Supplementary material for: Wearable-Based Digital Metrics Reveal Dynamic Dissociations Between Upper Limb Capacity and Real-World Use Over the First Six Months After Stroke

Eva Josse[1,2], Hsiao-Ju Cheng[1], Lay Fong Chin[3], Clara XJ Soh[3], Christopher W. K Kuah[1,3,4], Megan SE Lau[3], Chwee Yin Ng[3], Tegan K. Plunkett[3], Stephanie HZ Yeo[3], Catherine E. Lang[5,6,7], Karen SG Chua[1,3,4,8], Roger Gassert[1,2], and Olivier Lambercy[1,2]

[1]Singapore-ETH Centre, Future Health Technologies Programme, CREATE Campus, Singapore.

[2]Rehabilitation Engineering Laboratory, Department of Health Sciences and Technology, ETH Zurich, Switzerland.

[3] Institute of Rehabilitation Excellence (IREx), Tan Tock Seng Hospital Rehabilitation Centre, Singapore.

[4] Rehabilitation Research Institute of Singapore, Nanyang Technological University, Singapore.

[5] Program in Physical Therapy, Washington University School of Medicine, USA.

[6] Program in Occupational Therapy, Washington University School of Medicine, USA.

[7] Department of Neurology, Washington University School of Medicine, USA.

[8] Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore.

Corresponding author: Eva Josse, Future Health Technologies, 1 Create Way, CREATE Tower, #06-01, 138602 Singapore, Singapore.

eva.josse@sec.ethz.ch, relab.publications@hest.ethz.ch.

Appendix A Motor Evoked Potential: Assessment of Potential Bias

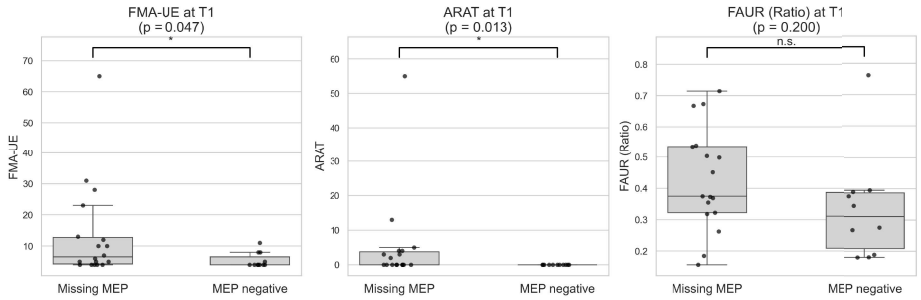

Fig. A1 Comparison of baseline motor impairment, activity capacity, and performance metrics between participants with missing motor evoked potential (MEP) data and those with negative MEP at T1. MEP status was determined either by direct measurement (in participants with SAFE < 5) or assumed positive for participants with SAFE ≥ 5. MEP data were missing in 18 participants due to refusal or medical contraindications. While participants with missing MEP data showed slightly higher baseline scores on some measures (e.g., ARAT), group distributions largely overlapped and differences were not consistent across metrics, suggesting limited evidence for systematic bias in MEP availability.

Appendix B Model of the Coupling Between Capacity and Performance

Table B1 OLS model of performance (FAUR) rate, with participant-clustered standard errors. N = 145 intervals, 57 participants; $R^2 = 0.17$, Adj. $R^2 = 0.15$.

Predictor	β **(Estimate)**	**95% CI**	**p-value**	**Interpretation**
Intercept	0.039	[0.012, 0.066]	0.005	Average rate of change in performance when both predictors = 0, i.e., capacity rate = 0 and time = average post-stroke time across all observations (centered time).
Capacity rate (β_1)	–0.0005	[–0.006, 0.007]	0.877	At the average post-stroke time across all observations (2.8 months, time centered at 0), changes in capacity are not associated with changes in performance.
$\text{Time}_{\text{centered}}$ (β_2)	–0.016	[–0.035, 0.003]	0.098	As time increases, performance gains become smaller (holding capacity rate constant).
Capacity rate $\times$ $\text{Time}_{\text{centered}}$ (β_3)	–0.003	[–0.007, 0.001]	0.098	The coupling between changes in capacity and performance weakens over time.

Appendix C Mixed-Effects Model Comparison for Capacity and Performance

Tables C2–C3 present comparisons of random- and fixed-effects structures for each outcome (ARAT, FAUR, FAUD, and FAUI). Model fit was assessed using Akaike's Information Criterion (AIC), Bayesian Information Criterion (BIC), and the log-likelihood (LogLik). When models showed comparable fit ($\Delta\text{AIC} < 4$), the less complex model was retained; if similar, preference was given to consistent time parameterization across fixed and random effects.

Table C2 Comparison of random-effects structures across outcomes. Fixed-effects structure (centered time) was held constant across models, and all models were estimated using restricted maximum likelihood (REML). Random intercepts and slopes on log-transformed time at the participant level were systematically preferred based on model fit.

Random effects	Converged	AIC	ΔAIC	BIC	LogLik
(1 + log1p(time)$_{centered}$ | ID)	**Yes**	**1651**	**0**	**1664**	**-821**
(1 | ID)	Yes	1710	59	1717	-853
(1 + time$_{centered}$ | ID)	Yes	1753	102	1766	-872

ARAT (N=54)

Random effects	Converged	AIC	ΔAIC	BIC	LogLik
(1 + log1p(time)$_{centered}$ | ID)	**Yes**	**-12**	**0**	**1**	**10**
(1 + time$_{centered}$ | ID)	Yes	-3	9	9	5
(1 | ID)	Yes	-2	10	4	3

FAUR (N=41)

Random effects	Converged	AIC	ΔAIC	BIC	LogLik
(1 + log1p(time)$_{centered}$ | ID)	**Yes**	**1149**	**0**	**1162**	**-571**
(1 + time$_{centered}$ | ID)	Yes	1158	9	1171	-575
(1 | ID)	Yes	1163	14	1169	-579

FAUD (N=43)

Random effects	Converged	AIC	ΔAIC	BIC	LogLik
(1 + log1p(time)$_{centered}$ | ID)	**Yes**	**690**	**0**	**702**	**-341**
(1 + time$_{centered}$ | ID)	Yes	699	9	711	-346
(1 | ID)	Yes	710	20	716	-353

FAUI (N=43)

Table C3 Comparison of fixed-effects structures across outcomes. Random-effects structure (random intercept and slope on log-transformed time) was held constant across models, and all models were estimated using maximum likelihood (ML). For ARAT, quadratic models provided the best fit. Given that the difference in AIC between the two quadratic models was small (ΔAIC < 4), the log-quadratic model was retained to ensure consistency with the selected random-effects structure. For FAUR, FAUD, and FAUI, among the best-fitting models (log-linear, log-quadratic, and quadratic), the log-linear model was retained for parsimony.

Model name	Fixed effects	Converged	AIC	ΔAIC	BIC	LogLik
Quadratic	$\text{time}_{centered}$ + $\text{time}_{centered}{}^2$	Yes	1625	0	1648	-805
Log-quadratic	**$\text{log1p(time)}_{centered}$ + $\text{log1p(time)}_{centered}{}^2$**	**Yes**	**1625**	**0**	**1649**	**-806**
Linear	$\text{time}_{centered}$	Yes	1659	34	1679	-823
Log-linear	$\text{log1p(time)}_{centered}$	No	1647	22	1667	-818

ARAT (N=54)

Model name	Fixed effects	Converged	AIC	ΔAIC	BIC	LogLik
Quadratic	$\text{time}_{centered}$ + $\text{time}_{centered}{}^2$	Yes	-29	0	-8	22
Log-quadratic	$\text{log1p(time)}_{centered}$ + $\text{log1p(time)}_{centered}{}^2$	Yes	-27	2	-6	21
Log-linear	**$\text{log1p(time)}_{centered}$**	**Yes**	**-25**	**4**	**-7**	**19**
Linear	$\text{time}_{centered}$	Yes	-20	9	-2	16

FAUR (N=41)

Model name	Fixed effects	Converged	AIC	ΔAIC	BIC	LogLik
Log-quadratic	$\text{log1p(time)}_{centered}$ + $\text{log1p(time)}_{centered}{}^2$	Yes	1140	0	1161	-563
Quadratic	$\text{time}_{centered}$ + $\text{time}_{centered}{}^2$	Yes	1141	1	1162	-563
Log-linear	**$\text{log1p(time)}_{centered}$**	**Yes**	**1143**	**3**	**1162**	**-566**
Linear	$\text{time}_{centered}$	Yes	1156	16	1174	-572

FAUD (N=43)

Model name	Fixed effects	Converged	AIC	ΔAIC	BIC	LogLik
Quadratic	$\text{time}_{centered}$ + $\text{time}_{centered}{}^2$	Yes	680	0	701	-333
Log-quadratic	$\text{log1p(time)}_{centered}$ + $\text{log1p(time)}_{centered}{}^2$	Yes	683	3	705	-335
Log-linear	**$\text{log1p(time)}_{centered}$**	**Yes**	**683**	**3**	**702**	**-336**
Linear	$\text{time}_{centered}$	Yes	690	10	708	-339

FAUI (N=43)

Appendix D Diagnostics and Performance of Best-Fitting Models for Capacity and Performance

D.1 ARAT

Model specification: Mixed-effects model (log-quadratic); No. of observations = 213; Groups = 54 (mean = 3.9 [3, 4] obs/group). Method: restricted maximum likelihood (REML).

Table D4 Summary of best-fitting mixed-effects model for ARAT.

Effect	Coef.	Std.Err.	z	p	[0.025, 0.975]
Intercept	28.7	3.1	9.1	<0.001	[22.5, 34.9]
log1p(time)$_{centered}$	7.5	1.3	6.0	<0.001	[5.1, 10.0]
log1p(time)$_{centered}$ [2]	-2.8	0.8	-3.4	0.001	[-4.5, -1.2]

Random effects: Var(Intercept) = 517.0; Var[log1p(time)$_{centered}$] = 74.9; Cov(Intercept, log1p(time)$_{centered}$) = -1.5; Residual Var = 21.1.

Convergence: Yes.

Nakagawa R^2 [60]: $R^2_{marginal}$ = 5%, $R^2_{conditional}$ = 97%

Leave-one-participant-out cross-validation[1]: RMSE$_{marginal}$[2] = 24.3, RMSE$_{conditional}$[3] = 13.7

[1] All outcomes except the baseline time point (T0) were predicted.
[2] Only the fixed effects were used to predict future outcomes.
[3] Participant-specific random effects were estimated via BLUPs derived from each individual's T0 measurement.

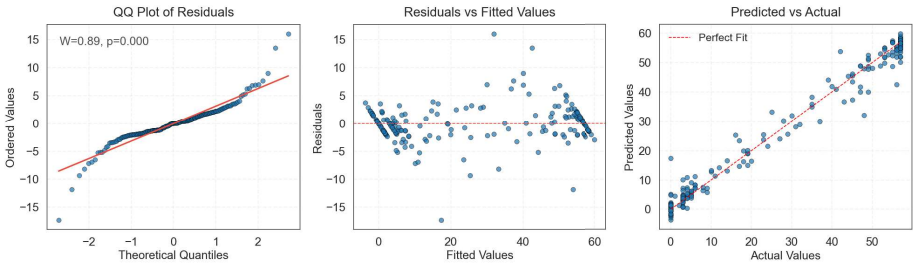

Fig. D2 Model diagnostics for the best-fitting ARAT model. The left plot shows the Q–Q plot assessing residual normality. Despite minor departures from normality due to floor and ceiling effects in the ARAT, linear mixed-effects models are generally robust to moderate violations of distributional assumptions, and fixed-effect estimates remain approximately unbiased [61]. Therefore, this model was considered appropriate. The middle plot displays residuals versus fitted values to check homoscedasticity, with no evident pattern. The right plot illustrates model fit quality through predicted versus actual values.

D.2 FAUR (Ratio)

Model specification: Mixed-effects model (log-linear); No. of observations: 153; Groups: 41 (mean: 3.7 [3, 4] obs/group); Method: restricted maximum likelihood (REML).

Table D5 Summary of best-fitting mixed-effects model for FAUR.

Effect	**Coef.**	**Std.Err.**	**z**	**p**	**[0.025, 0.975]**
Intercept	0.63	0.04	15.0	<0.001	[0.55, 0.72]
$\text{log1p(time)}_{centered}$	0.07	0.03	3.00	0.003	[0.03, 0.12]

Random effects: Var(Intercept) = 0.07; Var[$\text{log1p(time)}_{centered}$] = 0.02; Cov(Intercept, $\text{log1p(time)}_{centered}$) = 0.00; Residual Var = 0.02.

Convergence: Yes.

Nakagawa R^2 [60]: $R^2_{marginal}$ = 3%, $R^2_{conditional}$ = 83%

Leave-one-participant-out cross-validation[1]: $\text{RMSE}_{marginal}$[2] = 0.30 $\text{RMSE}_{conditional}$[3] = 0.23

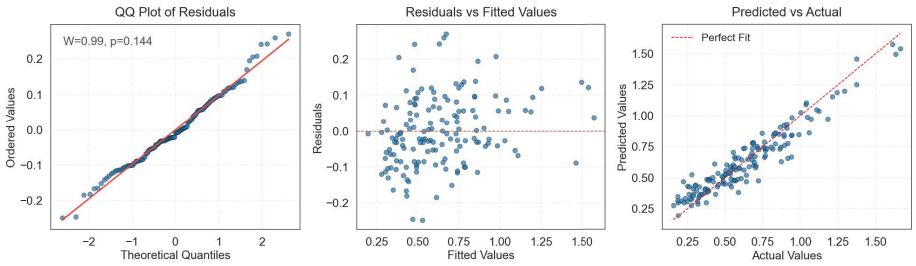

Fig. D3 Model diagnostics for the best-fitting FAUR model. The left plot shows the Q–Q plot assessing residual normality; the Shapiro–Wilk test rejected the non-normality hypothesis. The middle plot displays residuals versus fitted values to check homoscedasticity, with no evident pattern. The right plot illustrates model fit quality through predicted versus actual values.

D.3 FAUD (Duration)

Model specification: Mixed-effects model (log-linear); No. of observations: 160; Groups: 43 (mean: 3.7 [3, 4] obs/group); Method: restricted maximum likelihood (REML).

Table D6 Summary of best-fitting mixed-effects model for FAUD.

Effect	Coef.	Std.Err.	z	p	[0.025, 0.975]
Intercept	20.3	1.6	12.4	<0.001	[17.1, 23.6]
log1p(time)$_{centered}$	4.7	1.0	4.9	<0.001	[2.8, 6.5]

Random effects: Var(Intercept) = 109.7; Var[log1p(time)$_{centered}$] = 26.3; Cov(Intercept, log1p(time)$_{centered}$) = 18.7; Residual Var = 24.4.

Convergence: Yes.

Nakagawa R^2 [60]: $R^2_{marginal}$ = 7%, $R^2_{conditional}$ = 85%

Leave-one-participant-out cross-validation[1]: RMSE$_{marginal}$[2] = 12.2 RMSE$_{conditional}$[3] = 9.7

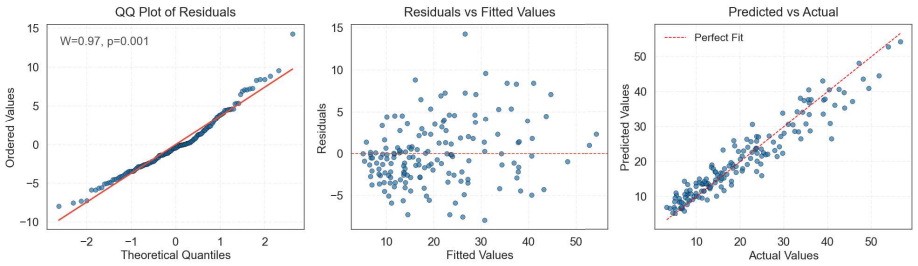

Fig. D4 Model diagnostics for the best-fitting FAUD model. The left plot shows the Q–Q plot assessing residual normality. Despite minor departures from normality in FAUD, linear mixed-effects models are generally robust to moderate violations of distributional assumptions, and fixed-effect estimates remain approximately unbiased [61]. The middle plot displays residuals versus fitted values to check homoscedasticity, with no evident pattern. The right plot illustrates model fit quality through predicted versus actual values.

D.4 FAUI (Intensity)

Model specification: Mixed-effects model (log-linear); No. of observations: 160; Groups: 43 (mean: 3.7 [3, 4] obs/group); Method: restricted maximum likelihood (REML).

Table D7 Summary of best-fitting mixed-effects model for FAUI.

Effect	Coef.	Std.Err.	z	p	[0.025, 0.975]
Intercept	8.3	0.3	24.8	<0.001	[7.7, 9.0]
log1p(time)$_{centered}$	1.0	0.2	4.0	<0.001	[0.5, 1.4]

Random effects: Var(Intercept) = 4.4; Var[log1p(time)$_{centered}$] = 1.6; Cov(Intercept, log1p(time)$_{centered}$) = 1.7; Residual Var = 1.7.

Convergence: Yes.

Nakagawa R^2 [60]: $R^2_{marginal}$ = 6%, $R^2_{conditional}$ = 77%

textbfLeave-one-participant-out cross-validation[1]: RMSE$_{marginal}$[2] = 2.8 RMSE$_{conditional}$[3] = 2.6

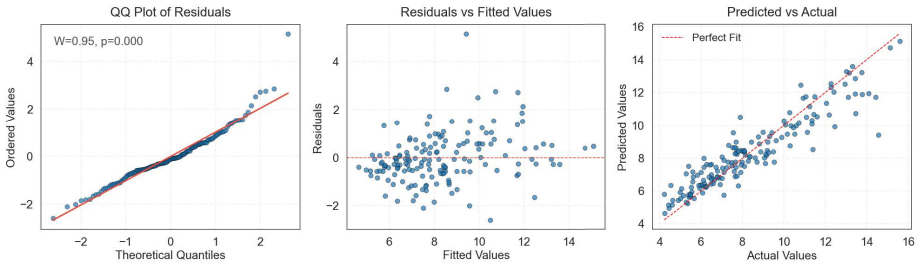

Fig. D5 Model diagnostics for the best-fitting FAUI model. The left plot shows the Q–Q plot assessing residual normality. Despite minor departures from normality in FAUI, linear mixed-effects models are generally robust to moderate violations of distributional assumptions, and fixed-effect estimates remain approximately unbiased [61]. The middle plot displays residuals versus fitted values to check homoscedasticity, with no evident pattern. The right plot illustrates model fit quality through predicted versus actual values.

Appendix E Simulated Distribution of Individual Changes Based on Best-Fitting Models

E.1 Simulation of Model-Consistent Recovery Trajectory after Stroke

Interpretation: The goal of these simulations was to determine whether the model-predicted change was significantly greater than zero, given the model-based uncertainty, for each participant.

Framework

1. Model specifications: The fitted mixed-effects model was defined as $y_{ij} = X_{ij}\beta + Z_{ij}u_i + \varepsilon_{ij}$, where y_{ij} is the outcome for participant i at time j; X_{ij} and Z_{ij} are the design matrices for the fixed and random effects, respectively; β represents the population-level coefficients; u_i the participant-specific random effects;

and ε_{ij} is the residual error.

2. Simulations: For each simulation s = 1, ..., 1000, fixed and random effects coefficients were randomly drawn from multivariate Gaussian distributions accounting for both population-level and individual-level uncertainty, and defined as:

$$\beta^{(s)} \sim \mathcal{N}(\hat{\beta}, \widehat{Var}(\hat{\beta}))$$

$$u_i^{(s)} \sim \mathcal{N}(E[\hat{u_i} \mid y_{i,pre}], Var(\hat{u_i} \mid y_{i,pre}))$$

where $\hat{\beta}$ are the estimated fixed effects, and $\widehat{Var}(\hat{\beta})$ is their variance-covariance matrix. $y_{i,pre}$ is the a priori observations (pre-6 months only). The posterior distribution of the random effects coefficients was estimated with BLUP described in section E.2.

Importantly, simulated outcomes were truncated within the valid range of the scale [a, b] (e.g., [0, 57] for ARAT, [0, 2] for FAUR, [0, 100] for FAUD, and [0, N.A.] for FAUI) to ensure clinically plausible values.

$$y_{ij}^{(s)} \sim \mathcal{N}_{trunc}(\mu_{ij}^{(s)}, \hat{\sigma_\varepsilon}^2; a, b), \text{with } \mu_{ij}^{(s)} = X_{ij}\beta^{(s)} + Z_{ij}u_i^{(s)}$$

where $\hat{\sigma_\varepsilon}^2$ is the residual variance of the original best-fitting model. This was equivalent to residuals $\varepsilon_{ij}^{(s)} = y_{ij}^{(s)} - \mu_{ij}^{(s)}$ being drawn from a truncated normal distribution centered at zero, conditional on the predicted mean.

3. Probability of change: Based on the posterior predictive simulations of individual change between baseline and 6 months post-stroke, the probability that the model-predicted change was significantly greater than 0, was defined as:

$$p_{predicted>0} = \frac{\sum_{s=1}^{1000} \mathbb{1}(\Delta_{simulated}^{(s)} > 0)}{1000}$$

4. Decision rule: Participants with $p_{predicted>0} \geq 0.84$ have 84% or more probability that the predicted change is significantly different from zero and positive. $p_{predicted>0} \leq 0.16$ have 84% or more probability that the predicted change is significantly different from zero and negative. $0.16 < p_{predicted>0} < 0.84$ are not significantly different from 0.

In comparison to Lang et al. (2023) [13], the predicted change score cannot be converted to a z-score. Indeed, the distribution of simulated changes is not Gaussian due to truncation, so the ratio $\frac{change_{predicted}}{SE(change_{simulated})}$ does not follow a standard Normal distribution.

E.2 Best Linear Unbiased Predictor (BLUP)

To estimate the Gaussian conditional distribution of the random effects in a mixed-effects model based only on *a priori* observations (pre-6 months), the **Best Linear Unbiased Predictor (BLUP)** was applied.

The model was defined as $y = X\beta + Zu + \varepsilon$, where X and Z are the design matrices for the fixed and random effects, respectively; β and u are the corresponding coefficient vectors; and ε is the vector of residual errors. By assumption, $u \sim \mathcal{N}(0, G), \quad \varepsilon \sim \mathcal{N}(0, R), \quad \text{and} \quad \text{Cov}(u, \varepsilon) = 0$, where G and R are the covariance matrices of the random effects and residuals, respectively.

BLUP computes the conditional expectation $E[u \mid y_{\text{pre}}]$ and variance $\text{Var}(u \mid y_{\text{pre}})$, where y_{pre} denotes the observations before 6 months.

Let $y_{\text{pre}} = y$ for simplicity.

From the definition of covariance,

$$\text{Cov}(u, y) = E\big[(u - E[u])(y - E[y])^T\big] = E\big[u(Zu + \varepsilon)^T\big] = GZ^T.$$

The joint distribution of y and u is also Gaussian and can be written as

$$\begin{pmatrix} y \\ u \end{pmatrix} = \begin{pmatrix} X\beta \\ 0 \end{pmatrix} + \begin{pmatrix} Z & I \\ I & 0 \end{pmatrix} \begin{pmatrix} u \\ \varepsilon \end{pmatrix}, \qquad \begin{pmatrix} y \\ u \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} X\beta \\ 0 \end{pmatrix}, \begin{pmatrix} ZGZ^T + R & ZG \\ GZ^T & G \end{pmatrix} \right),$$

where I is the identity matrix.

From this joint distribution, the marginal variance of y is $\text{Var}(y) = ZGZ^T + R$, and the conditional expectation and variance of u given y are derived using standard multivariate Gaussian identities [62]:

$$E[u \mid y] = \text{Cov}(u, y)\,\text{Var}(y)^{-1}(y - E[y]) + E[u] = GZ^T(ZGZ^T + R)^{-1}(y - X\beta),$$

$$\text{Var}(u \mid y) = \text{Var}(u) - \text{Cov}(u, y)\,\text{Var}(y)^{-1}\text{Cov}(y, u) = G - GZ^T(ZGZ^T + R)^{-1}ZG.$$

Therefore, under the Gaussian assumptions of a mixed-effects model, this defines the conditional (posterior) distribution of the random effects given only *a priori* observations.

E.3 Plots of the Simulated Distributions

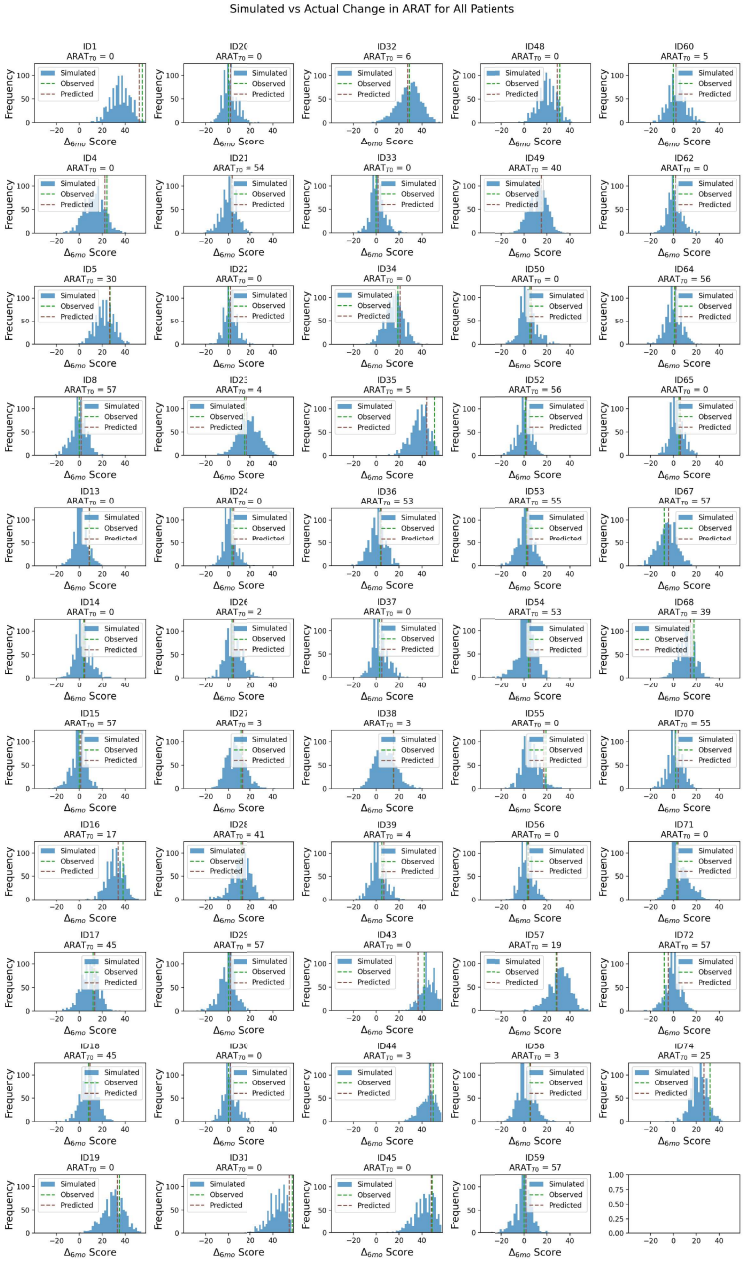

Fig. E6 Simulated ARAT changes per participant.

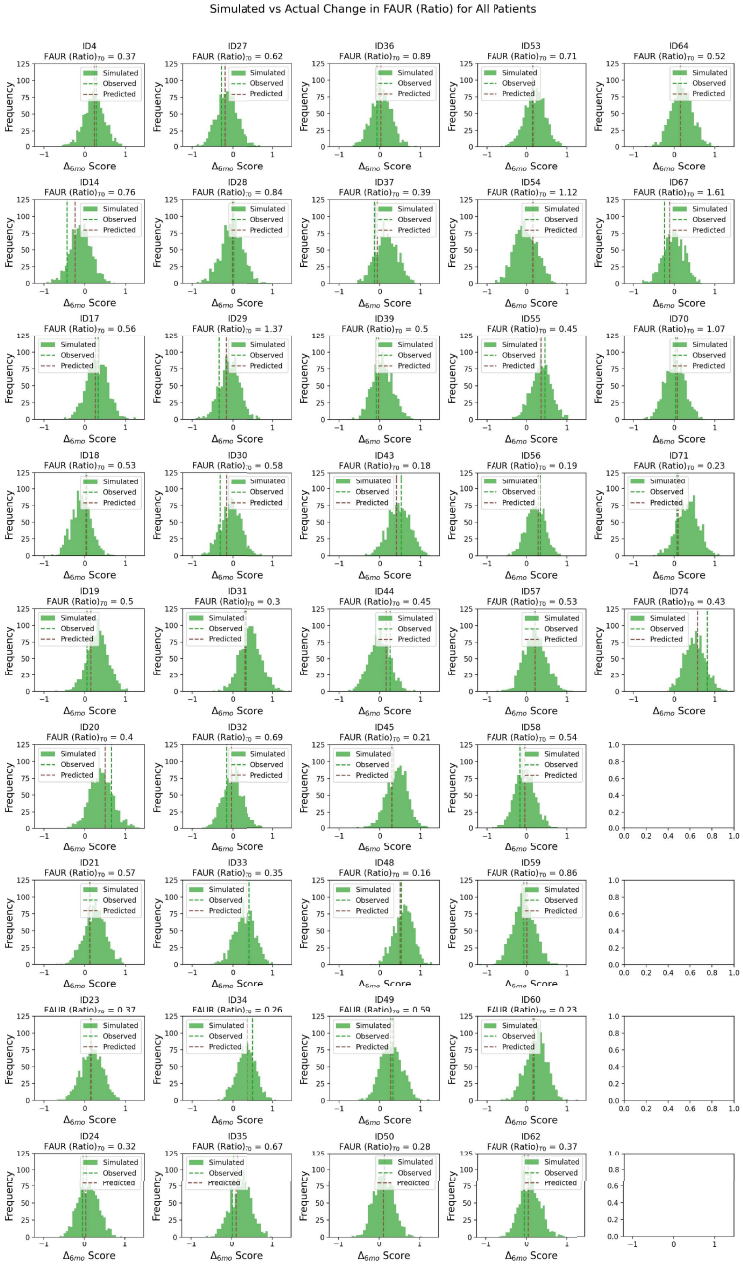

Fig. E7 Simulated FAUR changes per participant.

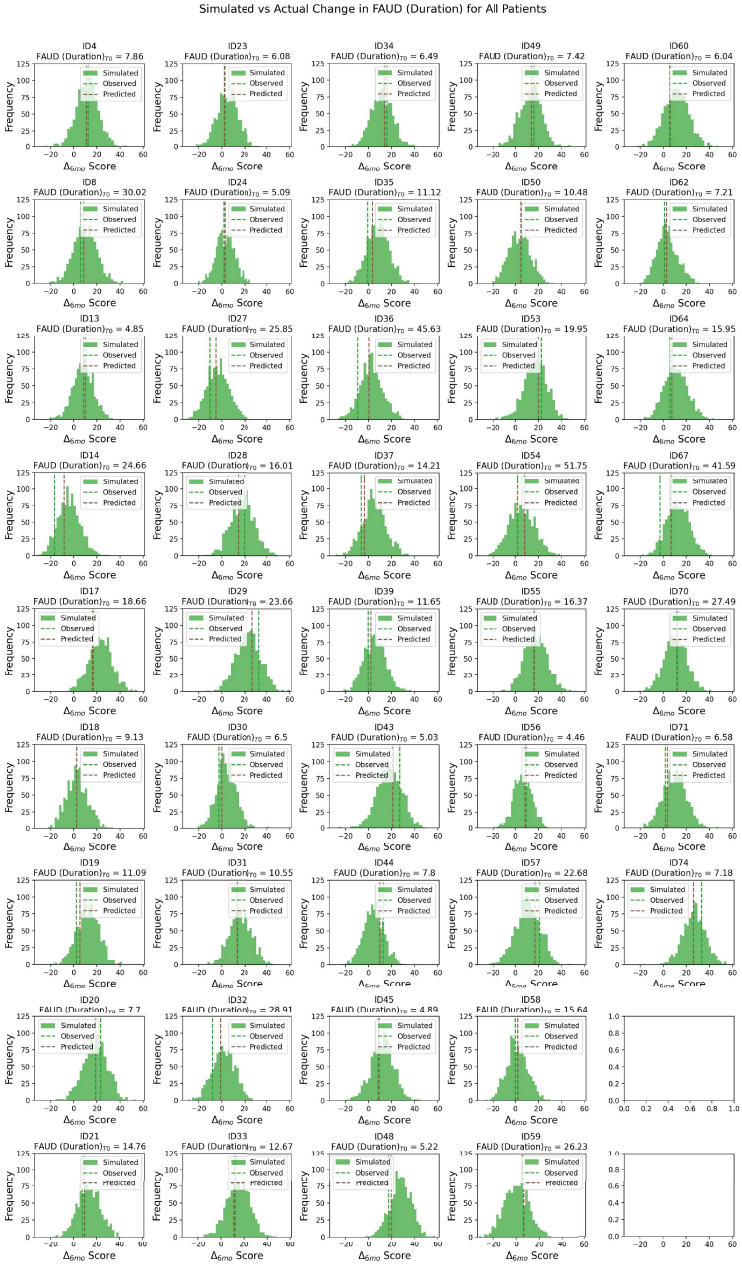

Fig. E8 Simulated FAUD changes per participant.

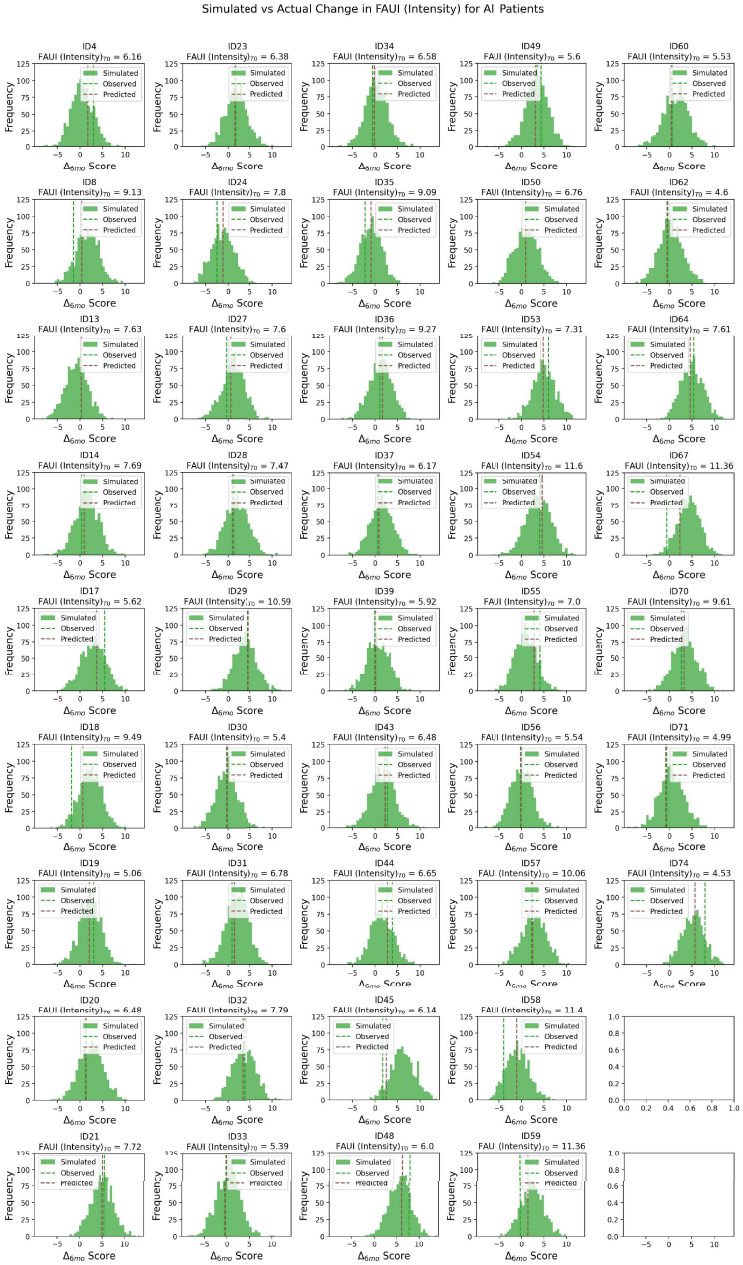

Fig. E9 Simulated FAUI changes per participant.

Appendix F Dissociation Between Changes in Capacity and Performance

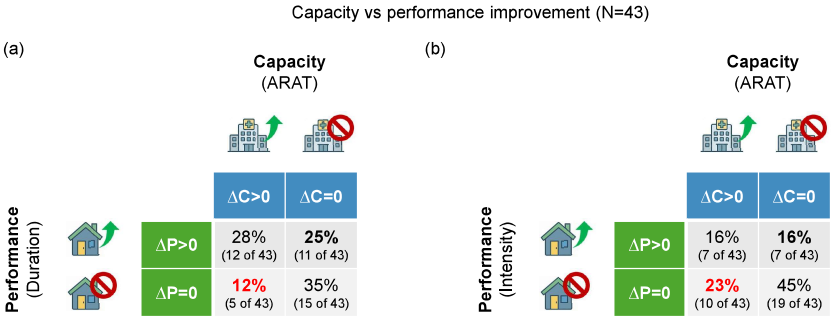

Fig. F10 Improvements in capacity versus improvements in functional arm use duration (a) and intensity (b) over 6 months.

Table F8 Descriptive distributions of participants stratified by capacity (**C: ARAT**) and performance (**P: FAUR**) improvement subgroups. Unless otherwise stated (Δ scores), all variables were measured at baseline. Values are mean ± SD [range] unless otherwise stated. These results are provided for contextual interpretation and should be interpreted with caution given the small sample size.

	$\Delta\mathbf{C}_{>0}\mathbf{P}_{>0}$	$\mathbf{C}_{=0}\mathbf{P}_{>0}$	$\mathbf{C}_{>0}\mathbf{P}_{=0}$	$\mathbf{C}_{=0}\mathbf{P}_{=0}$
N	9	6	8	18
Age [y]	51±12 [35, 75]	67±11 [49, 78]	49±11 [34, 64]	61±13 [38, 81]
Gender (F/M)	1F/8M	3F/3M	4F/4M	5F/13M
Stroke type (Isch./Hem.)	1I/8H	3I/3H	4I/4H	12I/6H
Paretic arm is dominant	11%	50%	50%	22%
MEP (Positive/ Negative/Not Tested)	6P/0N/3n.a.	2P/3N/1n.a.	6P/0N/2n.a.	10P/3N/5n.a.
MOCA [0–30]	22±4 [15, 28]	22±4 [14, 26]	23±2 [19, 26]	23±5 [15, 29]
SAFE [0–10]	3±3 [0, 8]	3±3 [0, 9]	4±4 [0, 10]	5±4 [0, 10]
ARAT [0–57]	12±19 [0, 45]	1±2 [0, 5]	15±18 [0, 45]	28±28 [0, 57]
ΔARAT [0–57]	32±14 [12, 57]	4±7 [0, 19]	27±15 [8, 51]	2±4 [-8, 11]
FMA-UE [0–66]	21±24 [4, 61]	11±9 [4, 26]	34±21 [8, 64]	34±29 [4, 65]
Grip force paretic [kgF]	5±10 [0, 31]	1±2 [0, 5]	6±9 [0, 23]	9±10 [0, 26]
MAS finger [0–4]	1.1±0.3 [1.0, 2.0]	1.7±1.2 [1.0, 4.0]	1.6±0.8 [1.0, 3.0]	1.4±0.8 [1.0, 4.0]
FAUR (Ratio) [-]	0.35±0.17 [0.16, 0.59]	0.31±0.11 [0.19, 0.45]	0.56±0.17 [0.37, 0.84]	0.73±0.37 [0.28, 1.61]
ΔFAUR (Ratio) [-]	0.41±0.20 [0.05, 0.81]	0.36±0.19 [0.07, 0.65]	0.10±0.14 [-0.16, 0.28]	-0.08±0.18 [-0.43, 0.17]
FAUD (Duration) - paretic [%]	8.5±4.43 [4.89, 18.66]	8.97±4.57 [4.46, 16.37]	13.7±8.21 [6.08, 28.91]	21.57±13.45 [5.09, 51.75]
ΔFAUD (Duration) - paretic [%]	16.89±8.47 [2.38, 33.14]	11.44±6.94 [1.84, 23.24]	7.76±9.62 [-8.19, 21.02]	2.72±11.31 [-16.64, 33.0]
FAUI (Intensity) - paretic [counts/s]	6±1 [5, 7]	6±1 [5, 7]	8±1 [6, 10]	8±2 [5, 12]
ΔFAUI (Intensity) - paretic [counts/s]	4±3 [-1, 8]	1±2 [-1, 4]	1±2 [-2, 4]	1±3 [-4, 6]

Appendix G Baseline Predictors of Capacity and Performance Improvements

Table G9 Primary models identifying baseline predictors of significant improvements in capacity and performance outcomes. The least correlated motor-function proxy (FMA-UE) was selected among the motor assessments (SAFE, grip force, FMA-UE). Odds ratios are reported from the final refitted model. Predictive performance (mean macro-F1 ± SD) reflects the nested cross-validation procedure. Predictor stability is quantified as the proportion of outer folds in which the predictor was selected and its mean rank based on the standardized coefficient magnitude. Aff. = Dom. indicate that the paretic arm is dominant.

Model variant		Predictors & Effects		Stability features		Predictive performance
Outcome	**Motor function proxy**	**Retained predictors**	**OR (95% CI)**	**% folds selected**	**Mean rank**	**Mean macro-F1 ± SD**
ARAT	FMA-UE	Age, MEP positive	0.95 [0.90–0.98], 2.63 [1.64–3.87]	100, 100	1.8, 1.2	0.72 ± 0.09
FAUR	FMA-UE	Inpatient duration	1.03 [0.99–1.10]	80	1.2	0.53 ± 0.31
FAUD	FMA-UE	Aff.=Dom., MEP positive	0.38 [0.20–0.70], 1.73 [0.88–3.26]	100, 100	1.2, 1.8	0.71 ± 0.18
FAUI	FMA-UE	FMA-UE	1.04 [1.02–1.07]	100	1	0.72 ± 0.19

Table G10 Secondary models for FAUR and FAUI, to probe secondary biomechanical and neural constraints, after exclusion of the primary predictors of improvements.

Model variation	Retained predictors	OR (95% CI)
No motor impairment-related predictor	MAS finger MEP positive	0.42 [0.26-0.77], 2.58 [1.59-3.79]

FAUR (N=41)

Model variation	Retained predictors	OR (95% CI)
No inpatient duration	Ischemic MoCA	0.42 [0.23-0.82], 0.91 [0.80-1.04]

FAUI (N=43)