ASNR AMERICAN SOCIETY OF NEUROREHABILITATION

What the Proportional Recovery Rule Is (and Is Not): Methodological and Statistical Considerations

Robinson Kundert, MSc[1,2,3] iD, Jeff Goldsmith, PhD[4*], Janne M. Veerbeek, PhD[1,2], John W. Krakauer, MD[5*], and Andreas R. Luft, MD[1,2]

Abstract

In 2008, it was proposed that the magnitude of recovery from nonsevere upper limb motor impairment over the first 3 to 6 months after stroke, measured with the Fugl-Meyer Assessment (FMA), is approximately 0.7 times the initial impairment ("proportional recovery"). In contrast to patients with nonsevere hemiparesis, about 30% of patients with an initial severe paresis do not show such recovery ("nonrecoverers"). Hence it was suggested that the proportional recovery rule (PRR) was a manifestation of a spontaneous mechanism that is present in all patients with mild-to-moderate paresis but only in some with severe paresis. Since the introduction of the PRR, it has subsequently been applied to other motor and nonmotor impairments. This more general investigation of the PRR has led to inconsistencies in its formulation and application, making it difficult to draw conclusions across studies and precipitating some cogent criticism. Here, we conduct a detailed comparison of the different studies reporting proportional recovery and, where appropriate, critique statistical methodology. On balance, we conclude that existing data in aggregate are largely consistent with the PRR as a population-level model for upper limb motor recovery; recent reports of its demise are exaggerated, as these excessively focus on the less conclusive issue of individual subject-level predictions. Moving forward, we suggest that methodological caution and new analytical approaches will be needed to confirm (or refute) a systematic character to spontaneous recovery from motor and other poststroke impairments, which can be captured by a mathematical rule either at the population or at the subject level.

Keywords

Introduction

It has been appreciated since Hippocrates that the strongest predictor of final motor impairment after stroke is initial impairment (Aphorisms of Hippocrates, Section 2: 42). A prominent poststroke motor impairment in humans is the intrusion of unwanted synergies, with synergy referring to a systematic pattern of either joint co-articulation or muscle co-activation. The Fugl-Meyer Assessment (FMA) was explicitly developed to track progression of recovery from such synergies. A seminal study tracking the recovery of patients using the upper extremity subscale of the Fugl-Meyer Assessment (FMA-UE) demonstrated that more severely affected patients saw greater recovery in this outcome, on average, than more mildly affected patients in the immediate poststroke recovery period[1]; however, the average final score of the FMA-UE among the severly affected still trailed behind the mildly affected. The authors of this study stated, "The most dramatic recovery in motor function occurred over the first 30 days, regardless of the initial severity of the stroke." On the basis of this study and other

considerations, Krakauer et al[2] sought to investigate the nature of this FMA-UE *change* early after stroke; work that led to the formulation of the proportional recovery rule (PRR).[2] The PRR states that patients recover approximately 70% of their maximal potential reduction in impairment as measured by the FMA.[2]

[1]University of Zurich and University Hospital Zurich, Zurich, Switzerland
[2]cereneo, Center for Neurology and Rehabilitation, Vitznau, Switzerland
[3]Swiss Federal Institute of Technology, Zurich, Switzerland
[4]Columbia University, New York, NY, USA
[5]Johns Hopkins University, Baltimore, MD, USA

*These authors are members of The American Society of Neurorehabilitation.

Corresponding Author:
Robinson Kundert, cereneo, center for Neurology and Rehabilitation, Seestrasse 18, Vitznau, 6354, Switzerland.
Email: robinson.kundert@hest.ethz.ch

Since it was introduced, the PRR has been applied in a broad range of studies that involve recovery from stroke, both for FMA-UE and for other outcomes. Claims related to the PRR have been made for upper and lower limb impairment measured by the FMA,[3-10] aphasia measured with the Western Aphasia Battery (WAB),[11] the resting motor threshold (RMT) of the extensor carpi radialis,[6] and visuospatial neglect measured with the Letter Cancellation Test (LCT),[12] among others. Applications of the PRR typically distinguish between two distinct subgroups of patients, referred to as "recoverers" and "nonrecoverers": the former subgroup is composed of patients who recover a significant amount of lost function, and the latter is composed of those who do not. The PRR is thought to usefully characterize the recovery process among recoverers only. Although the methods by which the PRR was applied and evaluated have differed substantially across publications, many authors have argued that their findings are evidence for a PRR that accurately describes an underlying biological process that arises across neurolocical domains. Recently, however, the PRR has been the subject of criticism related to the validity of the statistical methods underlying its implementation and to the degree to which data are consistent with claims in support of the PRR.[13,14] Much of the critique on the PRR articulated by these articles was focused on specific statements associated with the PRR followed by a general dismissal of all findings.

Our goal in this work is to provide a critical reexamination of the literature pertaining to the PRR. We focus first on the interpretation and implementation of PRR as a statistical model, and on data-driven concerns about the use of the PRR in studies of recovery. We then reexamine data reported in the literature and the extent to which past studies provide evidence for the PRR with these considerations in mind. Our hope is that this will serve as an instructive overview of issues that can arise in the application of the PRR to studies of recovery, aiming to improve future investigations into the PRR. Although our primary purpose is not to provide direct response to recent critiques,[13,14] we are mindful of the concerns raised and address these directly in the Discussion section.

The breadth of work on the PRR introduces a commensurate range of methodological concerns one might consider. We attempt to be complete in our discussion but prefer to focus on overarching concerns regarding the statistical validity of the PRR instead of point-by-point inspections of the existing literature. Two themes we will revisit while pursuing the main goals of this paper are the identification of recoverers and the distinction between describing biological mechanisms and making patient-level predictions. The manner in which nonrecoverers are identified is a point of legitimate concern, as some statistical approaches can artifactually create evidence for the PRR. The PRR was originally intended to describe biological mechanisms at the population level, although implicitly it is expected that the PRR may be useful for predicting recovery of individual patients. Both of these are related to recent concerns regarding the PRR.

The next section provides an overview of the statistical formulation of the PRR and introduces three simulated datasets to illustrate scenarios over which the PRR shows varying degrees of validity. Subsequent sections conduct a selective review of the literature, reevaluating specific articles in the light of the three scenarios, comment on recent criticisms of the PRR, and end with our current view on the veracity of the PRR.

Model Formulation and Simulated Examples

As a statistical model, the PRR can be expressed as a linear regression with the initial impairment as the main or only predictor and the observed recovery as the response. The slope of the regression line is interpreted as the proportion of recovery.

Notationally, we formulate the PRR using FMA-UE; the measure for which it was originally intended. We define initial impairment (FMA-UE$_{ii}$) by subtracting the initial measurement of the FMA-UE early after stroke (FMA-UE$_i$) from the ceiling of the FMA-UE (FMA-UE$_{ii}$ = 66 − FMA-UE$_i$). Change in FMA-UE (ΔFMA-UE) is calculated by subtracting initial impairment FMA-UE$_{ii}$ from the FMA-UE measured at the end of the subacute phase (FMA-UE$_f$), so that ΔFMA-UE = FMA-UE$_f$ − FMA-UE$_i$.

The PRR posits that, among those identified as "recoverers," patients are expected to regain a fixed proportion of the initial impairment:

$$\Delta \text{FMA-UE} = \text{prop} * \text{FMA-UE}_{ii} + \text{error}$$

When applying the PRR, a typical approach is to fit a linear regression model relating initial impairment to change in impairment to produce an estimate of the proportion recovered. Under the PRR there is no intercept; this can be constrained in the model fitting by excluding an intercept, or an intercept can be estimated and compared to the 0-value anticipated by the PRR. We leave aside the issue of identifying and removing "nonrecoverers" from analyses but return to this in the Discussion section.

As for other linear models, careful regression diagnostics using a strict examination of residuals is essential when using the PRR. We highlight two main areas for attention: possible heteroscedasticity in errors and potential nonlinear associations between recovery and initial impairment. Usual approaches to linear models assume constant variance of the residuals across levels of the predictor; when this assumption is not met, the usual techniques for estimation, inference, and summarizing goodness-of-fit will not
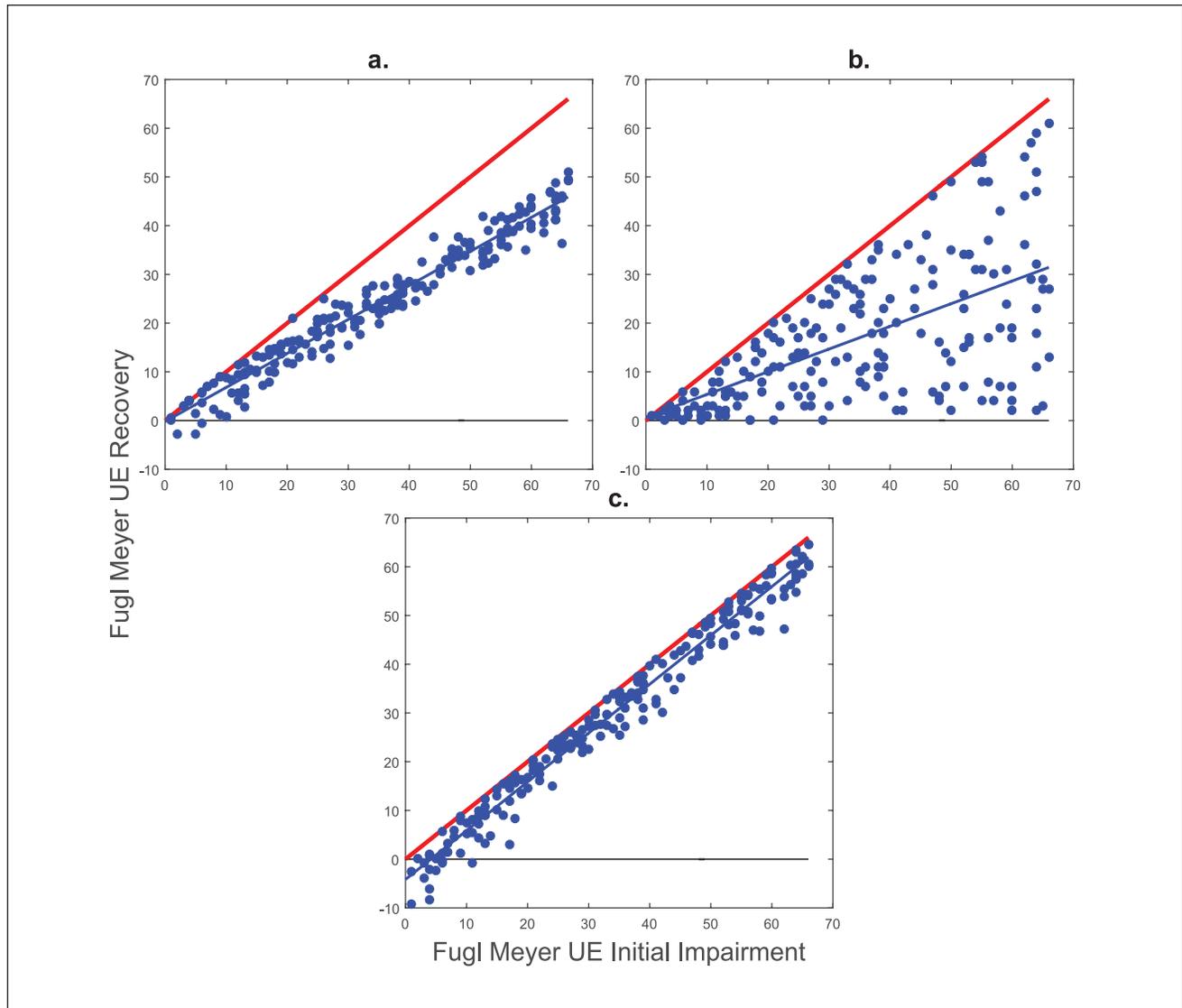
Figure 1. Three simulated datasets illustrating appropriate and nonappropriate application of the proportional recovery rule (PRR). Red line: Ceiling line, data points cannot lie above the ceiling line. Blue dots: Simulated datapoints. Blue line: Linear regression line of the simulated datapoints. (1a): Simulated canonical PRR. (1b): Randomly distributed data drawn from a uniform distribution. (1c): Simulated data of close to full recovery; UE, upper extremity subscale.

be suitable. Separately, the PRR suggests that the recovery proportion is constant across the range of initial impairment values; if this is inaccurate, the single estimated recovery proportion will oversimplify the true association. In both cases, departures from assumptions could be obscured through summary statistics like the estimated recovery proportion and its confidence interval.

We introduce three simulated examples to illustrate the PRR and to provide context for our discussion of the published literature (code generating each is available in a supplement). In each case, initial impairment is drawn uniformly between 0 and 66. The first simulated example assumes the 70% PRR with a relatively narrow error

distribution (Figure 1a). This simulation represents a canonical example of proportional recovery, wherein the recovery proportion can be interpreted biologically and is useful in making specific patient-level predictions. In the second simulated dataset, ΔFMA-UE is generated randomly between 0 and FMA-UE$_{ii}$—that is, between no recovery and full recovery (Figure 1b). This simulation is consistent with the PRR using a recovery proportion of 50% and an error distribution that violates usual assumptions of error homogeneity in linear regression models. The third simulated example implements a version of recovery to ceiling in which all patients approach full recovery and many patients do fully recover (Figure 1c).

The PRR is written and can be interpreted as a standard linear model. The expected ΔFMA-UE is a fixed proportion of FMA-UE$_{ii}$, and that proportion is shared across patients; in that sense, the PRR is appropriate for summarizing population-level patterns related to recovery. As noted, before, though, there is an implicit expectation that significant associations in the PRR will result in accurate predictions of subject-level recovery and, in turn, subject-level final outcomes. Individual predictions based on fitted values from the PRR can be obtained, but their accuracy will depend heavily on the error distribution. Indeed, direct applications of the PRR to both our first and second simulated datasets will produce estimated recovery proportions that differ significantly from zero but only the first simulated dataset could be used to make meaningful subject-level predictions. In settings where a primary goal is to make accurate subject-level predictions for tailoring rehabilitation and informing patients, pairing usual regression diagnostics with formal assessments of prediction accuracy is necessary.

Data Reported in the Literature

In this section, we revisit the datasets on which the PRR is based and then reexamine subsequent datasets that have applied the rule prospectively. We undertake this review with a particular emphasis on the methods applied and the interpretation of the data in order to clarify those that do and do not support the PRR. For this purpose, we examine articles that led to the formulation of the PRR and its subsequent popularity, specifically, we examine them through the lens of the simulated scenarios in the previous section.

The publications discussed in Krakauer and Marshall,[2] as well as publications directly building on these that were published since that time, were considered in the following. Data were extracted from the figures in the original publications. This extraction does not create an exact replica of the dataset because some data points overlap, and it is not possible to know which those are. However, there are not many overlapping data points and so our extraction is a fair representation of the reported data. Reported data were transformed to the same coordinate system as the simulated data to enable consistent comparison between datasets. We first consider examples from the literature that focus on the FMA-UE as the primary outcome and then move on to those that look at other outcomes.

FMA-UE Examples

The PRR was first reported by Prabhakaran et al[3] for the FMA-UE in patients suffering from a first-time ischemic stroke. A multivariate linear regression analysis with ΔFMA-UE as the outcome variable and several covariates
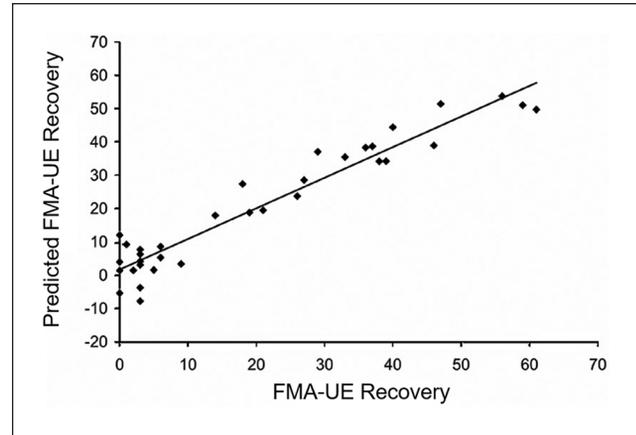

Figure 2. The relation between predicted FMA-UE (upper extremity subscale of the Fugl-Meyer Assessment) recovery and observed recovery shown in Prabhakaran et al.[3] The data shown represents the results of the multivariate linear regression and not the results of the univariate equation ΔFMA-UE = 0.7 * FMA-UE$_{ii}$ + 0.4 leading to the formulation of the proportional recovery rule (PRR). Adapted from Prabhakaran et al[3] copyright © 2008 by The American Society of Neurorehabilitation. Reprinted by permission of SAGE Publications, Inc.

was fit (see Figure 2). Outliers, formally identified, of the regression analysis were classified as nonrecoverers and excluded from the analysis; for the resulting data, backward selection identified subcortical lesion volume, age, time to reassessment and initial FMA-UE measured within three days after stroke as significant predictors of recovery. Fixing subcortical lesion volume, age and time to reassessment at their averaged values produced a univariate relationship interpretable as the association between ΔFMA-UE and FMA-UE$_{ii}$:

$$\Delta\text{FMA-UE} = 0.7 * \text{FMA-UE}_{ii} + 0.4,$$

which is the first formulation of the PRR.

This formulation of the PRR was, therefore, not a prespecified model with FMA-UE$_{ii}$ as the sole predictor and ΔFMA-UE as the outcome; an approach that evolved in subsequent work. Rather, the PPR resulted from a multivariate regression in which three of four covariates were evaluated at their mean value to emphasize the recovery proportion as a variable of interest. Many of the results in Prabhakaran et al,[3] including Figure 2, refer to the multivariate regression. Consequently, these can offer only limited support of the PRR: a more formal assessment of the PRR as a univariate model would give more formal evidence. Prabhakaran et al[3] concluded that as the intercept of the equation is close to zero, the relation between FMA-UE$_{ii}$ and ΔFMA-UE must be a proportional one; this statement is based on the model for an "average" subject, and thus may only be partially true in these data.
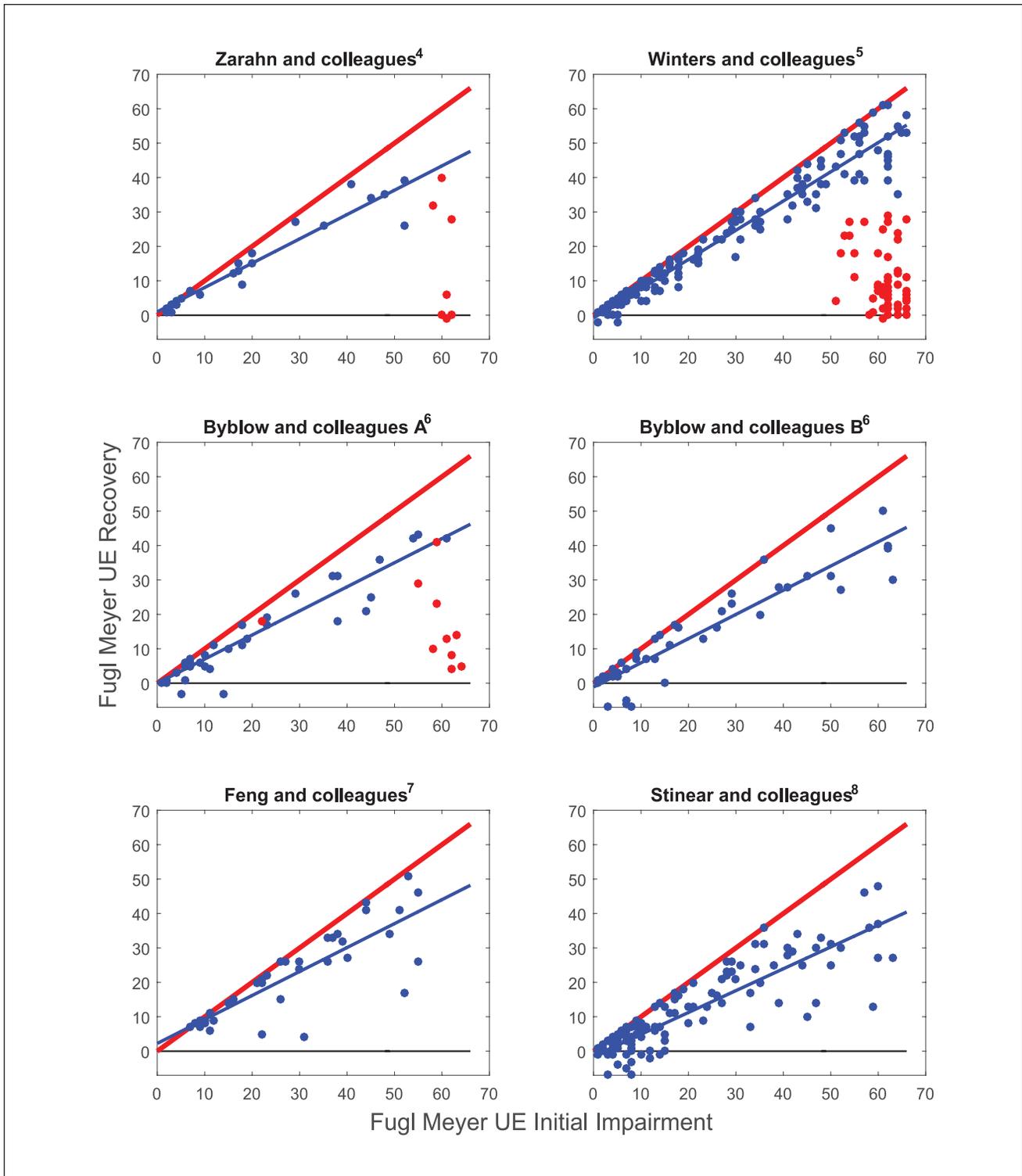
Figure 3. Overview of publications replicating and extending the results of the proportional recovery rule (PRR) proposed by Prabhakaran et al.[3] Red line: Ceiling line. Blue dots: Recoverers. Red dots: Nonrecoverers. Blue line: Linear regression line as computed from the measured datapoints. The data points were measured from the following table and figures in the respective articles: Zarahn et al[4] (Table 1), Winters et al[5] (Figure 2), Byblow et al (A and B)[6] (Figure 1e, Figure 3a), Feng et al[7] (Figure 5), Stinear et al[8] (Figure A); UE, upper extremity subscale.

After the initial formulation of the PRR, several publications sought to replicate and extend the results[4-8] (Figure 3 and Table 1). All publications looked at first-time ischemic stroke survivors, with Stinear et al[8] also considering hemorrhagic and recurrent stroke survivors. These studies all distinguish between recoverers and nonrecoverers although methods for identifying these groups differed; this has an impact on the interpretation of the data and will be discussed in later sections. In the following, we compare studies based only on models and findings for the recoverers.

In early work,[3-5] as the formulation of the PRR was proposed and refined, different methods were used to calculate the recovery proportion; it was only after Byblow et al[6] that the method using a linear regression with $FMA\text{-}UE_{ii}$ as the independent variable and $\Delta FMA\text{-}UE$ as the dependent variable solidified as an approach.[7,8] Building on the seminal work described above,[3] Zarahn et al[4] postulated a proportional model that estimated the conditional expectation of $\Delta FMA\text{-}UE$ given $FMA\text{-}UE_{ii}$ using a hierarchical framework. They assumed the population was comprised of a mixture of recovery groups and identified three subgroups with distinct recovery proportions. Winters et al[5] used a linear regression in which the response was predicted recovery (based on the PRR with the proportion set to 0.7) and the predictor was observed recovery.

Across these articles, the reported recovery proportion appears quite close to the proposed 70% (see Table 1), with some exceptions. After implementing their hierarchical model, Zarahn et al[4] report three different estimates (0.55, 0.81, and 0.93), from groups with close to equal size in the population. These estimates may suggest some underlying variability in the recovery proportion across subjects, or they could reflect a degree of heteroscedasticity in the error structure of the model (or both). Winters et al[5] report a 78% proportion in their linear regression; translating this to our standardized version of the PRR using data extracted from figures suggests an estimated recovery proportion as high as 85%, which is noticeably higher than others.

Looking only at the estimated recovery proportion can mask issues related to goodness of fit; as we argue above, inspections of the underlying data or use of regression diagnostics can clarify whether a method is valid for a given dataset. Heuristically, we compare each of the datasets in Figure 3 to our simulated examples as a way to qualitatively assess the appropriateness of the PRR for this collection of published studies. Several studies plausibly resemble the canonical version of the PRR as simulated in model 1, including Zarahn et al,[4] Winters et al,[5] and Byblow et al,[6] although issues of heteroscedasticity and ceiling effects are present in some cases as well. Our simulated model 2 exaggerates the issue of heteroscedasticity; the studies for which this is most obviously an issue are Feng et al[7] and Stinear et al.[8] That patients with low initial impairment are close to ceiling is trivial, but in the canonical PRR, moderate and severely affected patients are not at ceiling. Especially in

Winters et al[7] and Feng et al,[7] the ceiling effect, exaggerated in our simulated Model 3, has a pronounced impact on the observed data. Finally, visual inspection of several datasets supports the conclusion that there are "recoverers" and "nonrecoverers" with the latter designation being strongly related to severe initial impairment.[4-8]

A more formal summary of the datasets, estimated recovery proportions, statistics regarding homoscedasticity, and other key points of the discussed articles can be found in Table 1.

PRR for Other Outcomes

Although the PRR was originally posed for the FMA-UE, it has been applied to outcomes beyond FMA-UE, with the implication that the PRR denotes a more general biological recovery process after stroke. We now consider some illustrative examples of these analyses on other outcomes.

Smith et al[9] and Veerbeek et al[10] presented recovery data for the Fugl-Meyer Assessment Lower Extremity subscale (FMA-LE) for survivors of ischemic,[9,10] hemorrhagic and recurrent strokes[9] (Figure 4). Among recoverers, these studies fit linear regressions and reported estimated recovery proportions of 74% and 64%, respectively. In Smith et al,[9] corticospinal tract integrity was assessed in a subset of patients but did not lead to a clear separation of patients into recovers and nonrecoverers; instead, data points with the $\Delta FMA\text{-}LE$ differing 4 or more points from the predicted value were classified as outliers. In Veerbeek et al,[10] nonrecoverers were identified through hierarchical clustering.

Winters et al[12] reported the PRR in the LCT for visuospatial neglect in first-ever ischemic stroke patients (Figure 4). The slope of the regression line was 0.97, and nonrecoverers were identified through hierarchical clustering. Lazar et al[11] reported a proportional relation for language recovery as measured using the WAB in patients suffering from a first-time ischemic stroke (Figure 4). No distinction was made between responders and nonresponders, and it is unclear how the reported slope of 70% is derived.

For non-FMA-UE domains, there is no dataset that closely resembles the canonical PRR in our simulated Model 1, although data in Winters et al[12] and Byblow et al[6] contain some features of this model. In the other data, the high variance and heteroscedasticity accentuated in our simulated model 2 is an issue. All datasets in Figure 4 are also subject to ceiling effects, with the data from Winters et al[12] resembling a variable heavily affected by the ceiling as simulated in model 3 most clearly.

Discussion

What the PRR Is and What It Is Not

The PRR is a linear regression describing population-level recovery of patients in the subacute phase after stroke; it

Table 1.

Article	Measure	No. of Patients	No. of NR, n (%)	Outlier detection	Classification NR	Calculation of PRR	Slope (%)	Heteroskedasticity	Endpoint (Months)
Prabhakaran et al[3]	FMA-UE	41	7 (17)			Reduction of a multivariate linear regression to a linear relation	70	na	3-6
Zarahn et al[4]	FMA-UE	94	26 (27)	Initial FMA-UE <10		Maximum likelihood estimate	(55, 81, 93)	$P = .016$	3
Winters et al[5] (a)	FMA-UE	211	65 (30)	Hierarchical clustering		Linear regression*[1]	85*[1]	$P < .001$	6
Byblow et al[6]	FMA-UE	48	10 (21)	Corticospinal tract integrity		Linear regression	70	$P = .127$	6
Byblow et al[6]	FMA-UE	45	0*[2]	Corticospinal tract integrity		Linear regression	68	$P = .052$	3
Byblow et al[6]	RMT	37	un	Corticospinal tract integrity		Linear regression	74	$P = .444$	6
Feng et al[7]	FMA-UE	76	un	Initial FMA-UE <11		Linear regression	70	$P = .012$	3
Stinear et al[8]	FMA-UE	157	21 (13)	Corticospinal tract integrity		Linear regression	63	$P < .001$	3
Smith et al[9]	FMA-LE	32	0	Corticospinal tract integrity		Linear regression	74	$P = .015$	3
Veerbeek et al[10]	FMA-LE	202	27 (13)	Hierarchical clustering		Linear regression	64	$P = .016$	6
Winters et al[12] (b)	LCT	90	10 (11)	Hierarchical clustering		Linear regression	97	$P = .371$	6
Lazar et al[11]	WAB	21	na	na		un	70	$P = .019$	3

Summary of the key attributes of the discussed articles. The calculations are based on the data points shown in Figures 3 and 4 taken from the respective articles. P-values below 0.05 reject the null assumption that the dataset is homoscedastic (Breusch-Pagan test).

*1: after transformation to the format described in the section "Model formulation and simulated examples"

*2: nonrecoverers were a priori excluded.

Abbreviations: FMA-UE, upper extremity subscale of the Fugl-Meyer Assessment; NR, nonrecoverers; PRR, proportional recovery rule; LCT, Letter Cancellation Test; WAB, Western Aphasia Battery; na, not applicable; un, unknown.
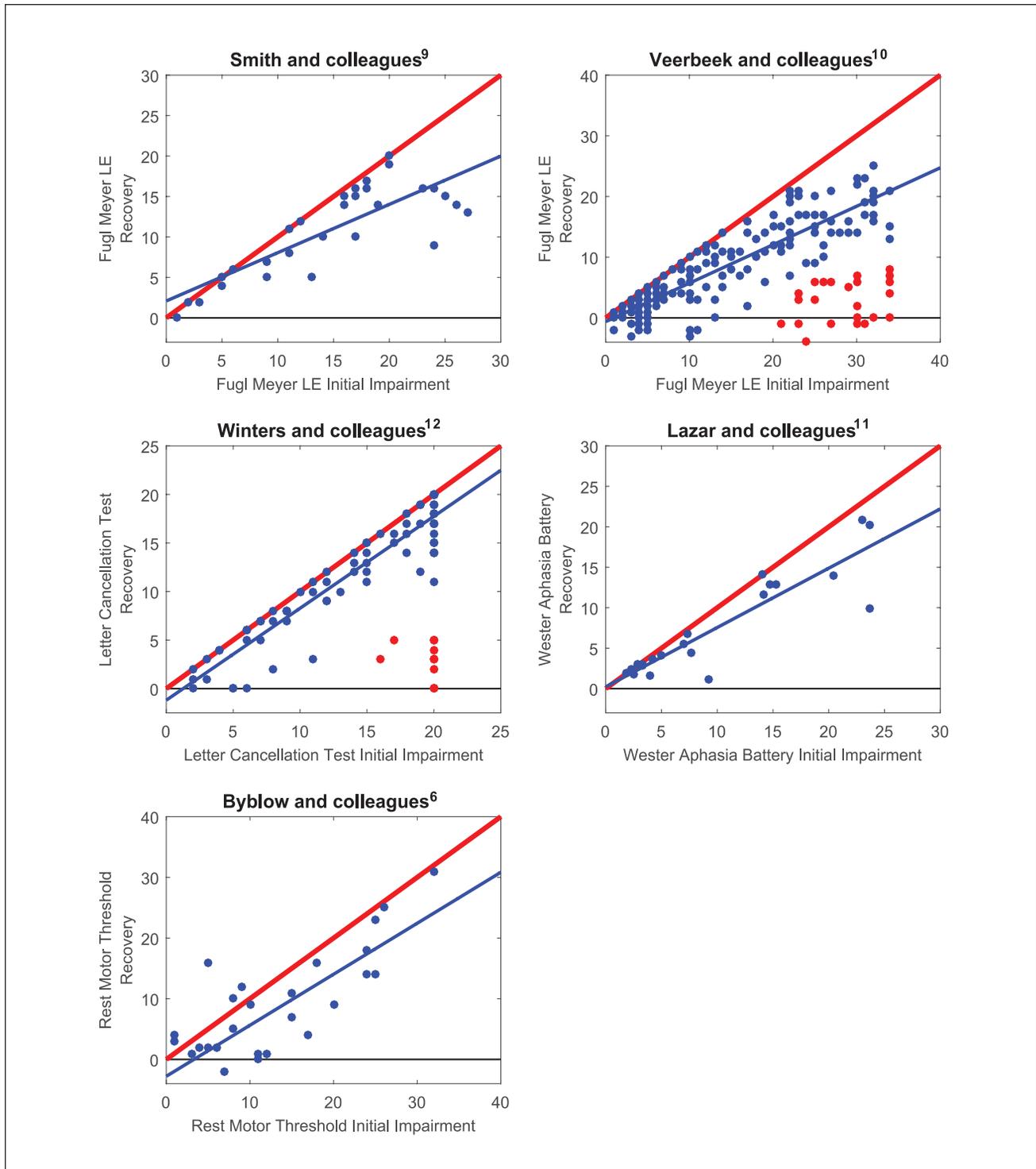
7

Figure 4. Proportional recovery rule (PRR) applied to lower extremity motor function, visuospatial neglect, aphasia and resting motor threshold. Red line: Ceiling line. Blue dots: Recoverers. Red dots: Nonrecoverers. Blue line: Linear regression line as computed from the measured datapoints. The data datapoints were measured from the following figures in the respective articles: Smith et al[9] (Figure B), Veerbeek et al[10] (Figure 2), Winters et al[12] (Figure 2), Lazar et al[11] (Figure), Byblow et al[6] (figure 2d).

was originally proposed for upper limb function measured with the FMA-UE but has since been extended to other domains. As outlined in the Introduction section, the PRR can be interpreted in a strictly statistical sense: from this perspective, the PRR is a particular model with parameters that can be estimated using standard tools. There are two major concerns. First, whether studies demonstrating or confirming the PPR have used appropriate statistical methods and paid sufficient attention to the data distribution and to associated underlying assumptions (eg, homoscedasticity)—failure to justify methods empirically or conduct adequate regression diagnostics may lead to unfounded or overly strong conclusions. Second, whether the implicit biological and clinical assertions of the PRR are supported by observed data—these are difficult to formulate precisely but generally relate to the idea that individual-level predictions made by the PRR should be meaningful.

In the rest of this section, we discuss remaining concerns related to the PRR and the implications of our literature review for biological questions about recovery. We also comment on recent critiques of the PRR; the most plausible of which address the subject-level accuracy of the PRR, in some detail.

Suggestions for Improved Analysis

A wide range of relationships might satisfy a statistical interpretation of the PRR, but only a narrower collection would satisfy the implicit biological and clinical expectations for the rule. It may therefore be helpful to acknowledge prediction accuracy as a distinct conceptual goal that can be assessed independently. With that in mind, we believe that support for the PRR would be bolstered through comparisons to other potential recovery models using out-of-sample prediction. Competing models might be naïve, such as using the out-of-sample mean change in impairment to predict recovery for new patients. More statistically complex methods might also be considered. A method may perform well relative to others and nonetheless provide poor overall prediction accuracy. For that reason, we also encourage authors to report quantities like the mean absolute prediction error for clinically meaningful patient populations (eg, mild, moderate, and severe initial impairment groups) or including prediction intervals, which combine uncertainty in parameter estimates with residual variance to provide the estimated recovery range for new patients, to improve on presenting only confidence intervals for the estimated slopes.

Throughout, we have focused on the PRR as it applies to "recoverers." This focus is predicated on the existence of biologically and statistically distinct subpopulations—those that recover according to the PRR in a way that allows reasonably accurate patient-level predictions, and those that do not recover predictably. A distinction of this

kind was originally made by Prabhakaran et al,[3] although it was framed in terms of initial severity: Initial impairment was strongly correlated with change in impairment for mild and moderate patients, but much less strongly correlated in severely affected patients. Building on the observation that some but not all severely affected patients recover as expected under the PRR led to the recoverer/ nonrecoverer partitions of the variety shown in Figure 3 (panel Winters et al[5]). Later work on the PPR for the FMA-UE suggested that in patients with initial severe impairments, nonrecoverers can be identified by examining integrity of the corticospinal tract, which has valuable implications for patient care.[6]

The methods that have been applied to identify nonrecoverers have been varied, which could influence results and hamper comparisons across studies. A useful thought experiment builds on the second simulated dataset. If one excludes "nonrecoverers" with high initial impairment and low recovery, the resulting data for "recoverers" may almost certainly be consistent with the PRR (in both the statistical and biological sense). Although nonrecoverers are not reported, the removal of outliers as in Smith et al[9] is questionable in the presence of heteroscedasticity as it might reshape the distribution of retained data. More concerningly, the direct application of more sophisticated analysis techniques (eg, hierarchical clustering or mixture modeling, as in Winters et al[5,12] and Veerbeek et al[10]) may effectively carry out the partitioning just described. Indeed, as shown by Hawe et al,[14] hierarchical clustering in data similar to our simulated model 2 can lead to patterns emergent solely from the properties of the data and methods that not reflect the PRR as an underlying mechanism. We emphasize that the possibility of these issues does not in itself invalidate the hypothesis of recoverers who follow the PRR and nonrecoverers who do not: The fact that one can generate data that reproduces some findings of the PRR does not mean that the PRR is invalid or that the observed data does not represent biologically meaningful associations. However, we do suggest rigor in identifying nonrecovers, as was done in the original PRR article[3] and as validated by analysis of motor-evoked potentials.[6]

Is Recovery Bounded and Domain-General?

To the extent that the 70% recovery rule has been supported by published studies, it has been interpreted by some as an upper limit to recovery given the current practice of rehabilitation medicine; hence, it has been suggested that there is substantial room for improvement for most patients that is currently not being achieved.[2] At the time of the publication of Krakauer and Marshall,[2] only slopes of around 70% were published, which made the assumption of a benchmark which has yet to be surpassed a reasonable proposition. Despite the statistical issues

discussed in the previous sections, the slope of 70% as hypothesized in the formulation of the PRR can still serve as a benchmark for different therapy interventions as proposed by Krakauer and Marshall.[2]

In the literature we reviewed, reported recovery proportions were generally near the 70% value hypothesized when the PRR was formed, but sometimes exceeded this value. In the articles by Winters et al,[5,12] the slopes appear to be somewhat higher, up to 90% or more, although variation in the applied analyses make direct comparisons difficult. These percentages are substantially different from 70% and cannot be ignored given the large sample size (N > 200). There are several potential explanations for the differences found in the slope, including a difference in the patient population, the possibility that patients received superior treatment, or possibly, systematic differences between study sites introducing the observed disparities. Regardless, the value of 70% as a recovery proportion is less important than whether recovery is systematic.

The PRR has also been thought by some to describe a domain-general underlying biological process of recovery. We first note that two (WAB, FMA-LE) out of four (WAB, RMT, visuospatial neglect, FMA-LE) measures which were not the FMA-UE either did not reliably distinguish between recoverers and nonrecoverers or did not report an identifiable cluster of patients recovering differently or poorly. For the WAB, no analysis of nonrecoverers was presented.[11] For the FMA-LE, Smith et al[9] found no identifiable subset of nonrecoverers based on corticospinal tract integrity, while Veerbeek et al[10] found a subset of nonrecoverers for the FMA-LE based on hierarchical clustering. Veerbeek et al[10] used the same data as Winters et al.[5] Interestingly, all the nonrecoverers found in the FMA-LE were also nonrecovers in the FMA-UE. However, only around 30% of the FMA-UE nonrecoverers were FMA-LE nonrecoverers. Veerbeek et al[10] proposed that this might be due to greater redundancy in descending pathways for lower extremities compared with upper extremities. For the RMT, it is not known if there is an identifiable cluster since the non-recoverers were excluded a priori. The lack of obvious nonrecoverers in the WAB, FMA-LE, and RMT could be viewed as evidence against a unified underlying biological process, although a systematic recovery process in recoverers (like the PRR) is likely attributable to a distinct mechanism from the reason for the existence of nonrecoverers. In the discussed literature, data for the FMA-UE more often resembled the "canonical" PRR than did other outcomes. This is perhaps unsurprising, because FMA-UE is the outcome the PRR was intended for in initial work. Overall, while there is some limited evidence for a domain-general recovery process, it appears that additional conceptual and methodological work is needed to draw conclusions about a domain-general recovery mechanism.

Recent Concerns About the PRR

As addressed above, the PRR can be seen as a statistical tool to understand population-level mechanisms and/or make subject-level predictions. In recent publications, Hope et al[13] and Hawe et al[14] raise concerns mainly concerned with single-subject-level prediction properties of the PRR.

The central argument of Hope et al[13] is that correlations between initial impairment and recovery, defined as the *change* between initial and follow-up values, are "spurious when (nontrivially) stronger than correlations between initial impairments and follow-up values." The authors raise this issue because of the common but inaccurate assumption that strong correlations between the former will necessarily imply strong correlations in the latter. Indeed, Hawe et al[14] seem to be making the same point when they say, "In theory, if proportional recovery can accurately predict change, it should also be able to accurately predict final score, since they are intrinsically linked." The scenario Hope et al[13] are most concerned with is one in which initial values have much higher variability than follow-up values, which is common in studies of stroke recovery. For example, if initial FM values are disturbed uniformly over [0, 66] and follow-up values are uniformly distributed over [60, 66] independently of initial values, one will observe strong correlations between initial impairment and recovery but a correlation approximately 0 between baseline and follow-up. This scenario would reflect a dataset where most patients recover close to the ceiling as in our simulated Model 3. This would still describe a proportional relationship with a slope close to 100% and therefore suggest a uniform prediction that each patient recovers almost completely from their initial impairment.

In contrast to the broader statistical concerns we have raised, Hope et al[13] are more explicitly focused on the ability to accurately predict individual patient outcomes based on initial values (and, potentially, other baseline data). They also emphasize correlations between variables of interest and the statistical properties of those correlations, rather than viewing the PRR as a linear regression subject to concerns like heteroscedasticity, nonlinearity, and width of confidence and prediction intervals. Hope et al[13] "[do not claim] that the proportional recovery rule is wrong", but also suggest that "empirical studies to date do not demonstrate that the rule holds." They propose a reevaluation of existing data and standard reporting for studies moving forward, keeping in mind the scenario described above and considering alternative hypotheses for the biological mechanisms underlying recovery. These are points we generally agree with, although we think that for the FMA-UE a case can be made that echoes our own more regression-minded suggestions in previous sections.

Hawe et al[14] emphasize the issue of mathematical coupling, which is a well-known source of concern in the

statistical literature. Briefly, coupling arises when one variable is directly or indirectly included in a second; correlations or associations between these may reflect their nonindependence rather than a "true" underlying relationship. They, and Hope et al,[13] mention the canonical example in which independent variables with equal variance are used to create "baseline" and "difference" variables; the correlation between the variables derived from independent samples will be approximately −0.71. In the context of the PRR, this issue appears to arise because the response (recovery) is computed in a way that seems to directly include the predictor (initial impairment).

Krakauer and Marshall[2] addressed this canonical version of mathematical coupling in the context of the PRR as applied to FMA-UE. In short, they argue that because (*a*) baseline and follow-up measurements assess a stable biological impairment and (*b*) FMA-UE has been shown to have relatively low measurement error, the canonical example of mathematical coupling is not a large concern in this setting. As a contrasting example, blood pressure may vary widely within a person in a short period of time and is subject to relatively large measurement error, both of which are more consistent with the canonical example of coupling. Hope et al[13] acknowledge the counterargument in Krakauer and Marshall,[2] and then emphasize the scenario described above as a source of "spurious" correlations; perhaps tellingly, a previous version of the article posted online was much more concerned about mathematical coupling in the canonical sense.

Hawe et al,[14] after introducing mathematical coupling as proposed by the canonical example, focus largely on simulations analogous to our second scenario. In that setting, the initial impairment creates an upper bound for possible recovery. This is a much more indirect example of coupling than one in which baseline is subtracted from follow-up to create recovery, and the statistical implications are much less clear. Instead of a derivation for the expected correlation between initial impairment and follow-up, Hawe et al[14] apply hierarchical clustering to the simulated datasets to partition recovers and nonrecoverers. The results of these analyses are bimodal distributions for estimated slopes and R^2 values with one mode roughly corresponding to the PRR. To the extent that these simulations quantify concerns raised above regarding identifying recoverers and nonrecoverers in highly heteroskedastic datasets with ceiling effects, we find them informative and useful. We do not, however, find these results compelling as an argument against findings based on applications of the PRR: The ability to produce results similar to the PRR via simulation does not preclude the applicability of the PRR as a linear regression model in real datasets. Indeed, as we have seen, while the statistical issues in our simulated model 2 are apparent in some cases, there are many datasets in which they are

not. Moreover, although Hawe et al[14] glancingly acknowledge that corticospinal tract integrity is a likely biomarker for nonrecoverers they do not discuss how this relates to their simulations[14].

After these simulations, Hawe et al[14] examine real datasets similar to (and overlapping somewhat with) those we have presented. Rather than arguing that these datasets sometimes exhibit the concerns raised in their preceding simulations (ie, that evidence in support of the PRR is possibly the result of heteroscedasticity and improper identification of nonrecoverers), they focus on new issues. First, they note that variability in observed recovery for individual patients is high, and that nearly half had recovery over 80%. Second, they note that the goodness of fit of the PRR is lower than the goodness of fit for a regression of follow-up values on initial impairment. These are important observations regarding (*a*) the ability of the PRR to explain observed recovery for each patient and (*b*) the difference between predicting recovery and predicting follow-up values. We agree that these should be considered both in reviewing the existing literature and in evaluating the PRR in future studies. We disagree that these observations are, in themselves, indictments of the PRR as a statistical and biological model for recovery.

These two critical articles[13,14] are valuable investigations into issues that can arise in studies of recovery and echo long-standing statistical concerns about relating baseline scores to changes. Our disagreements with these articles fall in two principal areas. First, we hold that models for recovery such as the PRR can be informative in themselves and do not depend on patient-level predictions, although certainly models that can do both are preferable. Patient-level prediction is the issue that both Hope et al[13] and Hawe et al[14] appear primarily interested in. Second, we view many (but not all) existing studies as largely consistent with the PRR as a statistical and biological model; Hope et al[13] are relatively agnostic on this point, but Hawe et al[14] conclude evidence supporting the 70% PRR is "too good to be true." We find the former article a good deal more convincing than the latter, which appears to be on somewhat of a crusade.

Conclusions

With the preceding concerns in mind, and using simulated datasets as points of reference, we critically reexamined published applications of the PRR. Visual inspection and formal analysis of reported data are often, although not uniformly, consistent with the "canonical" example of the PRR; see, for example, Figure 3 (panels Zarahn et al,[4] Winters et al,[5] Byblow et al[6] A and B). Other cases are suggestive of the simulated datasets featuring heteroscedastic errors and ceiling effects; examples include data in Figure 3 (panels Feng et al,[7] Stinear et al[8]). Inconsistencies in the

methods used in reporting the PRR make it difficult to draw conclusions across studies; in cases where ceiling effects and/or heteroscedasticity of variance of residuals in the underlying data are an issue, the validity of the linear regressions has to be questioned. In our view, these results neither conclusively demonstrate the existence of a universal PRR applicable across neurological modalities (which was never the claim in the original two articles[3,4]), nor do they refute the PRR and its usefulness in at least some settings. Instead, many of the examples support the PRR as a statistically and biologically meaningful model for spontaneous recovery, especially with regard to the FMA-UE. That said, we agree that caution is required with regard to other measures and non-motor impairments.

Future applications of the PRR should be conducted in a statistically uniform way, consistently using best practices for evaluating linear regression models, and include quantitative comparisons to alternative models of recovery to assess validity. More nuanced experimental and statistical methods will be needed to clarify the biological mechanisms involved in the recovery process. This is important, because at the very least there does seem to be a systematic nonartifactual relationship between initial impairment and the motor recovery process (ΔFMA-UE). Combining multiple datasets across sites could help to strengthen the arguments for the (non)existence of a PRR and may also reveal interesting differences in the effectiveness of rehabilitation in different settings. While the PRR was not intended to inform patients about their recovery potential or derive subject-level predictions, it has been implicitly assumed that the PRR should be useful in this way. This has emerged as an important consideration; in light of the heteroscedasticity in many of the underlying datasets, single subject-level recovery prediction should be evaluated as a distinct goal moving forward.

Declaration of Conflicting Interests

Funding

ORCID iD

Robinson Kundert   https://orcid.org/0000-0001-5162-0742

References

1. Duncan PW, Goldstein LB, Matchar D, Divine GW, Feussner J. Measurement of motor recovery after stroke. Outcome assessment and sample size requirements. *Stroke*. 1992;23:1084-1089.
2. Krakauer JW, Marshall RS. The proportional recovery rule for stroke revisited. *Ann Neurol*. 2015;78:845-847. doi:10.1002/ana.24537
3. Prabhakaran S, Zarahn E, Riley C, et al. Inter-individual variability in the capacity for motor recovery after ischemic stroke. *Neurorehabil Neural Repair*. 2008;22:64-71. doi:10.1177/1545968307305302
4. Zarahn E, Alon L, Ryan SL, et al. Prediction of motor recovery using initial impairment and fMRI 48 h poststroke. *Cereb Cortex*. 2011;21:2712-2721. doi:10.1093/cercor/bhr047
5. Winters C, van Wegen EE, Daffertshofer A, Kwakkel G. Generalizability of the proportional recovery model for the upper extremity after an ischemic stroke. *Neurorehabil Neural Repair*. 2015;29:614-622. doi:10.1177/1545968314562115
6. Byblow WD, Stinear CM, Barber PA, Petoe MA, Ackerley SJ. Proportional recovery after stroke depends on corticomotor integrity. *Ann Neurol*. 2015;78:848-859. doi:10.1002/ana.24472
7. Feng W, Wang J, Chhatbar PY, et al. Corticospinal tract lesion load: an imaging biomarker for stroke motor outcomes. *Ann Neurol*. 2015;78:860-870. doi:10.1002/ana.24510
8. Stinear CM, Byblow WD, Ackerley SJ, Smith MC, Borges VM, Barber PA. Proportional motor recovery after stroke: implications for trial design. *Stroke*. 2017;48:795-798. doi:10.1161/STROKEAHA.116.016020
9. Smith MC, Byblow WD, Barber PA, Stinear CM. Proportional recovery from lower limb motor impairment after stroke. *Stroke*. 2017;48:1400-1403. doi:10.1161/STROKEAHA.116.016478
10. Veerbeek JM, Winters C, van Wegen EEH, Kwakkel G. Is the proportional recovery rule applicable to the lower limb after a first-ever ischemic stroke? *PLoS One*. 2018;13:e0189279. doi:10.1371/journal.pone.0189279
11. Lazar RM, Minzer B, Antoniello D, Festa JR, Krakauer JW, Marshall RS. Improvement in aphasia scores after stroke is well predicted by initial severity. *Stroke*. 2010;41:1485-1488. doi:10.1161/STROKEAHA.109.577338
12. Winters C, van Wegen EE, Daffertshofer A, Kwakkel G. Generalizability of the maximum proportional recovery rule to visuospatial neglect early poststroke. *Neurorehabil Neural Repair*. 2017;31:334-342. doi:10.1177/1545968316680492
13. Hope TMH, Friston K, Price CJ, Leff AP, Rotshtein P, Bowman H. Recovery after stroke: not so proportional after all? *Brain*. 2019;142:15-22. doi:10.1093/brain/awy302
14. Hawe RL, Scott SH, Dukelow SP. Taking proportional out of stroke recovery. *Stroke*. 2019;50:204-211. doi:10.1161/STROKEAHA.118.023006