

# Predicting Upper Limb Motor Impairment Recovery after Stroke: A Mixture Model

Rick van der Vliet, MD,<sup>1,2</sup> Ruud W. Selles, PhD<sup>1,2,3,4</sup>, Eleni-Rosalina Andrinopoulou, PhD,<sup>5</sup> Rinske Nijland, PhD,<sup>6,7</sup> Gerard M. Ribbers, MD, PhD,<sup>2,4</sup> Maarten A. Frens, PhD,<sup>1</sup> Carel Meskers, MD, PhD,<sup>6</sup> and Gert Kwakkel, PhD<sup>6,7,8</sup>

**Objective:** Spontaneous recovery is an important determinant of upper extremity recovery after stroke and has been described by the 70% proportional recovery rule for the Fugl–Meyer motor upper extremity (FM-UE) scale. However, this rule is criticized for overestimating the predictability of FM-UE recovery. Our objectives were to develop a longitudinal mixture model of FM-UE recovery, identify FM-UE recovery subgroups, and internally validate the model predictions.

**Methods:** We developed an exponential recovery function with the following parameters: subgroup assignment probability, proportional recovery coefficient  $r_k$ , time constant in weeks  $\tau_k$ , and distribution of the initial FM-UE scores. We fitted the model to FM-UE measurements of 412 first-ever ischemic stroke patients and cross-validated endpoint predictions and FM-UE recovery cluster assignment.

**Results:** The model distinguished 5 subgroups with different recovery parameters ( $r_1 = 0.09$ ,  $\tau_1 = 5.3$ ,  $r_2 = 0.46$ ,  $\tau_2 = 10.1$ ,  $r_3 = 0.86$ ,  $\tau_3 = 9.8$ ,  $r_4 = 0.89$ ,  $\tau_4 = 2.7$ ,  $r_5 = 0.93$ ,  $\tau_5 = 1.2$ ). Endpoint FM-UE was predicted with a median absolute error of 4.8 (interquartile range [IQR] = 1.3–12.8) at 1 week poststroke and 4.2 (IQR = 1.3–9.8) at 2 weeks. Overall accuracy of assignment to the poor (subgroup 1), moderate (subgroups 2 and 3), and good (subgroups 4 and 5) FM-UE recovery clusters was 0.79 (95% equal-tailed interval [ETI] = 0.78–0.80) at 1 week poststroke and 0.81 (95% ETI = 0.80–0.82) at 2 weeks.

**Interpretation:** FM-UE recovery reflects different subgroups, each with its own recovery profile. Cross-validation indicates that FM-UE endpoints and FM-UE recovery clusters can be well predicted. Results will contribute to the understanding of upper limb recovery patterns in the first 6 months after stroke.

ANN NEUROL 2020;00:1–11

Longitudinal studies have repeatedly demonstrated the time-dependency of neurological recovery after stroke, including upper<sup>1,2</sup> and lower limb motor function,<sup>3,4</sup> visuo-spatial neglect,<sup>5</sup> and speech.<sup>6</sup> This suggests that recovery follows a predictable pattern, which is often described as spontaneous neurological recovery.<sup>7,8</sup> Understanding the mechanisms and individual dynamics that drive stroke recovery is vital for developing better

prognostic models and more effective, personalized therapeutic interventions.<sup>9–12</sup>

The proportional recovery rule has been instrumental in modeling spontaneous upper extremity recovery by linking baseline motor impairment, measured with the Fugl–Meyer assessment of the upper extremity (FM-UE),<sup>13</sup> to the observed motor recovery, defined as the difference between the measurements early and 3 to 6 months

View this article online at [wileyonlinelibrary.com](http://wileyonlinelibrary.com). DOI: 10.1002/ana.25679

Received Jan 18, 2019, and in revised form Jan 8, 2020. Accepted for publication Jan 8, 2020.

Address correspondence to Dr van der Vliet, Erasmus MC Room EE-1202, PO Box 2040, 3000 CA Rotterdam, the Netherlands.

E-mail: [r.vandervliet@erasmusmc.nl](mailto:r.vandervliet@erasmusmc.nl)

From the <sup>1</sup>Department of Neuroscience, Erasmus University Medical Center, Rotterdam, the Netherlands; <sup>2</sup>Department of Rehabilitation Medicine, Erasmus University Medical Center, Rotterdam, the Netherlands; <sup>3</sup>Department of Plastic and Reconstructive Surgery, Erasmus University Medical Center, Rotterdam, the Netherlands; <sup>4</sup>Rijnland Rehabilitation Center, Rotterdam, the Netherlands; <sup>5</sup>Department of Biostatistics, Erasmus University Medical Center, Rotterdam, the Netherlands; <sup>6</sup>Department of Rehabilitation Medicine, Amsterdam University Medical Centre, VU University Medical Center, Amsterdam Neurosciences and Amsterdam Movement Sciences, Amsterdam, the Netherlands; <sup>7</sup>Department of Physical Therapy and Human Movement Sciences, Northwestern University, Chicago, IL; and <sup>8</sup>Department of Neurorehabilitation, Amsterdam Rehabilitation Research Centre Reade, Amsterdam, the Netherlands

after stroke ( $\Delta$ FM-UE).<sup>14</sup> More specifically, the proportional recovery rule states that in 3 to 6 months, (1) the majority of patients (recoverers) gain a fixed proportion, estimated between 0.55 and 0.85,<sup>2</sup> of their potential recovery, calculated as the difference between baseline FM-UE and the scale's maximum score of 66, and (2) the minority of patients (nonrecoverers) show only very moderate improvement, which cannot be linked to potential recovery.<sup>1,2,14</sup> Mechanistically, the key underlying difference between recoverers and nonrecoverers is currently understood as the intactness of the corticospinal tract early after stroke.<sup>15–18</sup>

The proportional recovery rule has been criticized for a number of reasons. Recent analyses indicated that a strong correlation between baseline FM-UE and recovery can emerge even when baseline FM-UE is completely uncorrelated to endpoint FM-UE.<sup>19,20</sup> Therefore, even though the proportional recovery rule is not wrong,<sup>19</sup> it probably overstates the predictability of endpoint FM-UE.<sup>19,20</sup> In addition, the proportional recovery rule does not model the time course of recovery early poststroke, which means it cannot model the rate of recovery nor update predictions with repeated measurements in time. Finally, predictions of endpoint FM-UE based on the 70% proportional recovery rule for individual patients have not previously been reported.

To increase our understanding of upper extremity recovery after stroke, we need a model that relates the FM-UE to potential recovery as a function of time after stroke, with separate sets of parameters for different subgroups, including those that show no improvement early poststroke.<sup>21</sup> In this study, we developed and cross-validated a new longitudinal mixture model of FM-UE recovery that describes different patterns of recovery over time using exponential functions and identifies subgroups based on (1) the degree of recovery as a fraction of potential recovery, (2) the rate of recovery, and (3) the initial FM-UE score. Our goals were to estimate the number of recovery subgroups, the recovery parameters for each subgroup, and the predictability of endpoint FM-UE at 3 to 6 months poststroke, as well as subgroup assignment as a function of time poststroke. Results will contribute to the understanding and prediction of upper limb recovery patterns in the first 6 months after stroke.

## Patients and Methods

### Study Population

We combined FM-UE data of first-ever ischemic stroke patients collected in 4 different prospective cohort studies: the EXPLICIT,<sup>22</sup> EPOS,<sup>23</sup> 4D-EEG,<sup>24</sup> and EXPLORE studies. These data sets contain repeated measurements of the FM-UE scores and the exact measurement dates in days

poststroke, which also differ between patients assigned to the same follow-up scheme for practical reasons. Data collection and patient characteristics of the EXPLICIT and EPOS cohorts have been described extensively elsewhere.<sup>22,23</sup> The 4D-EEG and EXPLORE cohorts recruited patients with a first-ever ischemic stroke within 3 weeks poststroke. In the 4D-EEG study, patients were measured weekly during the first 5 weeks poststroke and after 8, 12, and 26 weeks. In the EXPLORE study, patients were measured 1, 2, 3, 5, 12, and 26 weeks poststroke. Inclusion criteria were comparable to the EPOS cohort. The majority of patients received standard rehabilitation treatment according to the Dutch rehabilitation guidelines, which are in agreement with current international rehabilitation guidelines.<sup>25,26</sup> In the EXPLICIT study, half of the patients with an unfavorable prognosis received electromyography-triggered neuromuscular stimulation, and half of the patients with a favorable prognosis received modified constrained-induced movement therapy.<sup>22</sup> Because both of these interventions did not affect the FM-UE at any time poststroke,<sup>22</sup> we disregarded therapeutic intervention as a factor in the analysis. The 4D-EEG and EXPLORE studies have been approved by the medical ethics committees of the VU University Medical Center (NL 47079 029 14, for 37 patients measured) and the Leiden University Medical Center (NL39323.058.12, for 11 patients measured), respectively.

We included a patient if at least 2 repeated measurements were available and if the first and last measurement were at least 12 weeks apart. This way, we maximized the number of included patients while still being able to cross-validate predictions of endpoint FM-UE. Additional patient data were age, gender, handedness, and dominant side affected; Bamford scale (lacunar anterior circulation infarct/partial anterior circulation infarct/total anterior circulation infarct);<sup>27</sup> administration of alteplase (recombinant tissue-type plasminogen activator); National Institutes of Health Stroke Scale (NIHSS; range, 0–42) with item 11, extinction and inattention (range, 0–2), reported separately;<sup>28</sup> motricity index (range, 0–99) with the shoulder abduction item listed separately (dichotomized as no shoulder abduction [0] and at least some shoulder abduction [1]);<sup>29,30</sup> and finger extension (dichotomized as no finger extension [0] and at least some finger extension [1]) as a separate item of the FM-UE (range, 0–66).<sup>31</sup>

### Longitudinal Mixture Model of FM-UE Recovery

We designed a longitudinal model of FM-UE recovery after stroke based on the principles of proportional recovery, which are (1) a proportional relation between observed recovery over time and potential recovery at baseline (longitudinal), and (2) the existence of clinically distinct subgroups of FM-UE recovery (mixture). Longitudinal,

therefore, refers to the ability of the model to handle repeated measurements over time, and mixture refers to the ability of the model to identify different subgroups. Because FM-UE recovery follows an exponential pattern,<sup>7</sup> we chose an exponential function as the time-dependent element of the model, with the asymptote defined as a proportion of the potential recovery and the time constant expressed in weeks. In addition, we included an intercept which represents the FM-UE early after stroke. The mathematical expression of our model is:

$$\mu_{ij|k} = \alpha_{i|k} + r_k * (66 - \alpha_{i|k}) * (1 - e^{-t_{ij}/\tau_k}) \quad (1)$$

$$y_{ij|k} \sim N(\mu_{ij|k}, \sigma_e^2) \quad (2)$$

Here,  $i$  is the patient identification number (1  $I$ ),  $j$  is the measurement identification number (1  $J$ ), and  $k$  is the subgroup identification number (1  $K$ ). The equation describes how the FM-UE ( $y_{ij|k}$ ) for a particular patient  $i$  and measurement  $j$  is determined by the (estimated) baseline FM-UE ( $\alpha_{i|k}$ ) plus an exponential term  $r_k * (66 - \alpha_{i|k}) * (1 - e^{-t_{ij}/\tau_k})$ , which increases over time  $t_{ij}$  as the patient recovers. We chose to express measurement dates poststroke in weeks by dividing the number of days poststroke by 7. The asymptote of the exponential term is determined by the potential recovery ( $66 - \alpha_{i|k}$ ) multiplied by the recovery coefficient  $r_k$  (0 1), which describes how much of the potential recovery is achieved. The rate of the exponential term (ie, how quickly the patient recovers) is defined by time constant  $\tau_k$  in weeks (1/7  $\infty$ ), which signifies the time point when recovery has reached a proportion of  $1 - e^{-1} \approx 0.63$  of the asymptotic value. Finally,  $\sigma_e^2$  is the residual error variance.

### Model Fitting

We chose a Bayesian approach to mixture modeling rather than expectation maximization, as Bayesian data analysis (1) focuses on parameter uncertainty rather than on point estimates, (2) estimates hidden variables (for example the subgroup identification number  $k$ ) simultaneously with the parameters, and (3) offers flexibility in specifying the form of the model (for example to constrain the recovery coefficient  $r_k$  between 0 and 1).<sup>32</sup> Modern Bayesian approaches rely on a family of algorithms called the Markov-chain Monte-Carlo (MCMC) algorithms.<sup>32</sup> These algorithms require defining a likelihood function (how the data would be generated if we knew the parameters) and the prior probability distributions for the parameters, and they return samples from the posterior joint-probability function of the parameters. We chose the following prior probability distributions for the model parameters:

$$\alpha_{i|k} \sim \frac{66}{1 + \exp(-N(\mu_{\alpha,k}, \sigma_{\alpha,k}^2))} \quad (3)$$

$$\mu_{\alpha,k} \sim N(0, 10^3) \quad (4)$$

$$\sigma_{\alpha,k}^2 \sim 1/\Gamma(10^{-3}, 10^{-3}) \quad (5)$$

$$r_k \sim \frac{1}{1 + \exp(-N(0, 10^3))} \quad (6)$$

$$\tau_k \sim \Gamma(10^{-3}, 10^{-3}) + 1/7 \quad (7)$$

$$k \sim \text{Cat}(K, p_k), p_k \sim \text{Dirichlet}(K, \gamma) \quad (8)$$

$$1/\sigma_e^2 \sim \Gamma(10^{-3}, 10^{-3}) \quad (9)$$

For the patient-specific baseline FM-UE  $\alpha_{i|k}$ , we defined a logistic normal prior distribution with the hyperparameters sampled from weakly informative normal and gamma distributions. This means that each subgroup is characterized by a specific distribution of the FM-UE early after stroke, which can be close to 0 or to the maximum of 66 or span the entire range with almost equal probability. The subgroup-specific prior distribution for the recovery coefficient  $r_k$  is also a logistic normal distribution, which spans the 0 to 1 range. Time constant  $\tau_k$ , specified separately for each subgroup, has a weakly informative gamma prior distribution, shifted by 1/7 to set the lower limit at 1 day. Subgroup labels  $k$  have a categorical prior distribution with hyperparameters for the subgroup assignment probability vector  $p_k$ , sampled from a Dirichlet distribution with concentration parameter  $\gamma$ . Finally, the precision  $1/\sigma_e^2$  has a weakly informative gamma prior distribution.

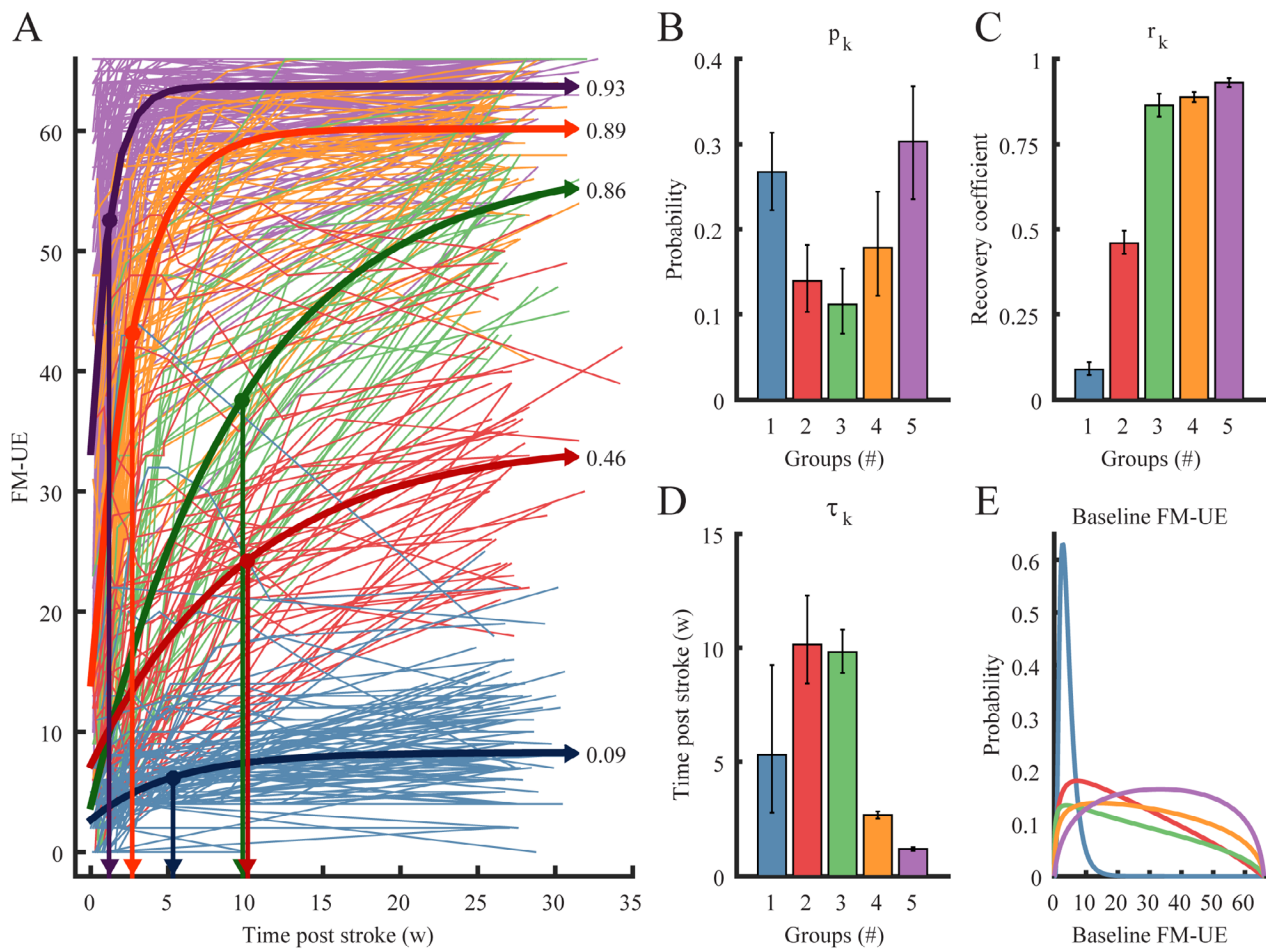
MCMC sampling was used to simultaneously calculate the number of subgroups in the data and the model parameters. We used the Rousseau and Mengersen<sup>33,34</sup> criterion as implemented by Nasserinejad et al<sup>35</sup> to select the number of subgroups  $K_{Optimal}$  setting the overfitted number of subgroups  $K$  at 10, the concentration parameter  $\gamma$  at  $0.9 * d/2$  (equal to 1.8 for our study), the cutoff value for the subgroup size at 5% of the number of patients, and the number of parallel chains to 10. From the parallel chains, we selected the solution which minimized the number of subgroups and maximized the total subgroup assignment probability. Subgroup assignment probabilities were normalized to 1. The subgroups were arranged according to the recovery coefficient  $r_k$ , making  $r_1$  the lowest and  $r_{K, optimal}$  the highest recovery coefficient. The ‘‘optimal FM-UE recovery cluster’’ was determined as the FM-UE recovery cluster a patient was assigned to most by the model. Goodness of fit was evaluated with the explained variance, which we calculated as 1 minus

the residual error variance ( $\sigma_e^2$ ) divided by the total FM-UE variance across patients and measurements.

### Cross-Validation

Predictability of  $\Delta$ FM-UE (the difference between the first and last measurements available for a particular patient) and endpoint FM-UE (last measurement available for a particular patient), as well as FM-UE recovery cluster assignment (poor, moderate, or good recovery; see Results section for the definitions), were estimated using the proposed model. We used cross-validation, which is a method for internal validation, to obtain correct estimates of the predictions. The study population was divided  $n$  (total number of patients) times into a prediction data set containing data from only 1 patient and a fitting data set containing data from all other patients. For all  $n$ -folds, we first ran the fitting data set with settings  $K = K_{Optimal}$  and  $\gamma = 1.8$  and randomly selected 100 samples from the

posterior distribution of the model parameters. In addition, we paired the 5 subgroups with 1 of the 3 FM-UE recovery clusters using a 1-nearest neighbor algorithm trained on the model parameters  $r_k$ ,  $\tau_k$ ,  $\mu_{\alpha,k}$  and  $1/\sigma_e^2$ . Next, MCMC sampling was performed for all 100 model parameter sets using the measurements available from the prediction data set in the first 1 to 12 weeks poststroke (12 time points). Only patients who had at least 1 measurement available were included in the analysis for a specific time interval. Therefore, the number of patients available for cross-validation increased with time poststroke. Outcome measures were (1) the predicted  $\Delta$ FM-UE between the first and last measurements of a patient, (2) the predicted FM-UE at the last measurement of a patient, and (3) the “predicted FM-UE recovery cluster,” defined as the FM-UE recovery cluster a patient was assigned to most by the model.



**FIGURE 1: Longitudinal mixture model of Fugl-Meyer motor upper extremity (FM-UE) recovery. (A) FM-UE recovery data of the 412 ischemic stroke patients in our data set. Individual patients are color-coded according to the subgroup they were assigned to most by the longitudinal mixture model of FM-UE recovery. The average subgroup recovery patterns are shown in bold. Estimated model parameters for the 5 different subgroups: subgroup assignment probability (B), recovery coefficient (C), time constant (D), and initial distribution of the FM-UE (E). Whiskers indicate 95% equal-tailed intervals.**

**TABLE 1. Model Parameters**

FM-UE Recovery Cluster	Poor	Moderate			Good
Subgroup	1	2	3	4	5
$p_k$	0.27 (0.22–0.31)	0.14 (0.10–0.18)	0.11 (0.08–0.15)	0.18 (0.12–0.24)	0.30 (0.24–0.37)
$r_k$	0.09 (0.07–0.11)	0.46 (0.43–0.50)	0.86 (0.83–0.90)	0.89 (0.87–0.90)	0.93 (0.92–0.94)
$\tau_k$	5.3 (2.8–9.2)	10.1 (8.4–12.3)	9.8 (8.9–10.8)	2.7 (2.5–2.8)	1.2 (1.1–1.3)
$\mu_{\alpha,k}$	-3.2 (-4.0 to 2.8)	-2.1 (-2.9 to 1.2)	-2.8 (-4.1 to 1.3)	-1.3 (-2.6 to 0.1)	0.0 (-0.6 to 0.6)
$\sigma_{\alpha,k}$	0.6 (0.3–1.5)	2.2 (1.5–3.3)	3.0 (1.7–4.8)	2.9 (2.0–4.0)	2.4 (1.9–3.0)

Subgroup mean model parameters, with 95% equal-tailed intervals calculated over all samples given in parentheses.  $p_k$  = subgroup assignment probability;  $r_k$  = recovery coefficient;  $\tau_k$  = time constant in weeks;  $\mu_{\alpha,k}$  = mean of the initial distribution of the FM-UE in the logistic space;  $\sigma_{\alpha,k}$  = standard deviation of the initial distribution of the FM-UE in the logistic space.  
FM-UE = Fugl–Meyer motor upper extremity.

To evaluate prediction accuracy, we calculated the absolute difference between the predicted and observed values, correlated the predicted and observed  $\Delta$ FM-UE and FM-UE, and determined the accuracy of the FM-UE recovery cluster assignment (proportion of patients in the study population who were correctly assigned), the positive predictive value (proportion of patients in 1 of the 3 “predicted FM-UE recovery clusters” who were correctly assigned), and the miss rate (proportion of patients in 1 of the 3 “optimal FM-UE recovery clusters” who were

incorrectly assigned). Note that accuracy is only defined for the entire study population, whereas the positive predictive value and miss rate are defined for the 3 FM-UE recovery clusters separately.

### Covariate Model

We compared the predictive accuracy of the model presented here to a model incorporating a set of static (not changing over time) covariates: age at stroke onset, gender, Bamford classification, and alteplase treatment. The static

**TABLE 2. Baseline Patient Clinimetric Scores**

FM-UE Recovery Cluster	Poor	Moderate			Good
Subgroup	1	2	3	4	5
Patients, n	111 (97–120)	56 (49–66)	44 (37–57)	72 (54–94)	126 (104–146)
Age, yr	63 (42–93)	65 (43–86)	60 (28–85)	64 (38–85)	66 (33–86)
Male, %	56	58	53	53	47
Right-handed, %	90	89	92	90	95
Dominant hand affected, %	27	46	52	49	43
Bamford LACI/PACI/TACI, %	28/47/25	50/37/13	55/31/14	70/22/8	64/26/9
Alteplase treatment, %	29	18	24	15	18
NIHSS	13 (6–21)	8 (2–18)	9 (2–18)	5 (1–18)	5 (0–14)
Motricity index	5 (0–34)	28 (0–84)	23 (0–92)	55 (0–100)	66 (0–100)
Shoulder abduction, %	23	69	51	94	95
Finger extension, %	2	24	24	69	88

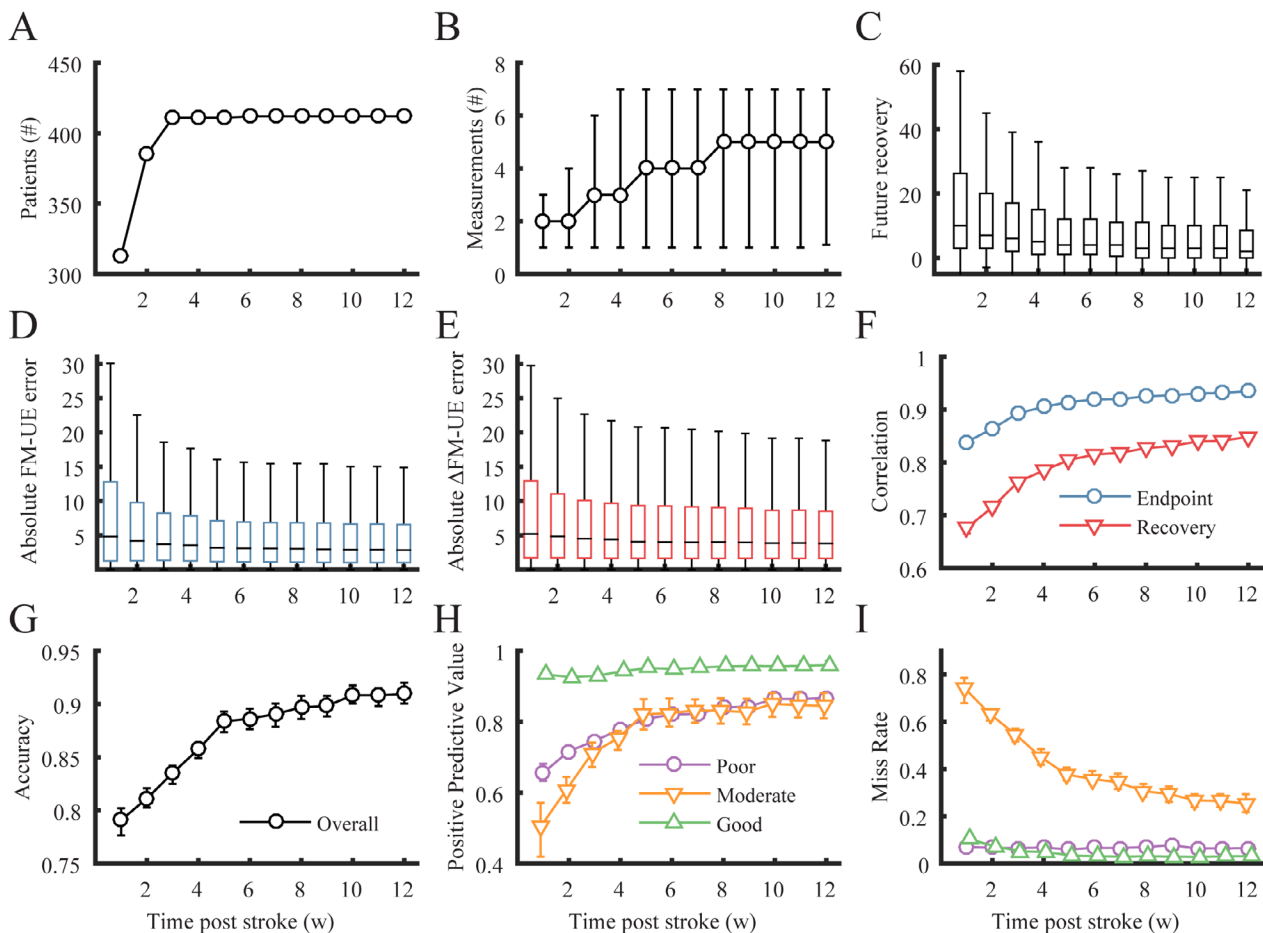
Subgroup mean clinimetric scores with 95% equal tailed intervals calculated per subgroup over all samples.

FM-UE = Fugl–Meyer motor upper extremity; LACI = lacunar anterior circulation infarction; NIHSS = National Institutes of Health Stroke Scale; PACI = partial anterior circulation infarction; TACI = total anterior circulation infarction.

covariates did not include right-handedness and dominant side affected, as these were biased by the inclusion criteria of the cohort studies. We modeled age as a normal distribution with hyperparameters sampled from normal and gamma distributions, gender and alteplase treatment as binomial distributions with the hyperparameters sampled from beta distributions, and Bamford classification as a categorical distribution with hyperparameters sampled from Dirichlet distributions. The cross-validated primary outcomes (absolute median error in endpoint FM-UE and  $\Delta$ FM-UE, correlations between actual and observed endpoint FM-UE and  $\Delta$ FM-UE, and mean accuracy of FM-UE recovery cluster assignment) of the models with and without covariates differed less than 10% at every time point poststroke. Therefore, we decided to present a simpler model without covariates.

**MCMC Sampling**

MCMC sampling was implemented in JAGS 4.3.0 (<https://sourceforge.net/projects/mcmc-jags/>). MATLAB 2015a (MathWorks, Natick, MA) and MATJAGS ([http://psiexp.ss.uci.edu/research/programs\\_data/jags/](http://psiexp.ss.uci.edu/research/programs_data/jags/)) were used for data and sample processing. Settings for determining the number of subgroups and calculating the model parameters were  $2.5 \times 10^4$  burn-in samples and  $2.5 \times 10^4$  posterior distribution samples, 10 parallel chains, and initial guesses for the model parameters. Settings for cross-validation were  $10^3$  burn-in samples,  $10^4$  posterior distribution samples, 1 parallel chain, and the mean model parameters estimated in step 1 as initial values for model fitting. All scripts can be accessed at <https://github.com/rickvandervliet/Bayesian-Proportional-Recovery>; this website also hosts scripts that can prospectively predict



**FIGURE 2: Cross-validation of model predictions.** (A) Number of patients who had at least 1 measurement at a specific time poststroke and were therefore included in the cross-validation. (B) Median number of measurements per patient available for cross-validation at a specific time poststroke. Error bars indicate 95% equal-tailed intervals [ETIs] across patients with at least 1 measurement. Whiskers represent 1.5 times the interquartile range; outliers not shown. (C) Future recovery, defined as endpoint Fugl-Meyer motor upper extremity (FM-UE) minus last available FM-UE for each patient at a specific time poststroke. (D, E) Boxplot of the absolute error across all 412 patients times 100 samplings of the endpoint FM-UE (A) and the  $\Delta$ FM-UE (B). Whiskers represent 1.5 times the interquartile range; outliers not shown. (F) Correlation between predicted and observed FM-UE (blue circles) and  $\Delta$ FM-UE (red triangles) with error bars indicating the 95% ETIs over the 100 samplings. FM-UE recovery cluster assignment accuracy (G), positive predictive value (H), and miss rate (I) with error bars indicating the 95% ETIs across the 100 samplings.

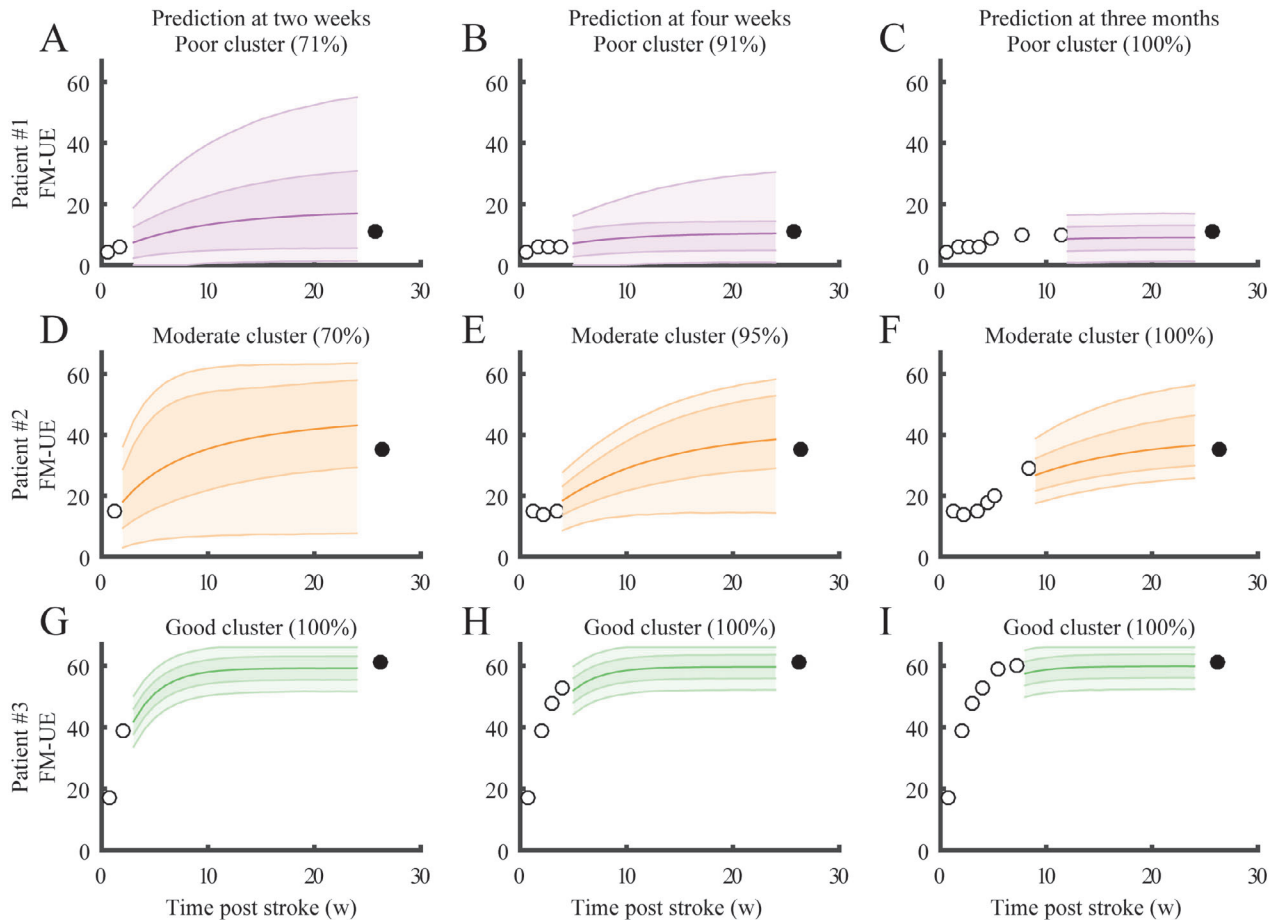
FM-UE recovery profiles for individual patients based on the model presented here. In addition, we have created an online application offering the same functionality in a user-friendly format: <https://emcbiostatistics.shinyapps.io/LongitudinalMixtureModelFMUE/>.

## Results

From the total of 479 patients in all 4 cohorts, we included data for 412 patients whose FM-UE had been measured at least 2 times, with the first and last measurements spaced at least 12 weeks apart. The 412 included patients were found to have a mean 6.1 measurements (standard deviation [SD] = 1.9) per patient, with an interval of 26.2 weeks (SD = 2.0) between the first and last measurements. The first FM-UE had been measured within the first 72 hours for 53% of patients, within the first week for 76% of patients, and within the first 2 weeks for 93% of patients.

The longitudinal mixture model of FM-UE recovery identified 5 different subgroups, with a residual error SD  $\sigma_e$  of 3.9 points (95% equal-tailed interval [ETI] = 3.7–4.0) on the FM-UE, corresponding to a variance explained of 0.97 (95% ETI = 0.97–0.98; Fig 1, Table 1). Patient characteristics (age, gender, and handedness) were comparable between subgroups. Baseline clinimetric scores correlated with the recovery coefficient as expected (Table 2); that is, more favorable clinimetric scores were associated with higher recovery coefficients. For example, subgroup 5, with the highest recovery coefficient, had the lowest score on the NIHSS and the highest scores on the motricity index and the finger extension item of the FM-UE, but the opposite was true for subgroup 1.

The number of patients included in the cross-validation increased with time poststroke as more patients with a baseline FM-UE became available (Fig 2). In addition, the median number of measurements per patient



**FIGURE 3:** Model Fugl-Meyer motor upper extremity (FM-UE) predictions for 3 typical patients. Model FM-UE predictions for example patients from the optimal (given all FM-UE data) poor (A–C), moderate (D–F), or good (G–I) FM-UE recovery cluster. The left column illustrates predictions made using data available at 2 weeks poststroke, the second column at 4 weeks poststroke, and the final column at 3 months poststroke. Open circles represent data used for prediction modeling. Filled markers indicate the actual endpoint FM-UE. The prediction is shown as the mean profile (dark line) with 68% equal-tailed intervals (dark shaded area) and 95% equal-tailed intervals (light shaded area). The figure titles and the colors of the credible intervals (poor [purple], moderate [orange], or good [green]) indicate the predicted FM-UE clusters as well as the probability of cluster assignment.

increased from 2 measurements at 1 week poststroke up to 5 measurements 8 weeks poststroke. Median future recovery, defined as endpoint FM-UE minus last available FM-UE for each patient, decreased with time poststroke from 10.0 (interquartile range [IQR] = 3.0–26.3) to 2.0 (IQR = 0.0–8.0) at 12 weeks poststroke. Reliability of endpoint FM-UE and  $\Delta$ FM-UE predictions increased with time poststroke and was higher for endpoint FM-UE than for  $\Delta$ FM-UE. The median absolute error for the predicted endpoint FM-UE was 4.8 (IQR = 1.3–12.8) at 1 week poststroke and 4.2 (IQR = 1.3–9.8) at 2 weeks poststroke, and the mean correlation between predicted and observed FM-UE was 0.84 (95% ETI = 0.83–0.84) at 1 week poststroke and 0.86 (95% ETI = 0.86–0.87) at 2 weeks poststroke. The median absolute error for the predicted  $\Delta$ FM-UE was 5.2 (IQR = 1.7–12.9) at 1 week poststroke and 4.8 (IQR = 1.7–11.0) at 2 weeks poststroke, and the mean correlation between predicted and observed  $\Delta$ FM-UE was 0.68 (95% ETI = 0.67–0.69) at 1 week poststroke and 0.71 (95% ETI = 0.71–0.72) at 2 weeks poststroke.

Based on the recovery coefficients ( $r_k$ ), time constants ( $\tau_k$ ), and initial distributions ( $\mu_{\alpha,k}$  and  $\sigma_{\alpha,k}$ ), we organized the 5 subgroups into 3 main FM-UE recovery clusters with poor (subgroup 1), moderate (subgroups 2 and 3), and good (subgroups 4 and 5) recovery profiles (see Tables 1 and 2). Mean accuracy of the FM-UE recovery cluster assignment was 0.79 (95% ETI = 0.78–0.8) at 1 week poststroke and 0.81 (95% ETI = 0.80–0.82) at 2 weeks (see Fig 2). Positive predictive value was high (>0.9) for the good FM-UE recovery cluster as early as 1 week poststroke and low to modest for the poor and moderate FM-UE recovery cluster at week 1 (0.66 [95% ETI = 0.63–0.68] and 0.50 [95% ETI = 0.42–0.57], respectively) and week 2 (0.72 [95% ETI = 0.70–0.73] and 0.61 [95% ETI = 0.57–0.64], respectively). The miss rate was lower than 0.1 for the poor and moderate FM-UE recovery cluster from week 1 onward, and the miss rate for the moderate cluster was much higher at 1 week (0.74 [95% ETI = 0.68–0.79]) and 2 weeks poststroke (0.63 [95% ETI = 0.60–0.65]).

FM-UE data for typical patients with model-based predictions of FM-UE recovery and FM-UE recovery clusters are shown in Figure 3. This figure illustrates how the credible intervals of the predictions decrease as more measurements become available and how individuals can initially be misclassified in terms of their FM-UE recovery cluster. Our prediction algorithm is available through a web-based application (Shiny App), which can be accessed at <https://emcbiostatistics.shinyapps.io/LongitudinalMixtureModelFMUE/>. This web-based application requires FM-UE scores and measurement dates, and outputs predict FM-UE profiles with credible intervals as well as the most likely FM-UE recovery cluster.

## Discussion

We have developed a longitudinal mixture model of FM-UE recovery that describes the time course of FM-UE recovery after a first-ever ischemic stroke and does not suffer from mathematical coupling.<sup>19,20</sup> Based on this model, we analyzed a large FM-UE data set of 412 first-ever ischemic stroke patients collected in prospective cohorts. Subsequently, we identified 5 subgroups, which we organized into 3 clinically relevant clusters of poor, moderate, and good FM-UE recovery. Based on a cross-validation, our research provides first-ever estimates of predictability of endpoint FM-UE between 3 and 6 months poststroke, as well as subgroup assignment as a function of time poststroke. These results contribute to the understanding of recovery patterns in the first 6 months after stroke.

Our current longitudinal mixture model of FM-UE recovery, as opposed to the proportional recovery model, cannot be confounded by mathematical coupling. Hope et al<sup>19</sup> showed that the correlations between baseline FM-UE score (distribution X) and the amount of recovery defined as endpoint FM-UE minus baseline FM-UE (distribution Y-X) found in proportional recovery research could be inflated by mathematical coupling. However, because mathematical coupling applies to correlations of data points (baseline and endpoint FM-UE) and not to models of longitudinal data, the recovery coefficients in our research represent nonconfounded measures of recovery as a proportion of potential recovery. In addition, mathematical coupling does not apply to the outcomes of the cross-validation, as we report correlations between the model predictions and the observed values for endpoint FM-UE and  $\Delta$ FM-UE rather than correlations of the form X and Y-X.

In contrast to studies relying on the proportional recovery rule, which have identified 2 subgroups of recoverers (fitters) and nonrecoverers (nonfitters),<sup>1,2,15–17,36</sup> our model distinguishes 5 subgroups, differing in the amount and rate of recovery as well as the distribution of the FM-UE early after stroke. Patients in subgroup 1, containing approximately 30% of patients, have a low baseline FM-UE and a small recovery coefficient, resulting in a poor outcome. These patients seem to overlap with the nonrecoverers from the proportional recovery rule. Subgroups 2 to 5 refine the recoverers in a more granular pattern. The majority of the recoverers (subgroups 4 and 5) regain close to 0.9 of their potential recovery in the first weeks after stroke, which is on the high end of previous estimates of 0.55 to 0.85,<sup>1,2,15–17,36</sup> whereas the recoverers in subgroups 2 (0.45) and 3 (0.86) also regain a fair amount of their potential recovery but over a much longer time frame. Because previous studies have identified disruption



of the corticospinal tract as the essential difference between recoverers and nonrecoverers,<sup>15–18</sup> we expect a similar contrast between patients from subgroup 1 and patients from subgroups 2 to 5. Indeed, the baseline Bamford classification shows a strikingly higher percentage of total anterior circulation infarctions in subgroup 1 compared with the other 4 subgroups. Further definition of the structural and possibly also the genetic characteristics of the 5 subgroups might lead to a better understanding of FM-UE recovery.

Our study provides first-ever cross-validated estimates of individual endpoint FM-UE and  $\Delta$ FM-UE prediction errors. Theoretically, it is possible to predict endpoint FM-UE at baseline using the proportional recovery rule as well. One approach could be to first identify recoverers and nonrecoverers using measurements of corticospinal intactness (transcranial magnetic stimulation [TMS]<sup>16</sup> and diffusion tractography [DTI]<sup>15</sup>) and then estimate endpoint FM-UE for the recoverers as the baseline FM-UE plus a proportional recovery term and for the nonrecoverers as just the baseline FM-UE. However, even though TMS<sup>37,38</sup> and DTI<sup>17,39</sup> have been validated as markers of recoverers and nonrecoverers, the absolute error of predicted FM-UE or  $\Delta$ FM-UE scores for a population of first-ever ischemic stroke patients based on this combined approach has never been cross-validated. We found the median absolute error of endpoint FM-UE to be 4.8 at the first week poststroke and 4.2 at the second week poststroke, which is at the low end of what is deemed to be a clinically important difference (4.25–7.25).<sup>40</sup> Therefore, our model can provide a satisfactory prognosis to patients as early as 1 week poststroke. In the future, further reduction in prediction errors may be achieved by adding time-dependent covariates such as the recently recommended performance assays<sup>41</sup> and biomarkers of corticospinal integrity (eg, TMS<sup>16</sup> or DTI<sup>15</sup>) to improve the accuracy of subgroup assignment early after stroke. Interested researchers can apply our model to predict FM-UE recovery and the FM-UE recovery cluster by accessing a web-based application at <https://emcbiostatistics.shinyapps.io/LongitudinalMixtureModelFMUE/>. This application requires 1 or multiple FM-UE measurements (dates and scores) from a single patient to predict upper limb recovery within the first 6 months. Predictions are presented as the expected recovery with 68% and 95% credibility intervals to express uncertainty.

Currently, we do not yet recommend that clinicians implement our model in clinical practice or provide FM-UE recovery predictions based on our model. First, future studies should externally validate the model with different stroke rehabilitation data sets. Outcomes of these studies might also be that the precision of the model needs to be

increased (using some of the recommendations listed previously) before clinical implementation is realistic. Second, guidelines need to be developed on responsible communication of stroke recovery prognoses to patients and health-care professionals, with special emphasis on the uncertainty of the model predictions. Finally, it is necessary to investigate whether knowledge of the FM-UE prognosis actually improves rehabilitation efficiency or outcome.

Based on the 5 subgroups identified by the model, we defined poor, moderate, and good FM-UE recovery clusters (similar to the lower, middle, and upper band groups identified in the classic descriptive cohort study of Garraway et al<sup>42</sup> almost 40 years ago). These clusters could, in the future, be relevant for personalizing therapeutic interventions as well as supporting decisions on the discharge policy after admission to acute and subacute stroke units. For example, patients in the poor FM-UE recovery cluster (subgroup 1) will show very limited motor recovery and might therefore benefit from learning compensation strategies<sup>7</sup> or early started neuropharmacological interventions<sup>43</sup> aimed at promoting neural repair.<sup>9</sup> In contrast, patients in the moderate FM-UE recovery cluster (subgroups 2 and 3) recover reasonably well over an extended period and might benefit from early started intensive therapeutic interventions aimed at behavioral restitution.<sup>7</sup> Patients in the good FM-UE recovery cluster (subgroups 4 and 5) are expected to require support in regaining advanced skills such as writing.<sup>7</sup>

In a research setting, the present model can be used to select patients for interventions designed for a specific FM-UE recovery cluster (eg, interventions designed specifically for the moderate recovery cluster). Patient selection can be achieved by predicting the cluster for a patient based on the patient's early FM-UE scores using the web-based application of our model early after stroke. The efficiency of this approach depends critically on the positive predictive value and the miss rate of cluster assignment. Positive predictive value in this context indicates the proportion of patients from a predicted cluster who have been assigned to their optimal cluster and will therefore receive the personalized intervention specifically designed for their cluster. In the current model, the positive predictive value is high for the good FM-UE recovery cluster, fair for the poor cluster, and relatively low for the moderate cluster. We therefore expect that an intervention designed for good recoverers will be regularly offered to patients with good FM-UE recovery, and an intervention designed for moderate recoverers will be regularly offered to patients with poor or good recovery. The miss rate is the proportion of incorrectly assigned patients from an optimal cluster who will therefore receive a personalized intervention designed for another cluster. We found that the miss rate is low for

the good and poor cluster, and we therefore expect that patients from these clusters will often get the intervention designed for their cluster; we found the miss rate is high for the moderate cluster and therefore expect that patients from this cluster will often receive an intervention designed for another cluster. Identification of patients in the poor and moderate FM-UE recovery cluster might benefit from additional repeated FM-UE measurements over time. Of particular interest would be to design a decision algorithm that identifies patients in whom the cluster prediction is uncertain and advises on specific measurements to achieve sufficient accuracy. An additional option to increase assignment accuracy would be to incorporate clinical markers as explained previously.

Another future application of the longitudinal mixture model of FM-UE recovery could be to detect intervention effects in recovery and rehabilitation trials with more statistical power.<sup>14</sup> To estimate an intervention effect, the model would need to be amended with an additional term to capture differences in the extent or possibly also the rate of recovery. This amended model could be fitted to serially collected clinical data to establish the added value of an innovative therapeutic intervention above usual care either for the entire study population or for the 3 FM-UE recovery clusters separately. Given that all serial measurements are analyzed, rather than just the baseline and endpoint FM-UE, as is true for stroke recovery and rehabilitation trials, this approach could significantly promote study power. Studies investigating therapies specifically designed for either poor, moderate, or good stroke recoverers could additionally use the model's predicted FM-UE recovery cluster early after stroke to select patients, as explained previously. This way, the proportion of patients from a certain FM-UE recovery cluster and the power to detect an intervention effect in that FM-UE recovery cluster will both increase, with the positive predictive value determining study homogeneity and the miss rate the study inclusion efficiency.<sup>44,45</sup> Further quantification of these approaches will be one of the main targets of our future work.

Limitations of the present study include the lack of severely affected patients with a hemiparesis in the dominant hand, the restricted generalization to patients with upper limb motor impairment after a first-ever ischemic stroke, and the focus on stroke recovery rather than deterioration. The language center is localized in the left hemisphere for most left-handed and right-handed people.<sup>46</sup> Therefore, severely affected patients with left hemisphere lesions often have language impairments that hinder providing informed consent and therefore participating in a clinical study. This explains the low percentage of patients with a hemiparesis on the dominant side in subgroup

1 (severely affected patients). In addition, we cannot conclude whether hemorrhagic stroke patients have similar recovery patterns or investigate how spontaneous neurological recovery is affected by recurrent stroke. Finally, our model is not equipped to predict FM-UE deterioration after stroke. As recently emphasized, the next step is to start an international collaboration for building data sets large enough to address these questions and move recovery and rehabilitation studies forward.<sup>44</sup> These databases could also be used to model recovery of lower limb impairment<sup>3,4</sup> as well as other nonmotor modalities such as speech<sup>6</sup> and visuo-spatial neglect<sup>5</sup> after stroke.<sup>21</sup>

---

## Acknowledgment

This work was supported by ZonMw (grants 10-10400-98-008 and 104003008) and Amsterdam Movement Sciences. Data collection in the 4 prospective cohorts was supported by a number of grants: the Netherlands Organization for Health Research and Development (ZonMw grant 89000001), the European Research Council (ERC) under the European Union Seventh Framework Program (FP/2007–2013)/ERC Advanced grant 291339, the Dutch Society of Physical Therapy (grant 33368), and the Dutch Brain Foundation (Hersenstichting Nederland, grant F2011(1)-25).

## Author Contributions

R.v.d.V., R.W.S., E.-R.S., G.M.R., M.A.F., and G.K. contributed to the concept and study design. R.v.d.V., R.W.S., E.-R.S., R.N., C.M., and G.K. contributed to the data acquisition and analysis. R.v.d.V., R.W.S., and G.K. drafted the manuscript and figures. All authors approved the final version.

## Potential Conflicts of Interest

Nothing to report.

---

## References

1. 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.
2. Winters C, van Wegen EEH, 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.
3. Smith M-C, Byblow WD, Barber PA, Stinear CM. Proportional recovery from lower limb motor impairment after stroke. *Stroke* 2017;48:1400–1403.
4. 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.

5. Nijboer TCW, Kollen BJ, Kwakkel G. Time course of visuospatial neglect early after stroke: a longitudinal cohort study. *Cortex* 2013; 49:2021–2027.
6. Lazar RM, Speizer AE, Festa JR, et al. Variability in language recovery after first-time stroke. *J Neurol Neurosurg Psychiatry* 2008;79: 530–534.
7. Kwakkel G, Kollen B, Lindeman E. Understanding the pattern of functional recovery after stroke: facts and theories. *Restor Neurol Neurosci* 2004;22:281–299.
8. Gresham GE. Stroke outcome research. *Stroke* 1986;17:358–360.
9. Ward NS. Restoring brain function after stroke—bridging the gap between animals and humans. *Nat Rev Neurol* 2017;13:244–255.
10. Boyd LA, Hayward KS, Ward NS, et al. Biomarkers of stroke recovery: consensus-based core recommendations from the Stroke Recovery and Rehabilitation Roundtable. *Int J Stroke* 2017;12:480–493.
11. Bernhardt J, Godecke E, Johnson L, Langhorne P. Early rehabilitation after stroke. *Curr Opin Neurol* 2017;30:48–54.
12. Byblow W, Schlaug G, Wittenberg G. What's the perfect dose for practice to make perfect? *Ann Neurol* 2016;80:339–341.
13. Gladstone DJ, Danells CJ, Black SE. The Fugl-Meyer assessment of motor recovery after stroke: a critical review of its measurement properties. *Neurorehabil Neural Repair* 2002;16:232–240.
14. Krakauer J, Marshall R. The proportional recovery rule for stroke revisited. *Ann Neurol* 2015;78:845–847.
15. 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.
16. Byblow WD, Stinear CM, Barber PA, et al. Proportional recovery after stroke depends on corticomotor integrity. *Ann Neurol* 2015;78: 848–859.
17. Buch ER, Rizk S, Nicolo P, et al. Predicting motor improvement after stroke with clinical assessment and diffusion tensor imaging. *Neurology* 2016;86:1924–1925.
18. Marshall RS, Zarahn E, Alon L, et al. Early imaging correlates of subsequent motor recovery after stroke. *Ann Neurol* 2009;65:596–602.
19. Hope TMH, Friston K, Price CJ, et al. Recovery after stroke: not so proportional after all? *Brain* 2019;142:15–22.
20. Hawe RL, Scott SH, Dukelow SP. Taking proportional out of stroke recovery. *Stroke* 2019;50:204–211.
21. Kundert R, Goldsmith J, Veerbeek JM, et al. What the proportional recovery rule is (and is not): methodological and statistical considerations. *Neurorehabil Neural Repair* 2019;33:876–887.
22. Kwakkel G, Winters C, van Wegen EEH, et al. Effects of unilateral upper limb training in two distinct prognostic groups early after stroke. *Neurorehabil Neural Repair* 2016;30:804–816.
23. Nijland RHMM, van Wegen EEHH, Harmeling-van der Wel BC, Kwakkel G. Presence of finger extension and shoulder abduction within 72 hours after stroke predicts functional recovery: early prediction of functional outcome after stroke: the EPOS cohort study. *Stroke* 2010;41:745–750.
24. Kwakkel G. 4D-EEG: uncovering the neurological mechanisms behind upper limb functional recovery after stroke. Available at: <https://www.trialregister.nl/trial/4084>. Accessed January 18, 2019.
25. Duncan PW, Zorowitz R, Bates B, et al. Management of adult stroke rehabilitation care: a clinical practice guideline. *Stroke* 2005;36: e100–143.
26. Quinn T, Paolucci S, Sunnerhagen K, et al. Evidence-based stroke rehabilitation: an expanded guidance document from the European Stroke Organisation (ESO) guidelines for management of ischaemic stroke and transient ischaemic attack 2008. *J Rehabil Med* 2009;41: 99–111.
27. Bamford J, Sandercock P, Dennis M, et al. Classification and natural history of clinically identifiable subtypes of cerebral infarction. *Lancet* 1991;337:1521–1526.
28. Meyer BC, Hemmen TM, Jackson CM, Lyden PD. Modified National Institutes of Health Stroke Scale for use in stroke clinical trials: prospective reliability and validity. *Stroke* 2002;33:1261–1266.
29. Bohannon R. Motricity index scores are valid indicators of paretic upper extremity strength following stroke. *J Phys Ther Sci* 1999;11: 59–61.
30. Collin C, Wade D. Assessing motor impairment after stroke: a pilot reliability study. *J Neurol Neurosurg Psychiatry* 1990;53:576–579.
31. Sanford J, Moreland J, Swanson LR, et al. Reliability of the Fugl-Meyer assessment for testing motor performance in patients following stroke. *Phys Ther* 1993;73:447–454.
32. Gelman A, Carlin JB, Stern HS, Rubin DB. *Bayesian data analysis*. 3rd ed. Boca Raton, FL: Chapman and Hall/CRC Press, 2013.
33. Malsiner-Walli G, Frühwirth-Schnatter S, Grün B. Model-based clustering based on sparse finite Gaussian mixtures. *Stat Comput* 2016; 26:303–324.
34. Rousseau J, Mengersen K. Asymptotic behaviour of the posterior distribution in overfitted mixture models. *J R Stat Soc Series B StatMethodol* 2011;73:689–710.
35. Nasserinejad K, van Rosmalen J, de Kort W, Lesaffre E. Comparison of criteria for choosing the number of classes in Bayesian finite mixture models. *PLoS One* 2017;12:e0168838.
36. 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.
37. Stinear CM, Byblow WD, Ackerley SJ, et al. Proportional motor recovery after stroke. *Stroke* 2017;48:795–798.
38. Schambra HM, Xu J, Branscheidt M, et al. Differential poststroke motor recovery in an arm versus hand muscle in the absence of motor evoked potentials. *Neurorehabil Neural Repair* 2019;33: 568–580.
39. Guggisberg AG, Nicolo P, Cohen LG, et al. Longitudinal structural and functional differences between proportional and poor motor recovery after stroke. *Neurorehabil Neural Repair* 2017;31: 1029–1041.
40. Page SJ, Fulk GD, Boyne P. Clinically important differences for the upper-extremity Fugl-Meyer Scale in people with minimal to moderate impairment due to chronic stroke. *Phys Ther* 2012;92:791–798.
41. Kwakkel G, Van Wegen E, Burridge JH, et al. Standardized measurement of quality of upper limb movement after stroke: consensus-based core recommendations from the Second Stroke Recovery and Rehabilitation Roundtable. *Int J Stroke* 2019;14:783–791.
42. Garraway WM, Akhtar AJ, Smith DL, Smith ME. The triage of stroke rehabilitation. *J Epidemiol Community Health* 1981;35:39–44.
43. Pekna M, Pekny M, Nilsson M. Modulation of neural plasticity as a basis for stroke rehabilitation. *Stroke* 2012;43:2819–2828.
44. Bernhardt J, Borschmann K, Boyd L, et al. Moving rehabilitation research forward: developing consensus statements for rehabilitation and recovery research. *Int J Stroke* 2016;11:454–458.
45. Winters C, Heymans MW, van Wegen EEH, Kwakkel G. How to design clinical rehabilitation trials for the upper paretic limb early post stroke? *Trials* 2016;17:468.
46. Knecht S, Dräger B, Deppe M, et al. Handedness and hemispheric language dominance in healthy humans. *Brain* 2000;123:2512–2518.