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

Bruno Bernardini, Luigi Baratto, Costanza Pizzi, Annibale Biggeri, Giovanna Cerina, Viviana Colantonio, Carla Corsini, Sara Ghirmai, Marco Pagani, Stefania Fracchia, Marisa Gardella, Dolores Catelan, Maria Luisa Malosio, Elisa Malagamba

PII: S0895-4356(22)00271-2
DOI: https://doi.org/10.1016/j.jclinepi.2022.10.021
Reference: JCE 10945

To appear in: Journal of Clinical Epidemiology

Received Date: 15 March 2022
Revised Date: 24 August 2022
Accepted Date: 31 October 2022


This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2022 Published by Elsevier Inc.
Indicators for Performance Evaluation in Rehabilitation (IPER-2.0) Dataset

N = 8,796

Training Set
N = 5,162 (58.7%)

Validation Set
N = 3,634 (41.3%)

Dependence in Ambulation at Discharge (DAD) from rehabilitation

Outcome

Temporal split

Validation Set
N = 3,634 (41.3%)

IPER-2.0 indicators profiling patient complexity for history and the burden of care at baseline for medical and functional adverse syndromes.

IPER-2.0 system indicators

FRIDA nomogram scoring system

The total FRIDA score ranged from 0 (no complexity) to 39 (extreme complexity) with an associated individual risk for DAD increasing from 2.4% to 99.1%.

Within each RIC, decision curve analysis showed the a decision strategy based on FRIDA score far outperformed the default strategy of “treat all”.

Advanced age, premorbid disability, and medical and functional adverse syndromes affect all post-acute patients with non-negligible prevalence and uniform prognostic magnitude. By quantifying these complexities, the FRIDA score generates 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 post-acute rehabilitation.
A multicenter prospective study validated a nomogram to predict individual risk of dependence in ambulation after rehabilitation

Bruno Bernardini\textsuperscript{a,\*}, Luigi Baratto\textsuperscript{b}, Costanza Pizzi\textsuperscript{c}, Annibale Biggeri\textsuperscript{d,f}, Giovanna Cerina\textsuperscript{a}, Viviana Colantonio\textsuperscript{a}, Carla Corsini\textsuperscript{a}, Sara Ghirmai\textsuperscript{a}, Marco Pagani\textsuperscript{a}, Stefania Fracchia\textsuperscript{e}, Marisa Gardella\textsuperscript{b}, Dolores Catelan\textsuperscript{f}, Maria Luisa Malosio\textsuperscript{g}, Elisa Malagamba\textsuperscript{h}.

Author Affiliations

\textsuperscript{a} IRCCS Humanitas Research Hospital, Neurocenter-Neurorehabilitation Unit, Rozzano, Milan (MI), Italy.
\textsuperscript{b} Department of Rehabilitation, La Colletta Hospital, Arenzano (GE), Italy.
\textsuperscript{c} Department of Medical Sciences, University of Turin, Turin (TO), Italy.
\textsuperscript{d} Department of Statistics, Computer Science, Applications “G. Parenti”, University of Florence (FI), Italy.
\textsuperscript{e} Geriatric Internal Medicine Unit, Garbagnate Hospital, Garbagnate Milanese, Milan (MI), Italy.
\textsuperscript{f} Unit of Biostatistics, Epidemiology and Public Health, Department of Cardiac, Thoracic, Vascular Sciences and Public Health, University of Padua (PD), Italy.
\textsuperscript{g} Institute of Neuroscience, National Research Council (CNR), Laboratory of Pharmacology and Pathology of the Nervous System, IRCCS Humanitas Research Hospital, Rozzano, Milan (MI), Italy.
\textsuperscript{h} Department of Health Services of Liguria Region, Genoa (GE), Italy.

*Corresponding Author: Bruno Bernardini, MD, IRCCS Humanitas Research Hospital, Neurocenter-Neurorehabilitation Unit, Via Alessandro Manzoni 56, 20089 Rozzano, Milan (MI), Italy. Email address: bruno.bernardini@humanitas.it Phone number +39 02 82246738

Co-authors email address:
Luigi Baratto, luigibaratto48@gmail.com ; Costanza Pizzi, costanza.pizzi@unito.it ; Annibale Biggeri, annibale.biggeri@unipd.it ; Giovanna Cerina, giovanna.cerina@humanitas.it ; Viviana Colantonio, viviana.colantonio@humanitas.it ; Carla Corsini, carla.corsini@humanitas.it ; Sara Ghirmai, sara.ghirmai@humanitas.it ; Marco Pagani, marco.pagani@humanitas.it ; Stefania Fracchia, stefania.fracchia@gmail.com ; Marisa Gardella, m.a.gardella@virgilio.it ; Dolores Catelan, Dolores.Catelan@ubep.unipd.it ; Maria Luisa Malosio, marialuisa.malosio@in.cnr.it ; Elisa Malagamba, elisa.malagamba@gmail.com

Declaration of interest: none

Approximate word count of the manuscript: 3900
Abstract

Objective

To develop the FRIDA (Functional Risk Index for Dependence in Ambulation) score, a nomogram to predict individual risk of dependence in ambulation at discharge (DAD) from post-acute 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 8796 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 3026 patients (34.4%) were discharged with DAD. In the training set of 5162 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 a set of 3234 patients (41.3%), meta-analysis showed that the FRIDA score had good homogeneous discrimination ($I^2 = 0.0\%$) (synthetic AUC 0.84, 95% CI= 0.83-0.86) combined with accurate calibration (synthetic Log O/E ratio 0.02, 95% CI -0.16-0.19).

These performances remained stable at spatial validation obtained on 3626 patients from 9 facilities, with higher heterogeneity. Decision curve analyses showed that a FRIDA score-supported strategy far outperformed the usual "treat all" approach in each impairment categories.

Conclusions

The FRIDA score is a new clinically useful tool to predict individual risk for dependence in ambulation at rehabilitation discharge across many different disabilities.

Abstract word count: 229
Keywords

Post-acute rehabilitation, Adverse Clinical Syndromes, Lasso regression, Prognostic Nomogram, Decision Curve Analysis.

Abbreviations

DAD, Dependence in Ambulation at Discharge; DCA, Decision Curve Analysis; FRIDA, Functional Risk Index to predict Dependence in Ambulation; IMCs, Indicators of Medical Complexity; IPER-2.0, Indicators for Performance Evaluation in Rehabilitation, version 2.0; mRS, modified Rankin Scale; RICs, Rehabilitation Impairment Categories.
What is New?

Key findings

- In patients admitted to post-acute rehabilitation, the individual risk for dependence in ambulation at discharge increased from 2.4% to 99.1% due to the combined effect of advanced age, premorbid disability, count of indicators of medical complexity, communicative disability, the dependence in eating and in five key tasks of basic mobility at baseline.
- The resulting FRIDA (Functional Risk Index to predict Dependence in Ambulation) score showed accurate and clinically useful prediction across multiple disabilities.

What this adds to what is known?

- The features of clinical complexity we addressed are already known as distinct risk factors for adverse health outcomes in rehabilitation. However, the FRIDA score is the first bedside tool to summarize the prognostic impact of shared medical and functional adverse syndromes into a single individual risk measure.

What is the implication, what should change now?

- The FRIDA score may be useful in supporting patient-centered decision making in general post-acute rehabilitation, regardless of the major disabling disease. Special areas of application may be rehabilitation triage, case-mix adjustment, and care processes monitoring.
1. Introduction

Many of the patients admitted to rehabilitation are frail elderly with chronic multimorbidity and disabilities pre-existing the acute event. Depending on the impact of acute illness, complex clinical phenotypes are generated in the post-acute 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 [1–3] and paucity of validated measures. An instrument, known as the Rehabilitation Complexity Scale, was proposed by Turner-Stokes and colleagues 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 [8–10]. 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 post-acute 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 IPER-2.0 (Indicators for Performance Evaluation in Rehabilitation, version 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 FRIDA (Functional Risk Index to predict Dependence in Ambulation), 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 (SIMFER) from January 2012 to March 2016, initially enrolling 7 intensive rehabilitation facilities in Liguria, to which another 4 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, cardio-respiratory or severe brain injury units were not included in the IPER-2.0 study.

We considered all 8796 consecutive patients aged 18 years and older admitted for rehabilitation included in the database, without exclusion criteria. We controlled for case-mix composition by classifying the main disabling condition according to 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 7, namely stroke, hip fracture, lower limb joint replacement, debility, other neurological, orthopaedic, or miscellaneous conditions (Supplementary Table 1).

2.2. Ethical Approval

The IPER-2.0 study was conducted according to 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, 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 de-identified data.

2.3. Data collection

All baseline characteristics were collected within 24 hours of patient admission according to 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 de-identified 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 sub-core 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 sub-score is a 5-level categorical scale ranging from 0 (unable to ambulate) to 15 (completely independent in ambulation for at least 50 meters). Patients identified as "dependent" were unable to walk or needed the assistance of a person (sub-score 0 to 8). "Independent" patients were able to walk independently for less (sub-score 12) or more (sub-score 15) than 50 meters.

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, 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, 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 TRIPOD 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 5162 patients (58.7%) discharged from January 2012 to May 2014 and a validation set consisting of 3234 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-square 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 (IMC) 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 5 to 9 possible IMCs. Pain and depression were introduced as individual covariates during modelling.

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 3 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 ten-fold cross-validation [23].

2. Internal model validation by cluster logistic bootstrapping (1000 replications) on data clustered by RICs, rehabilitation centers, patient provenance (hospital wards versus other provenance), and four-month time periods. According to 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 ROC 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 (geographic validity) performance. Discrimination was assessed by the c-statistic, reporting the areas under the curve (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 finally performed pooled meta-analyses of point estimates of AUCs and observed/expected (O/E) ratios with their respective 95% confidence intervals, 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 under- or over-prediction, respectively [27].
All meta-analyses were conducted according to a random effects model with Der Simonian-Laird inverse variance weighting. We reported the effect size with 95% confidence interval (CI) and an approximate 95% prediction interval (PI). Heterogeneity was assessed by the $I^2$ statistic and examined by sensitivity analysis for values greater 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 under- or over-treatment [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, TX, USA). Reported P values are 2-sided, and statistical significance was set at $P < 0.05$. 
3. Results

3.1. Descriptive statistic and bivariate analyses

We analyzed 8796 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 7790 patients (88.6%) were admitted by direct transfer from acute hospital wards, whereas 1006 (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 meters (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 8342 patients (94.8%) were discharged in a planned manner, while 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 3026 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 super-categories: history and complexity of care. Within history, severe chronic organ failure (cardiac, respiratory, hepatic, 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: 1. IMCs count, 2. communicative disability, dependence in 3. eating, 4. supine-to-sitting transfer, 5. sitting balance, 6. bed-to-chair transfer, 7. sit-to-stand, and 8. 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 ($I^2 = 0.00\%$). The 95% PI was 0.821 - 0.860
A bedside score to predict dependence in ambulation after rehabilitation.

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 – 0.190; P = 0.842), with evidence of substantial between-RICs heterogeneity ($I^2 = 88.43\%$, P <0.001). The 95% PI was [-0.576, 0.611].

(Figure 3.A). Sensitivity analysis (Supplementary Figure 4) showed that removing the categories of joint replacement and debility was sufficient to drop the heterogeneity in calibration performance ($I^2 = 16.70\%$), producing an overall significant underestimation effect (summary O/E ratio = 0.083; 95% CI, 0.015 to 0.151; P = 0.016). The 95% CI was [-0.068, 0.234].

(Figure 3.B).

Sensitivity analysis (Supplementary Figure 4) showed that removing the categories of joint replacement and debility was sufficient to drop the heterogeneity in calibration performance ($I^2 = 16.70\%$), producing an overall significant underestimation effect (summary 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. Geographic validation

Geographic validation involved 9 facilities with a total of 3626 patients. Structure "A" which had no patients in the test set and structure "C" which had only 8 patients in the test set were excluded (Supplementary Figure 5). Discrimination performance between structures 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] (Figure 3.C). The summary calibration was 0.016 (95% CI, -0.135 - 0.167; P = 0.835), with a 95% PI of -0.487 to 0.519 (Figure 3.D). 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 (Figure 4). In each RIC, excluding the joint replacement and miscellaneous categories, the net benefit emerged from a probability threshold between 10%-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 single measure of individual risk for poor outcome.

Two main considerations emerged from our analyses. First, baseline complexity indicators are the true "active multimorbidity" of the post-acute phase, which is much more powerful in generating outcome prediction than chronic multimorbidities preexisting the acute event. This is a departure from current tools, which generally overlook post-acute 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 post-acute 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 post-stroke 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 post-acute 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 post-acute 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 geographic 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 post-acute rehabilitation. In our country, inpatient rehabilitation activities are still monitored using the acute Diagnosis-Related Group (DRG) system, with expert opinion-based adaptations for reimbursement. Transferring the FRIDA score to a DRG-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 post-acute 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.

Last, we are aware that temporal validation is not completely equal to external validation because the target population are 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. Conclusions

Advanced age, premorbid disability, and medical and functional adverse syndromes affect all post-acute 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 post-acute rehabilitation.
REFERENCES


A bedside score to predict dependence in ambulation after rehabilitation

TABLE/FIGURE TITLES AND LEGENDS

Table 1 title: Patient characteristics and outcomes
Legend: a Urinary incontinence was detected in patients without a bladder catheter.
* DAD includes patients transferred to acute hospital wards or died during their rehabilitation stay.
Abbreviations: RICs, rehabilitation impairment categories; mRS, modified Rankin Score; DAD, dependence in ambulation at rehabilitation discharge.

Figure 1 title: Bivariate association between the indicators of complexity of the IPER-2.0 system
Legend: Heat map showing correlation coefficients between candidate predictors. Pseudocolor bar show 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.

Figure 2 title: The FRIDA scoring system to predict dependence in ambulation at rehabilitation discharge
Legend: The Forest plot (A) shows the adjusted odd ratios of each predictive factor in the multivariable logistic regression model and the partial score corresponding to the prognostic impact. The nomogram (B) reproduces the scoring system in graphical form. The total score, calculated by summing the partial scores, is matched with the probability of dependence in ambulation.

Figure 3 title: Temporal and geographic validation of FRIDA score in predicting dependence in ambulation at discharge from rehabilitation
Legend: Forest plots show the effect-size estimates and associated confidence intervals for discrimination and calibration across the impairment categories and rehabilitation facilities. The overall effect size (the diamond) shows the 95% prediction interval. The calibration was reported as log risk ratio with the value of 0 indicating perfect calibration.
**Figure 4 title: Clinical Utility of FRIDA score Across Rehabilitation Impairment Categories**

Legend: 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.
<table>
<thead>
<tr>
<th></th>
<th>Training Set</th>
<th>Validation Set</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N = 5162</td>
<td>N = 3634</td>
<td></td>
</tr>
<tr>
<td>Rehabilitation Impairment Categories</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stroke</td>
<td>1228 (23.8)</td>
<td>616 (17.0)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Other Neurologic conditions</td>
<td>577 (11.2)</td>
<td>348 (9.6)</td>
<td>0.016</td>
</tr>
<tr>
<td>Hip Fracture</td>
<td>1351 (26.2)</td>
<td>941 (25.9)</td>
<td>0.786</td>
</tr>
<tr>
<td>Lower Extremity Joint Replacement</td>
<td>1393 (27.0)</td>
<td>1061 (29.2)</td>
<td>0.023</td>
</tr>
<tr>
<td>Other Orthopedic Conditions</td>
<td>333 (6.5)</td>
<td>326 (9.0)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Debility</td>
<td>161 (3.1)</td>
<td>255 (7.0)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Miscellaneous Conditions</td>
<td>119 (2.3)</td>
<td>87 (2.4)</td>
<td>0.830</td>
</tr>
<tr>
<td>Provenance from Acute Hospital Wards</td>
<td>4536 (87.9)</td>
<td>3254 (89.5)</td>
<td>0.016</td>
</tr>
<tr>
<td>Age yrs, median (IQR)</td>
<td>75 (66-82)</td>
<td>77 (69-83)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Female sex</td>
<td>3150 (61.0)</td>
<td>2310 (63.6)</td>
<td>0.016</td>
</tr>
<tr>
<td>History</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Severe Organ System Failure</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heart</td>
<td>468 (9.1)</td>
<td>308 (8.5)</td>
<td>0.340</td>
</tr>
<tr>
<td>Respiratory</td>
<td>181 (3.5)</td>
<td>152 (4.2)</td>
<td>0.112</td>
</tr>
<tr>
<td>Liver</td>
<td>64 (1.2)</td>
<td>52 (1.4)</td>
<td>0.449</td>
</tr>
<tr>
<td>Kidney</td>
<td>118 (2.3)</td>
<td>79 (2.2)</td>
<td>0.770</td>
</tr>
<tr>
<td>Dementia</td>
<td>276 (5.4)</td>
<td>189 (5.2)</td>
<td>0.772</td>
</tr>
<tr>
<td>Chronic Multimorbidity</td>
<td>2633 (51.0)</td>
<td>1787 (49.2)</td>
<td>0.091</td>
</tr>
<tr>
<td>Cancer in the last year</td>
<td>176 (3.4)</td>
<td>140 (3.9)</td>
<td>0.295</td>
</tr>
</tbody>
</table>
Table 1. Patient characteristics and outcome (continue)

<table>
<thead>
<tr>
<th></th>
<th>Training Set</th>
<th>Validation Set</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>N = 5162</td>
<td>N = 3634</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Pre-morbid Disability (mRS Score)</strong></td>
<td></td>
<td></td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>0. No symptoms</td>
<td>1,891 (36.6)</td>
<td>968 (26.6)</td>
<td></td>
</tr>
<tr>
<td>1. No significant disability</td>
<td>1,644 (31.8)</td>
<td>1,304 (35.9)</td>
<td></td>
</tr>
<tr>
<td>2. Slight disability</td>
<td>722 (14.0)</td>
<td>625 (17.2)</td>
<td></td>
</tr>
<tr>
<td>3. Moderate</td>
<td>577 (11.2)</td>
<td>525 (14.4)</td>
<td></td>
</tr>
<tr>
<td>4. Moderate-Severe</td>
<td>275 (5.3)</td>
<td>190 (5.2)</td>
<td></td>
</tr>
<tr>
<td>5. Severe</td>
<td>53 (1.0)</td>
<td>22 (0.6)</td>
<td></td>
</tr>
<tr>
<td><strong>Social Frailty</strong></td>
<td></td>
<td></td>
<td>0.265</td>
</tr>
<tr>
<td>434 (8.4)</td>
<td>331 (9.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Indicators of Medical Complexity</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reduced alertness</td>
<td>137 (2.6)</td>
<td>88 (2.4)</td>
<td>0.537</td>
</tr>
<tr>
<td>Delirium</td>
<td>141 (2.7)</td>
<td>87 (2.4)</td>
<td>0.341</td>
</tr>
<tr>
<td>Medical instability</td>
<td>629 (12.2)</td>
<td>426 (11.7)</td>
<td>0.527</td>
</tr>
<tr>
<td>Ongoing infection</td>
<td>846 (16.4)</td>
<td>518 (14.2)</td>
<td>0.006</td>
</tr>
<tr>
<td>Depression</td>
<td>1651 (32.0)</td>
<td>1049 (28.9)</td>
<td>0.002</td>
</tr>
<tr>
<td>Pain</td>
<td>3219 (62.4)</td>
<td>2123 (58.4)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Dysphagia</td>
<td>711 (13.8)</td>
<td>474 (13.0)</td>
<td>0.326</td>
</tr>
<tr>
<td>Malnutrition</td>
<td>826 (16.0)</td>
<td>499 (13.7)</td>
<td>0.003</td>
</tr>
<tr>
<td>Pressure sore</td>
<td>639 (12.4)</td>
<td>473 (13.0)</td>
<td>0.379</td>
</tr>
<tr>
<td>Urinary catheter</td>
<td>1193 (23.1)</td>
<td>727 (20.0)</td>
<td>0.001</td>
</tr>
<tr>
<td>Urinary incontinence (no catheter)</td>
<td>862/3969 (21.7)</td>
<td>741/2907 (25.5)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Tracheostomy</td>
<td>57 (1.1)</td>
<td>30 (0.8)</td>
<td>0.229</td>
</tr>
</tbody>
</table>
Table 1. Patient characteristics and outcomes (continued).

<table>
<thead>
<tr>
<th>Indicators of Functional Dependence</th>
<th>Training Set</th>
<th>Validation Set</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communicative Disability</td>
<td>1015 (19.7)</td>
<td>727 (20.0)</td>
<td>0.704</td>
</tr>
<tr>
<td>Dependence in Eating</td>
<td>1153 (22.3)</td>
<td>764 (21.0)</td>
<td>0.149</td>
</tr>
<tr>
<td>Dependence in Basic Mobility</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transfer from Supine to Seated</td>
<td>2551 (49.4)</td>
<td>1970 (54.2)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Sitting Balance</td>
<td>1193 (23.1)</td>
<td>781 (21.5)</td>
<td>0.073</td>
</tr>
<tr>
<td>Transfer from Bed-to-chair</td>
<td>3368 (66.2)</td>
<td>2379 (65.5)</td>
<td>0.838</td>
</tr>
<tr>
<td>Sit-to-stand</td>
<td>3487 (67.5)</td>
<td>2515 (69.2)</td>
<td>0.104</td>
</tr>
<tr>
<td>Standing</td>
<td>3380 (65.5)</td>
<td>2476 (69.1)</td>
<td>0.009</td>
</tr>
<tr>
<td>Walk for ≥ 3 meters</td>
<td>4133 (80.1)</td>
<td>2972 (81.8)</td>
<td>0.045</td>
</tr>
</tbody>
</table>

| Outcomes                           |              |                |         |
| Days of stay in rehabilitation, median (IQR) | 26 (16-43)  | 25 (16-45)     | 0.162   |
| Planned discharged                  | 4906 (95.0)  | 3436 (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 prevalence<sup>b</sup>          | 1724 (33.4)  | 1302 (35.8)    | 0.019   |

Abbreviations: mRS, modified Rankin Scale; DAD, dependence in ambulation at discharge.

*Urinary incontinence was detected in patients without a bladder catheter.

<sup>b</sup> Patients transferred to acute wards and those who died during their rehabilitation stay were included.
Author Statement

Author Contributions
Concept and design: Bernardini, Baratto, Biggeri.
Analysis, and interpretation of data: All authors.
Methodological supervision: Biggeri.
Statistical analysis: Pizzi, Bernardini.
Drafting the manuscript: Bernardini, Pizzi, Fracchia.
Revision of the manuscript: All authors.
Critical revision of the manuscript for important intellectual content: Catelan, Malosio, Malagamba.

Declaration of Interest: The authors have no conflicts of interest to declare.

Funding/Support: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The IPER-2.0 Rehabilitation Quality Improvement Study was funded by the Region of Liguria, Department of Health Services.

Bernardini and Baratto had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Acknowledgements
The authors would like to thank Dr. Francesco Benvenuti and Dr. Sante Giardini for their support and strong geriatric expertise during the construction of the IPER-1.0 Indicator System, Marco Bressi and Chiara Bellia for creating the IPER-2.0 web platform and serving as data managers, and Francesco Copello for statistical support and reporting during the IPER-2.0 Rehabilitation Quality Improvement study of the Liguria Region.

Data Availability Statement
The data of this study can be available upon reasonable request to the corresponding author.

The Italian IPER-2.0 group
Bagnoli Roberto, Azienda di Servizi alla Persona (ASP) Pio Albergo Trivulzio, (Milano); Benevolo Emilio, Fondazione S. Maugeri, Genova Nervi (Genova); Benvenuti Francesco, UOC Cura e Riabilitazione delle Fragilità, Dipartimento Territorio-Fragilità AUSL11, Empoli; Checchia Giovanni Antonio, SC Recupero e Rieducazione Funzionale Ospedale Santa Corona, Pietra Ligure (Savona); Forni Marco, Fondazione don Carlo Gnocchi Ospedale S. Bartolomeo, Sarzana (La Spezia); Gaggero Lorenza, Presidio Riabilitativo “Presentazione”, Loano (Savona); Gardella Marisa, SC Recupero e Rieducazione Funzionale Ospedale “La Colletta”, Arenzano (Genova); Giardini Sante, Unità di Ortoperiatria, Ospedale Santa Maria Annunziata, Azienda Sanitaria di Firenze; Grosso Vittorio, SSD Rieducazione e Riabilitazione Funzionale Ospedale di Cairo Montenotte (Savona); Leoni Valeria, SC Medicina Fisica e Riabilitazione Ospedaliera, Sestri Levante (Genova); Mayer Federico, Casa di Cura Villa Ulivella (Firenze); Panella Lorenzo, Dipartimento di Riabilitazione Integrata Ospedale Territorio, ASL Vercelli (Vercelli); Prina Roberto, Istituto Geriatrico P. Redaelli, Vimodrone (Milano); Spaghetti Ilaria, UO Recupero e Rieducazione Funzionale, Azienda USL Prato (Prato); Ventura Francesco, SC di Riabilitazione e Rieducazione Funzionale IRCCS AOU San Martino – IST, (Genova).