Figure 6 shows in full detail the discrimination and calibration of FRIDA score to temporal and geographic validation.

3.4. Clinical utility
The DCA analysis showed that a decision strategy based on the FRIDA score far outperformed the default strategy of "treat all" within each RIC (Fig. 4). In each RIC, excluding the joint replacement and miscellaneous categories, the net benefit emerged from a probability threshold between 10% and 25%, reaching approximately 25% at the point of overall DAD incidence.

4. Discussion
4.1. Strengths
We converted a system of indicators of clinical complexity into a score that accurately predicts individual risk for dependence in ambulation after rehabilitation across multiple disabilities. The FRIDA score includes as qualifiers advanced age, premorbid disability, count of medical complexity indicators, communicative disability, and six indicators related to immobility at baseline that are easily detected during a standard bedside consultation. Most of these features are already known as stand-alone risk factors for poor health outcomes after hospitalization [[30], [31], [32]], however, to our knowledge, the FRIDA score is the first tool to quantify the joint effect of adverse medical and functional syndromic conditions into a unified measure of individual risk for poor outcomes.

Two main considerations emerged from our analyses. First, baseline complexity indicators are the true "active multimorbidity" of the postacute phase, which is much more powerful in generating outcome prediction than chronic multimorbidities pre-existing the acute event. This is a departure from current tools, which generally overlook postacute syndromes despite their known prognostic importance [11,33]. Second, in our clinical model, complexity indicators at baseline drive the flow of care processes and are used as benchmarks to set and monitor individual patient goals and treatment plans. The review of indicators at discharge indicates a reduction in medical complexity and improvement in patient communication and motor dependence. Thus, the FRIDA score is a summary index that quantifies at the patient level the care needs and their changes and at the facility level the amount and effectiveness of care provided.

4.2. Implications

We believe that the FRIDA score may be of special value in two closely related areas of postacute care, such as the triage process and case-mix adjustment. Rehabilitation triage overlaps with the concepts of prognosis and resource commitment. There is consensus on the importance of having powerful predictors of outcome to guide the transition and care delivery in rehabilitation and there is some literature especially in specific subgroups such as stroke patients [34]. In poststroke rehabilitation, early screening for admission to rehabilitation programs is a standard of quality care but it is essentially limited to the patient's actual ability to successfully participate in the program.

Our results strongly suggest that triage practice should expand beyond this kind of approach to include all postacute conditions. The FRIDA score stratified ambulation dependence risk from 2.4% to 99.1% within each impairment category with good and completely homogeneous accuracy. Our data suggest that the prediction horizon is approximately 30–60 days after the first assessment, depending on the macro-category of functional impairment (orthopedic or neurological).

Excluding the lower limb replacement category, this good performance can also be expected for future patient groups, as suggested by the meta-analyses we conducted, providing a decision-making advantage in identifying patients for treatment far superior to the usual "treat all" strategy. To gain further confidence in appropriately transitioning patients to postacute care services, it will be sufficient to calculate for each category of impairment one or more optimal cutoff threshold(s) for the FRIDA score that maximize its utility.

The spatial validation we conducted showed that the FRIDA score varies with case-mix heterogeneity among rehabilitation facilities, maintaining good discrimination and appreciable calibration. Thus, it is plausible that the FRIDA score could be a metric for case-mix adjustment in general postacute rehabilitation. In our country, inpatient rehabilitation activities are still monitored using the acute Diagnosis-Related Group system, with expert opinion-based adaptations for reimbursement. Transferring the FRIDA score to a Diagnosis-Related Group–like system could generate homogeneous risk-adjusted groups across multiple diagnoses, allowing comparative effectiveness between facilities, as the most careful literature suggests [35,36].

4.3. Limitations
First, the FRIDA score was derived from a model designed for sustainable monitoring in bedside clinical routines of general postacute rehabilitation. For this reason, predictors, even relevant to specific diseases, may have been omitted.

Second, we treated IMCs as equivalent in prediction by testing only their unconditional associations. This simplification may have masked a selection bias for the most important IMCs. More in-depth analyses under causal assumptions could maximize the IMCs selection, while also providing evidence on processes of care that are causally related to failure to recover.

Third, we cannot rule out that the IPER-2.0 study included all patients admitted during the enrollment period or that opportunistic coding was used to complete online data entry. However, we are confident that these potential biases are minor due to the clinical and validation protocols of the IPER 2.0 study and we believe that our results are generalizable because the case-mix of our sample is representative of the inpatient rehabilitation population in our country.

Finally, we are aware that temporal validation is not completely equal to external validation because the target population is from the same facilities. Before recommending the application of the FRIDA score in current practice, we need to confirm our results with studies on truly external patient groups.

5. Conclusion

Advanced age, premorbid disability, and medical and functional adverse syndromes affect all postacute patients with non-negligible prevalence and uniform prognostic magnitude. By quantifying these complexities, the FRIDA score generated an accurate prediction of individual risk for dependence in ambulation at the end of rehabilitation that is transferable across multiple disabilities. The FRIDA score may be a new clinically useful tool for patient-centered decision-making in postacute rehabilitation.

References

[1] A.K. Schaink, K. Kuluski, R.F. Lyons, M. Fortin, A.R. Jadad, R. Upshur, et al. A scoping review and thematic classification of patient complexity: offering a unifying framework.J Comorbidity, 2(1) (2012), pp. 1-9

[2] F.J. Huyse, F.C. Stiefel, P. de Jonge. Identifiers, or "red flags," of complexity and need for integrated care Med Clin North Am, 90 (2006), pp. 703-712

[3] N.D. Shippee, N.D. Shah, C.R. May, F.S. Mair, V.M. Montori. Cumulative complexity: a functional, patient-centered model of patient complexity can improve research and practice. J Clin Epidemiol, 65 (2012), pp. 1041-1051

[4] L. Turner-Stokes, H. Williams, R.J. Siegert. The Rehabilitation Complexity Scale version 2: a clinimetric evaluation in patients with severe complex neurodisability. J Neurol Neurosurg Psychiatry, 81 (2010), pp. 146-153

[5] F. Roda, M. Agosti, A. Merlo, M. Maini, F. Lombardi, C. Tedeschi, et al. Psychometric validation of the Italian rehabilitation complexity scale-extended version 13. PLoS One, 12 (2017)

[6] R.J. Siegert, O. Medvedev, L. Turner-Stokes. Dimensionality and scaling properties of the patient categorisation tool in patients with complex rehabilitation needs following acquired brain injury. J Rehabil Med, 50 (5) (2018), pp. 435-443

[7] D. Wade. Measuring case complexity in neurological rehabilitation. J Neurol Neurosurg Psychiatry, 81 (2) (2010), p. 127

[8] S.J. Lee, M.A. Schonberg, E.W. Widera. Prognostic indices for older adults A systematic review JAMA, 307 (2012), pp. 182-192

[9] S.B. Angleman, G. Santoni, A. Pilotto, L. Fratiglioni, A.K. Welmer. Multidimensional prognostic index in association with future mortality and number of hospital days in a population-based sample of older adults: results of the EU Funded MPI-AGE project. PLoS One, 10 (2015), pp. 1-11

[10] A. Zucchelli, D.L. Vetrano, G. Grande, A. Calderón-Larrañaga, L. Fratiglioni, A. Marengoni, et al. Comparing the prognostic value of geriatric health indicators: a population-based study. BMC Med, 17:185. (2019)

[11] M.E. Tinetti, T. Fried. The end of the disease era. Am J Med, 116 (2004), pp. 179-185

[12] B. Bernardini, M. Gardella, L. Baratto, A. Banchero. IPER 2 Indicatori di Processo Esito in Riabilitazione (versione 2): uno strumento per l'audit clinico e il controllo di gestione. Quaderno n° 10. Genova (2012) Available at: https://www.alisa.liguria.it/, Accessed 15th Dec 2022

[13] G. Bellelli, B. Bernardini, M. Pievani, G.B.G.B. Frisoni, A. Guaita, M. Trabucchi. A score to predict the development of adverse clinical events after transition from acute hospital wards to post-acute care settings Rejuvenation Res, 15 (6) (2012), pp. 553-563

[14] J. Gassaway, S.D. Horn, G. DeJong, R.J. Smout, C. Clark, R. James. Applying the clinical practice improvement approach to stroke rehabilitation: methods used and baseline results. Arch Phys Med Rehabil, 86 (2005), pp. S16-S33

[15] M.G. Stineman, J.A. Shea, A. Jette, C.J. Tassoni, K.J. Ottenbacher, R. Fiedler, et al. The functional independence measure: tests of scaling assumptions, structure, and reliability across 20 diverse impairment categories. Arch Phys Med Rehabil, 77 (1996), pp. 1101-1108

[16] S. Shah, F.C.B. Vanclay. Improving the sensitivity of the Barthel index for stroke rehabilitation. J Clin Epidemiol, 42 (1989), pp. 703-709

[17] T.J. Quinn, J. Dawson, M.R. Walters, K.R. Lees. Exploring the reliability of the modified rankin scale. Stroke, 40 (3) (2009), pp. 762-766

[18] V.M.T. de Jong, K.G.M. Moons, M.J.C. Eijkemans, R.D. Riley, T.P.A. Debray. Developing more generalizable prediction models from pooled studies and large clustered data sets. Stat Med, 40 (2021), pp. 3533-3559

[19] E.W. Steyerberg, Y. Vergouwe. Towards better clinical prediction models: seven steps for development and an ABCD for validation. Eur Heart J, 35 (2014), pp. 1925-1931

[20] K.G.M. Moons, D.G. Altman, J.B. Reitsma, J.P.A. Ioannidis, P. Macaskill, E.W. Steyerberg, et al. Transparent reporting of a multivariable prediction model for individual prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med, 162 (2015), pp. W1-W73

[21] E.W. Steyerberg, F.E. Harrell. Prediction models need appropriate internal, internal-external, and external validation. J Clin Epidemiol, 69 (2016), pp. 245-247

[22] M.C. Johnston, M. Crilly, C. Black, G.J. Prescott, S.W. Mercer. Defining and measuring multimorbidity: a systematic review of systematic reviews. Eur J Public Health, 29 (2019), pp. 182-189

[23] T.T. Hastie. The elements of statistical learning second edition. Math Intell, 27 (2) (2017), pp. 83-85

[24] A. Zlonik, V. Abraira. A general-purpose nomogram generator for predictive logistic regression models Stata J, 15 (2015), pp. 537-546

[25] B. Van Calster, D. Nieboer, Y. Vergouwe, B. De Cock, M.J. Pencina, E.W. Steyerberg. A calibration hierarchy for risk models was defined: from utopia to empirical data. J Clin Epidemiol, 74 (2016), pp. 167-176

[26] K.I.E. Snell, J. Ensor, T.P.A. Debray, K.G.M. Moons, R.D. Riley. Meta-analysis of prediction model performance across multiple studies: which scale helps ensure between-study normality for the C-statistic and calibration measures? Stat Methods Med Res, 27 (11) (2018), pp. 3505-3522

[27] K. Ming Ho. Forest and funnel plots illustrated the calibration of a prognostic model: a descriptive study J Clin Epidemiol, 60 (2007), pp. 746-751

[28] A.J. Vickers, E.B. Elkin. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Mak, 26 (6) (2006), pp. 565-574

[29] A.J. Vickers, B. van Calster, E.W. Steyerberg. A simple, step-by-step guide to interpreting decision curve analysis. Diagn Progn Res, 3 (2019), p. 18

[30] S.F. Simmons, S. Bell, A.A. Saraf, C.S. Coelho, E.A. Long, J.M.L. Jacobsen, et al. Stability of geriatric syndromes in hospitalized medicare beneficiaries discharged to skilled nursing facilities. J Am Geriatr Soc, 64 (10) (2016), pp. 2027-2034

[31] R.G. Holloway, C.G. Benesch, W.S. Burgin, J.B. Zentner. Prognosis and decision making in severe stroke. JAMA, 294 (2005), pp. 725-733

[32] E.J. Roth, L. Lovell, R.L. Harvey, R.K. Bode, A.W. Heinemann Stroke rehabilitation: indwelling urinary catheters, enteral feeding tubes, and tracheostomies are associated with resource use and functional outcomes. Stroke, 33 (7) (2002), pp. 1845-1850

[33] M. Clerencia-Sierra, A. Calderón-Larrañaga, N. Martínez-Velilla, et al. Multimorbidity patterns in hospitalized older patients: Associations among chronic diseases and geriatric syndromes. PLoS One, 10 (2015), pp. 1-14

[34] S. Hakkennes, K.D. Hill, K. Brock, J. Bernhardt, L. Churilov Selection for inpatient rehabilitation after severe stroke: what factors influence rehabilitation assesor decision making. J Rehabil Med, 45 (1) (2013), pp. 24-31

[35] K.E. Covinsky, A.C. Justice, G.E. Rosenthal, R.M. Palmer, C.S. Landefeld
Measuring prognosis and case mix in hospitalized elders: the importance of functional status. J Gen Intern Med, 12 (1997), pp. 203-208

[36] M. Hopfe, G. Stucki, R. Marshall, C.D. Twomey, T.B. Üstün, B. Prodinger
Capturing patients' needs in casemix: a systematic literature review on the value of adding functioning information in reimbursement systems
BMC Health Serv Res, 16 (2015), p. 40

Table 1. Patient characteristics and outcomes

|	Frequency (%)		
	Training set	**Validation set**	**P value**
	N = 5,162	**N = 3,634**	
Rehabilitation Impairment Categories			
Stroke	1,228 (23.8)	616 (17.0)	<0.001
Other Neurologic conditions	577 (11.2)	348 (9.6)	0.016
Hip Fracture	1,351 (26.2)	941 (25.9)	0.786
Lower Extremity Joint Replacement	1,393 (27.0)	1,061 (29.2)	0.023
Other Orthopedic Conditions	333 (6.5)	326 (9.0)	<0.001
Debility	161 (3.1)	255 (7.0)	<0.001
Miscellaneous Conditions	119 (2.3)	87 (2.4)	0.830
Provenance from Acute Hospital Wards	4,536 (87.9)	3,254 (89.5)	0.016
Age y, median (IQR)	75 (66-82)	77 (69-83)	<0.001
Female sex	3,150 (61.0)	2,310 (63.6)	0.016
History			
Severe Organ System Failure			
Heart	468 (9.1)	308 (8.5)	0.340
Respiratory	181 (3.5)	152 (4.2)	0.112
Liver	64 (1.2)	52 (1.4)	0.449
Kidney	118 (2.3)	79 (2.2)	0.770
Dementia	276 (5.4)	189 (5.2)	0.772
Chronic Multimorbidity	2,633 (51.0)	1,787 (49.2)	0.091
Cancer in the last year	176 (3.4)	140 (3.9)	0.295
Premorbid Disability (mRS Score)			<0.001
No symptoms	1,891 (36.6)	968 (26.6)	
No significant disability	1,644 (31.8)	1,304 (35.9)	
Slight disability	722 (14.0)	625 (17.2)	
Moderate	577 (11.2)	525 (14.4)	
Moderate-Severe	275 (5.3)	190 (5.2)	
Severe	53 (1.0)	22 (0.6)	
Social Frailty	434 (8.4)	331 (9.1)	0.265
Indicators of Medical Complexity			
Reduced alertness	137 (2.6)	88 (2.4)	0.537
Delirium	141 (2.7)	87 (2.4)	0.341
Medical instability	629 (12.2)	426 (11.7)	0.527
Ongoing infection	846 (16.4)	518 (14.2)	0.006
Depression	1,651 (32.0)	1,049 (28.9)	0.002
Pain	3,219 (62.4)	2,123 (58.4)	<0.001
Dysphagia	711 (13.8)	474 (13.0)	0.326
Malnutrition	826 (16.0)	499 (13.7)	0.003
Pressure sore	639 (12.4)	473 (13.0)	0.379
Urinary catheter	1,193 (23.1)	727 (20.0)	0.001
Urinary Incontinence (no catheter)a	862/3,969 (21.7)	741/2,907 (25.5)	<0.001
Tracheostomy	57 (1.1)	30 (0.8)	0.229
Indicators of Functional Dependence			
Communicative Disability	1,015 (19.7)	727 (20.0)	0.704
Dependence in Eating	1,153 (22.3)	764 (21.0)	0.149
Dependence in Basic Mobility			
Transfer from Supine to Seated	2,551 (49.4)	1,970 (54.2)	<0.001
Sitting Balance	1,193 (23.1)	781 (21.5)	0.073
Transfer from Bed-to-chair	3,368 (66.2)	2,379 (65.5)	0.838
Sit-to-stand	3,487 (67.5)	2,515 (69.2)	0.104
Standing	3,380 (65.5)	2,476 (69.1)	0.009
Walk for ≥ 3 m	4,133 (80.1)	2,972 (81.8)	0.045
Outcomes			
Days of stay in rehabilitation, median (IQR)	26 (16–43)	25 (16–45)	0.162
Planned discharged	4,906 (95.0)	3,436 (94.6)	0.328
Transferred to acute hospital wards	226 (4.4)	153 (4.2)	0.709
Deceased	30 (0.6)	45 (1.2)	0.001
DAD prevalenceb	1,724 (33.4)	1,302 (35.8)	0.019
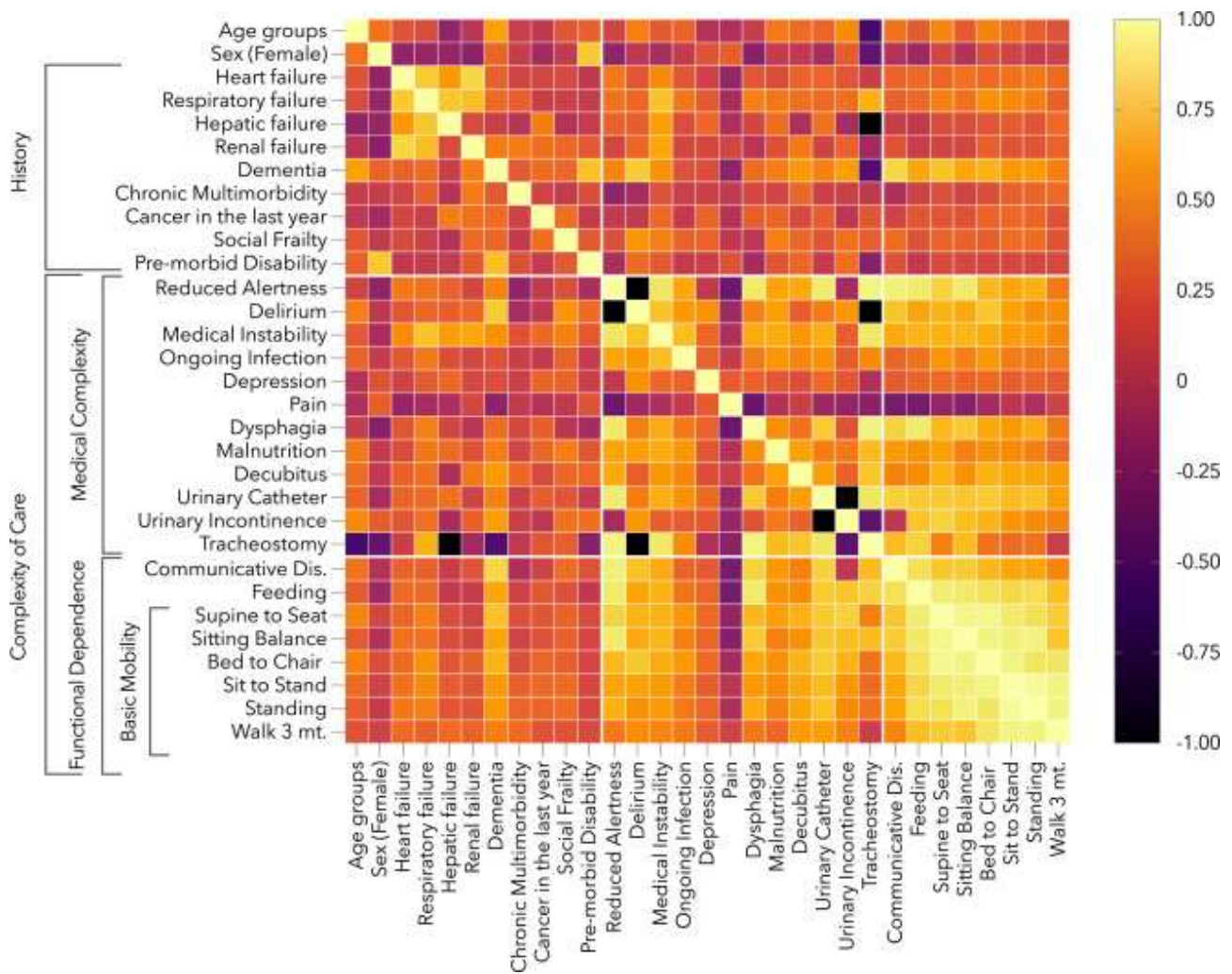
Fig. 1. Bivariate association between the indicators of complexity of the IPER-2.0 system. Heat map showing correlation coefficients between candidate predictors. Pseudocolor bar shows the strength of the association, from directly associated (1.00) to inversely associated (−1.0). The threshold for statistical significance is a correlation coefficient of ± 0.50). Abbreviation: IPER-2.0, Indicators to evaluate PErformance in Rehabilitation version 2.0.
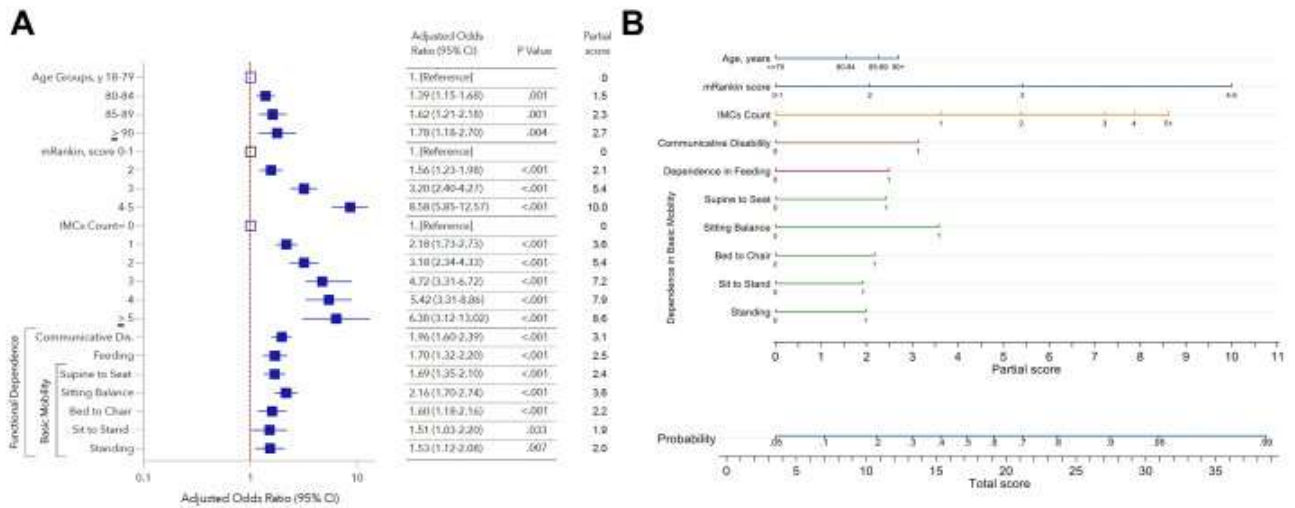
Fig. 2. The FRIDA scoring system to predict dependence in ambulation at rehabilitation discharge. Panel A shows the Forest plot of the adjusted odds ratios of each predictive factor in the multivariable logistic regression model and the partial score corresponding to their prognostic impact. Panel B shows the nomogram reproducing the scoring system in graphical form. The total score, calculated by summing the partial scores, is matched with the probability of dependence in ambulation.
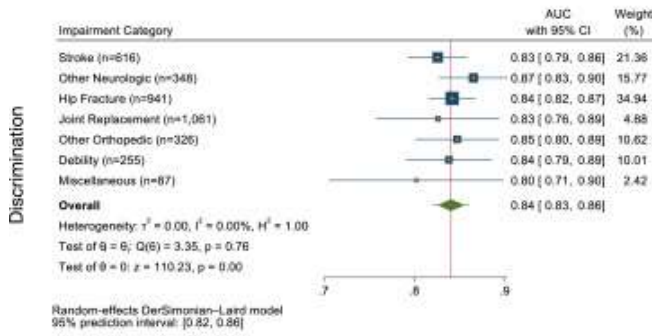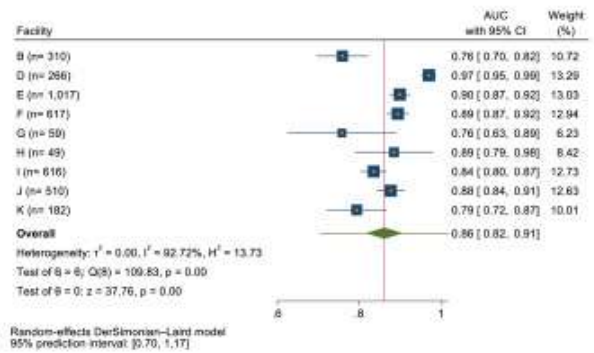
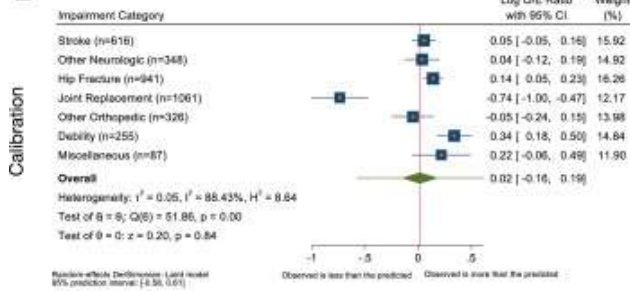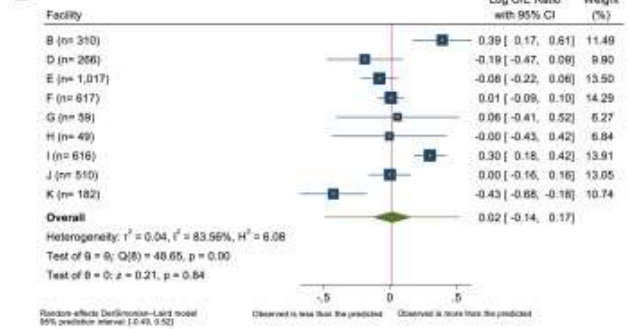Fig. 3. External validation of FRIDA score in predicting dependence in ambulation at discharge from rehabilitation. Forest plots show the effect-size estimates and associated confidence intervals for discrimination and calibration across the impairment categories (temporal validation) and rehabilitation facilities (spatial validation). The overall effect size (the diamond) shows the 95% prediction interval. The calibration was reported as log O/E ratio with the value of 0 indicating perfect calibration.
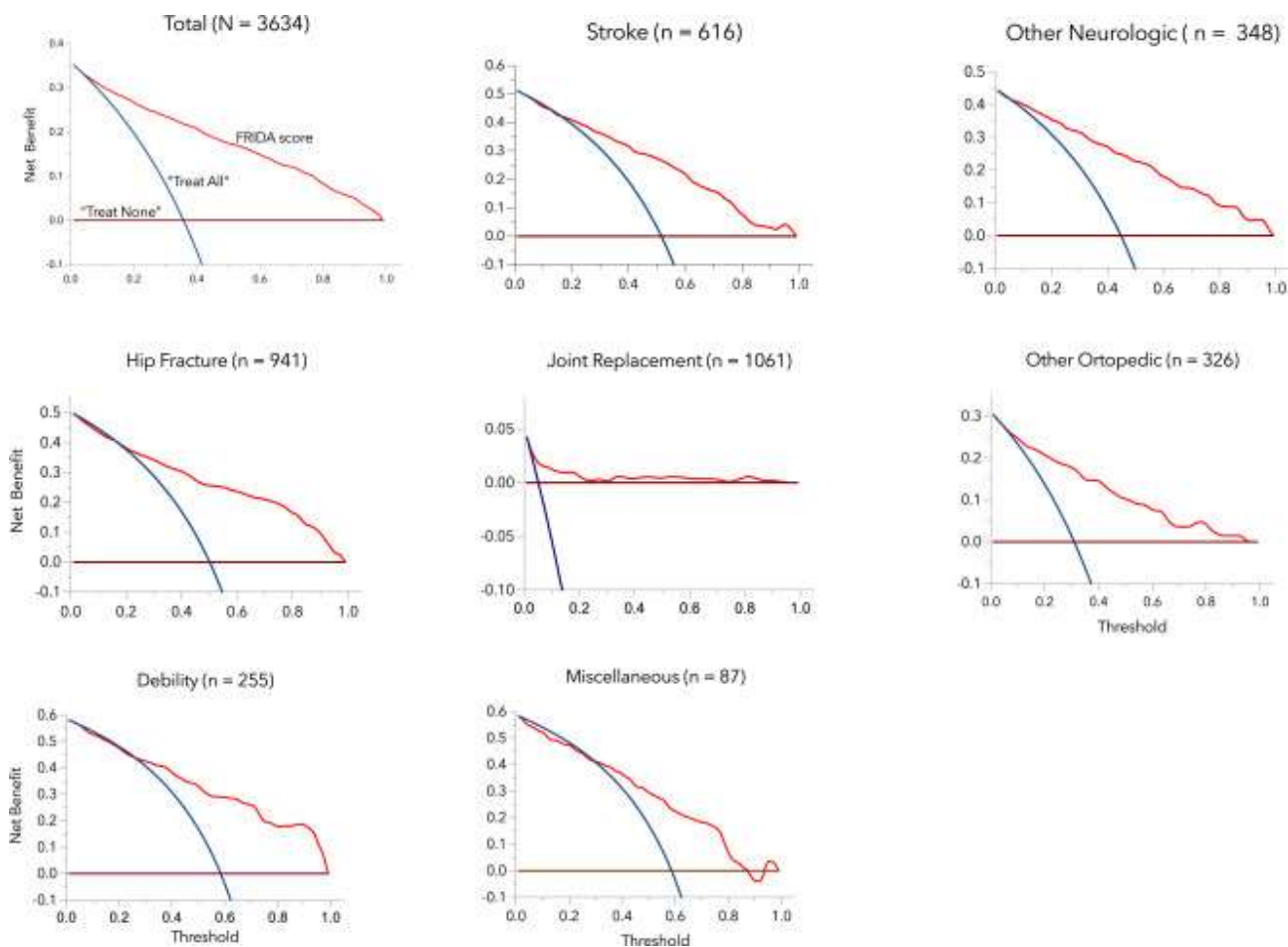
Fig. 4. Clinical utility of FRIDA score across rehabilitation impairment categories. Decision curves showing the net benefit (y-axes) as a function of risk thresholds (x-axes) of a FRIDA-based strategy (red curve) compared to strategies based on "treat all" (blue curve) or "treat none" (brown line). The "treat all" approach assumes that all patients will be dependent in ambulation at discharge from rehabilitation, whereas the "treat none" approach assumes that no patients will be dependent in ambulation at discharge from rehabilitation.