Chapter 4

Accuracy of physical therapists' early predictions of upper limb capacity at hospital stroke units: the EPOS study

Rinske HM Nijland, Erwin EH van Wegen,
Barbara C Harmeling-van der Wel, Gert Kwakkel;
on behalf of the EPOS Investigators. A complete list of the EPOS
Investigators is provided in the acknowledgements.

Accepted for publication in *Phys Ther*

ABSTRACT

Background: Early prediction of outcome after stroke is becoming increasingly important, since most patients are discharged from hospital stroke units within several days after stroke.

Objectives: 1) To determine the accuracy of physical therapists (PTs) prediction at hospital stroke unit regarding upper limb (UL) capacity; 2) to develop a Computational Prediction Model (CPM) and 3) to compare the accuracy of PTs' and CPMs' prediction. Secondary objectives were to investigate the impact of timing on the accuracy of PTs and CPMs' predictions and to investigate the direction of the difference between predicted and observed outcome. Finally, we investigated if the accuracy of PTs' prediction was affected by their experience in stroke rehabilitation.

Design: Prospective cohort study.

Methods: PTs made predictions at T_{72h} and $T_{discharge}$ about UL capacity after 6 months in 3 categories, derived from the Action Research Arm Test. At the same time, clinical variables were measured to derive a CPM. The accuracy of PTs and CPMs' predictions was evaluated by calculating Spearman rank correlation coefficients (r_s) between predicted and observed outcome.

Results: 131 patients and 20 PTs participated. For T_{72h}, r_s between predicted and observed outcome was 0.63 (P<0.01) for PTs' predictions, and 0.75 (P<0.01) for the CPM. For $T_{discharge}$, r_s for PTs' prediction improved to 0.75 (P<0.01), the r_s for CPMs' predictions improved slightly to 0.76.

Limitations: PTs administered a test battery every 3 days, which might enhance accuracy of prediction.

Conclusions: Accuracy of PTs' predictions at T_{72h} is lower than that of the CPM. At $T_{discharge}$, PTs and CPMs' predictions are about equally accurate.

4

INTRODUCTION

Stroke is the leading cause of disabilities in developed countries.[1] One of the main problems after stroke is impaired upper limb (UL) function, which strongly affects the ability to perform activities of daily living (ADL),[2] as well as quality of life.[3, 4]

The time course of functional recovery after stroke is difficult to predict for individual patients because of substantial inter-individual variability.[5] Nevertheless, clinical decision-making with respect to discharge policies and intervention planning is often based on the expected functional improvements within a certain time window post stroke. Considering that most patients are typically discharged from hospital stroke units to rehabilitation centers or nursing homes within the first week after stroke onset,[6] early prediction about the degree of functional improvement is becoming increasingly important.

The main focus of stroke rehabilitation, and in particular the work of physiotherapists and occupational therapists, is on mobility, ADL independency and UL training. For UL training, we distinguish between patients with severe, moderate and mild impairments based on prognostic determinants for outcome of dexterity.[7] Especially patients with mild to moderate impairments present an opportunity to optimize the function of the affected arm.[7] This group of patients will probably benefit most from intensive rehabilitation interventions, such as Constraint Induced Movement Therapy.[8-10] Conversely, rehabilitation for patients with severe UL impairments should focus more on learning compensation- and coping strategies to deal with the impairments. To distinguish between these groups already in the first week after stroke, is therefore essential for clinical decision making and might reduce the economic burden of stroke by optimally allocating resources. To distinguish patients with favorable or unfavorable prognosis for recovery, several compuational prediction models (CPMs) for UL outcome have been developed.[5, 11] Recently, we showed for instance that it is possible to accurately predict UL capacity at 6 months by administering two simple clinical tests, finger extension and shoulder abduction, within 72 hours post stroke.[7] The positive likelihood ratio of this CPM, when applied within 72 hours was 5.123 (95% CI, 2.945–9.800), whereas the negative likelihood ratio was 0.132 (95% CI, 0.084–0.219) Over the first nine days post stroke the positive likelihood ratio increased to 5.436 (95% CI, 3.252–9.094) and the negative likelihood ratio decreased to 0.066 (95% CI, 0.030–0.135) This improved accuracy may be related to a complex array of processes involving several underlying mechanisms, such as restitution of penumbral tissues, elevation of diaschisis as well as homeostatic, and Hebbian and anti-Hebbian forms of neuroplasticity.[12] These processes contribute to the non-linear spontaneous neurological recovery and operate in different, partly overlapping time-frames.[12] As a consequence, the accuracy of a CPM is heavily dependent on the moment in time when the clinical determinants are assessed.[7, 13, 14]

The aim of CPMs is to provide objective estimates of outcome probabilities, in order to complement clinical intuition, expertise and guidelines. However, most of these models are seldom implemented in clinical practice, due to a lack of usability. Application of a CPM requires for instance unambiguous definitions of the predictors and availability of reproducible methods in clinical practice,[15] which is not always the case with for instance sophisticated techniques such as neuro-imaging and neurophysiologic assessments.[15, 16] Further, many CPMs won't be implemented easily due to doubts about generalizability. Models may for example be restricted to a hospital population, a certain age-group or severity of stroke, whereas patients with a previous history of stroke are often excluded.[17] These restrictions increase the precision of the CPM, however limit its generalizability.[18]

Current knowledge about processes underlying the forecasting of clinical outcome and clinical decision making is still limited. Two approaches have been described in the literature to improve our understanding of decision -making processes, the intuitive approach and the analytical approach.[16] The intuitive approach leans heavily on the experience and insights of the therapist, who recognizes overall patterns and acts accordingly. The analytical approach is more scientifically oriented and involves aspects like scrutiny and hypothetico-deductive reasoning, which is characterized by collecting relevant data and then testing hypotheses.[16, 19] A combination of both approaches is probably used in routine clinical practice.[16, 19] In either approach, it is expected that the accuracy of prediction will improve by observing progress over time,[20] and therefore one may hypothesize that therapists can, like CPMs, make more accurate predictions regarding recovery after stroke as time progresses.

So far, however, the accuracy of predictions made by physical therapists (PTs) during the first days at a hospital stroke unit is unknown. Results by Kwakkel and colleagues[21] suggest that predictions about functional outcomes made by physical and occupational therapists in the second and fifth weeks after stroke onset are within the 95% confidence interval (CI) of CPMs and hence considered to be equally accurate. As expected, their results showed that therapists make more accurate predictions of final outcomes at week five after stroke than at week two after stroke.21 It was also shown that therapists' predictions of final outcome tended to be too pessimistic at both points in time.

In view of the early discharge from the hospital, the first aim of the present study was to determine the accuracy of predictions made by PTs in a hospital stroke unit about UL capacity at 6 months post stroke. The second aim was to develop a CPM with data of the same group of patients. The third aim was to compare the accuracy of predictions made by PTs with predictions made by the CPM. A secondary aim was to investigate the impact of timing post stroke on the accuracy of prediction made by PTs and by the CPM, by comparing predictions made within 72 hours after stroke onset with predictions made at discharge from the hospital stroke unit. Another secondary aim was to investigate the direction of the difference between predicted and observed outcome. With that, we investigated whether PTs and the CPM tended to be too pessimistic or too optimistic. Finally, we investigated if the accuracy of PTs' prediction was affected by their experience in stroke rehabilitation and their engagement in continuing education.

METHODS

Design

The EPOS (Early Prediction of functional Outcome after Stroke) study is a prospective cohort study with a repeated measurements design. Over a period of 36 months, 131 patients with stroke were recruited from 9 acute hospital stroke units in The Netherlands. Stroke was defined according to the World Health Organization criteria.[22] Patients recruited for the EPOS study were assessed within 72 hours (T_{72h}) after stroke onset and re-assessed every 3 days until the moment of discharge from the hospital stroke unit ($T_{discharge}$). Finally, all patients were assessed again at 6 months after stroke onset. Assessments were performed by trained assessors working as PTs in the hospital stroke units of the participating centers. The PTs were asked at T_{72h} and $T_{discharge}$ to make predictions about UL capacity at 6 months after stroke.

Subjects

Patients meeting the following admission criteria were included: (1) a first-ever ischemic anterior circulation stroke; (2) monoparesis or hemiparesis within the first 72 hours after stroke onset; (3) no disabling medical history (i.e., a premorbid Barthel Index (BI) score ≥19); (4) at least 18 years of age; (5) no severe deficits in communication, memory or understanding that impede proper measurement performance; (6) signed informed consent. Patients in the current study did not participate in other studies and all patients received usual treatment according to the Dutch rehabilitation guidelines.[23] The study was approved by the ethics committees of all participating hospitals, and all affiliated rehabilitation centers and nursing homes approved local feasibility.

Outcome variable

The outcome in terms of UL capacity at 6 months was assessed with the Action Research Arm Test (ARAT).[24] This one-dimensional hierarchical test consists of 19 functional tasks that are divided into four domains, i.e., grasp, grip, pinch and gross movement, with a maximum total score of 57 points. The clinimetric properties of the ARAT are excellent.[24, 25] All ARAT measurements were performed in a standardized manner according to Yozbatiran et al.,[26] together with the time limits used in the study by Van der Lee et al.[25]

PTs' prediction

The PTs were asked to make a prediction about UL capacity at 6 months in 3 categories: (1) the patient will not regain any dexterity (<10 points on ARAT); (2) the patient will regain some dexterity (10–56 points on ARAT) and (3) the patient will regain full recovery and will score the maximum of 57 points on the ARAT. The cut-off score of 10 points was chosen because a score of 9 points or lower may reflect only gross arm movements, whereas a score above 9 points always represents some hand function. Patients with more than 10 points on the ARAT are particularly eligible for rehabilitative therapy. However, from patients' perspective, there is a big difference between 10 points on the ARAT and complete recovery. For this reason, patients who regained full upper limb recovery were distinguished from patients who regained some dexterity by adding the third category (57 points). Although this last category appears to be small, this category represents a large group of patients with normal UL capacity after stroke, which should be classified separately.

At the time the EPOS study started at the nine acute Hospital Stroke Units in 2006, and during the 36 months of patient recruitment, PTs were still unaware of the prognostic determinants (i.e. finger extension and shoulder abduction) that were included in the CPM of UL function, which was published in 2010.[7]

Computational Prediction Model

In constructing the CPM, ARAT scores were divided into the same 3 categories as those used by the therapists to predict outcome (i.e., <10 points, 10–56 points or 57 points).

The following independent variables were selected for the development of the prediction model: (1) gender; (2) age; (3) hemisphere of stroke; (4) type of stroke (Bamford classification)[27]; (5) days between stroke and first assessment; (6) recombinant tissue plasminogen activator (rTPA); (7) Comorbidity (Cumulative Illness Rating Scale [CIRS])[28]; (8) visual inattention (National

Institutes of Health Stroke Scale [NIHSS][29] item 11); (9) deviation conjugee (NIHSS item 2); (10) hemianopia (NIHSS item 3); (11) sensory loss (NIHSS item 8); (12) urinary incontinence (BI item 2 [30]); (13) severity and extent of paresis of the arm and leg according to the upper and lower extremity parts of the Motricity Index (MI)[31] and the Fugl-Meyer (FM) assessment[32]; and (14) Sitting Balance (Trunk Control Test [TCT])[33] The variables are presented in Table 4.1. The selection was based on our previous study[7] and existing literature.[34-37]

Table 4.1 Candidate determinants, measured within 72 hours, associated with the return of some dexterity at 6 months after stroke, as determined by ordinal logistic regression

	Parameter estimates			
	α-ARAT Cat 1: <10 points	α-ARAT Cat 2: 10–56 points	β-predictor (underlined outcome)	Significance level
Gender (M/F)	-0.905	0.688	0.210	0.516
Age (0<70; 1≥70)	-0.803	0.803	0.421	0.195
Hemisphere of Stroke (L/R)	-0.896	0.700	0.260	0.427
BAMFORD (0:TA/PA; 1:LA)	-1.498	0.158	-0.839	0.012*
Days till assessment (0≤3d; 1≥4d)	-0.795	0.795	0.222	0.813
rTPA (y/n)	-1.172	0.442	-0.572	0.116
CIRS total score (0≥1; 1=0)	-1.492	0.113	-0.562	0.219
Visual inattention (0≥1; 1=0)	-1.645	0.089	-1.236	0.000*
Deviation conjugee (0≥1; 1=0)	-1.800	0.129	-2.055	0.000*
Hemianopia (0≥1; 1=0)	-1.760	0.128	-1.896	0.000*
Sensory loss (0≥1; 1=0)	-1.531	0.114	-797	0.021*
Urinary incontinence (0<1; 1≥1)	-1.891	0.027	-1.897	0.000*
MI shoulder abduction (0=0; 1≥9)	-2.274	0.102	-3.451	0.000*
MI leg (0<25; 1≥25)[47]	-2.000	0.193	-3.336	0.000*
FM Shoulder elevation (0<1; 1≥1)	-2.100	0.114	-2.995	0.000*
FM Finger extension (0<1; 1≥1)	-3.825	-0.487	-4.156	0.000*
TCT: Sitting balance 0<25; 1=25)	-2.025	0.120	-2.797	0.000*

α, threshold value; β, regression coefficient.
The values between the brackets represent the cut-off scores.
BAMFORD:TA, Total Anterior Cerebral Infarction; PA, Partial Anterior Cerebral infarction; LA, Lacunar Anterior Cerebral Infarction; rTPA, recombinant tissue plasminogen activator; CIRS: Cumulative Illness Rating Scale; MI: Motricity Index; FM, Fugl-Meyer; TCT: Trunk Control Test.

Data analysis

PTs' predictions of final outcome

Accuracy of prediction

Spearman rank correlation coefficients (r_s) were used to evaluate the association between the PTs' predictions of final outcome on the one hand, and observed outcome after 6 months on the other. To explore if the accuracy of prediction improves as time progresses, the r_s for prediction at T_{72h} and at $T_{discharge}$ were compared.

Direction of the difference between predicted and observed outcome

To test the hypothesis that therapists are too pessimistic in their prediction, like they were in the study of Kwakkel et al.,[21] the Wilcoxon signed-rank test was used to evaluate the direction of the difference between predicted and observed outcome. A P-value of <0.05 on the Wilxocon signed-rank test was used to indicate if therapists were more often too pessimistic or too optimistic in their prediction.

Influence of PTs' experience in stroke rehabilitation

4

The influence of the PTs' skills was explored using the number of years of experience and the number of courses attended in the field of stroke rehabilitation as covariates in ordinal logistic regression analyses. The characteristics that are tested are presented in Table 4.2.

Computational Prediction Model

Development of the CPM

The association between the independent variables measured at T_{72h} and the outcome at 6 months in terms of UL capacity (assessed with the ARAT, i.e., <10 points, 10–56 points or 57 points) was at first investigated using bivariate ordinal logistic regression analysis, using the following equation: Prob (event *j*) = $1 / (1+e^{-(\alpha j-\beta x)})$. An example calculation of probabilities is presented in Box 4.1. Ordinal or ratio-scaled determinants were dichotomized on clinical grounds or based on the optimal cut-off point in a receiver-operating characteristic curve (ROC).[7] Based on bivariate ordinal logistic regression analysis, significant determinants assessed at T_{72h} (with liberal P-value of ≤0.10) were selected for the subsequent development of a multivariate ordinal regression model to predict the outcome in terms of UL capacity at 6 months. Subsequently, the same determinants, together with the dichotomized change scores on these determinants from T_{72h} to $T_{discharge}$, were used to develop the model for $T_{discharge}$. The

Box 4.1 Example calculation using ordinal logistic regression

Ordinal logistic regression

Ordinal logistic models produce cumulative probabilities, using the following equation:

$Prob\ (event\ j) = 1 / (1 + exp^{-(\alpha j - \beta x)})$.

Example: probabilities calculated with the parameter estimates for 10–56 points on the ARAT represent cumulative probabilities for <10 points and 10–56 points. As a consequence, probabilities to score 10–56 points on the ARAT need to be calculated as follows:

probabilities (score <10 points + 10–56 points) - probabilities (score <10 points).

Example based on the parameter estimates for inattention. For this example we assume that inattention is present (i.e. score 0):

Prob (ARAT <10) = 1 / 1 + exp $^{(1.645–1.236)}$) = 0.399
Prob (ARAT <10 points or 10–56 point) = 1 / 1 + exp $^{(-0.089–1.236)}$) = 0.790
Prob ARAT 10–56 points = 0.790–0.399 = 0.391
Prob ARAT <10 points or 10–56 points or 57 points = 1

When inattention is not present (i.e. score 1):
Prob (ARAT <10) = 1 / 1+ exp $^{(1.645)}$) = 0.162

α, threshold value for each category; β, regression coefficient.

4

multivariate ordinal logistic models thus derived were used to calculate probabilities of scoring <10 points, 10–56 points or 57 points on the ARAT at 6 months.

Accuracy of prediction

Spearman rank correlation coefficients (r_s) were used to evaluate the association between the predictions of final outcome made by the CPM (at T_{72h} and $T_{discharge}$) and the observed outcome at 6 months. To explore if the accuracy of prediction improves from T_{72h} to $T_{discharge}$, the r_s for predictions at T_{72h} and $T_{discharge}$ were compared.

Direction of the difference between predicted and observed outcome

The Wilcoxon signed-rank test was used to determine whether there was a significant difference between too optimistic and too pessimistic predictions of final outcome.

Table 4.2 Patient characteristics assessed within 72 hours after stroke

Patient characteristics	Total
N	131
Gender, M/F	66/65
Mean (SD) age, years	66.4 (13.9)
Hemisphere of stroke, L/R	55/76
rTPA, N/Y	95/36
Mean (SD) time interval (days) between stroke and:	
assessment/prediction within 72 h	2.03 (1.20)
assessment/prediction at discharge	10.10 (4.36)
assessment at 6 months	188.02 (10.81)
Length of stay in hospital stroke unit (days)*	10 (7–12)
Type of stroke (BAMFORD)	
LACI	57
PACI	46
TACI	28
NIHSS*	8 (4–13)
Cognitive disturbance	
Inattention, N/Y	70/61
Disorientation, N/Y	93/38
Impairments of vision	
Hemianopia, N/Y	89/42
Deviation conjugee, N/Y	89/42
Sensory loss, N/Y	47/84
TCT (0–100)*	74 (37–100)
MI arm (0–100)*	47 (9–76)
MI leg (0–100)*	53 (32–75)
FM arm (0–66)*	23 (6–52)
FM leg (0–34)*	23 (11–29)
FAC (0–5)*	1 (0–3)
ARAT total score (0–57)*	6 (0–38)
BI total score* (0–20)	8 (3–13)
Characteristics of physical therapists	
Number of participating therapists	20
Years of experience in stroke rehabilitation*	24 (23–27)
Type of courses in the field of stroke rehabilitation	
NDT, N/Y	7/13
Dutch course for clinimetrics post stroke, N/Y	8/12
Other courses in this field, N/Y	13/7

N, number of patients; M/F, male/female; SD, standard deviation; L/R, left/right; rTPA, recombinant tissue plasminogen activator; N/Y, no/yes; Type of stroke: LACI, Lacunar Anterior Cerebral Infarction; PACI, Partial Anterior Cerebral Infarction; TACI, Total Anterior Cerebral Infarction; NIHSS, National Institutes of Health Stroke Scale; TCT, Trunk Control Test; MI, Motricity Index; FM, Fugl-Meyer; FAC, Functional Ambulation Categories Score; ARAT, Action Research Arm Test; BI, Barthel Index; NDT: neuro-development treatment. * Median values (interquartile ranges).

Comparison between the accuracy of PTs' predictions of final outcome and the CPM

The r_s that were calculated to evaluate the accuracy of PTs' predictions were compared with the r_s that was calculated to evaluate the accuracy of the CPM. This was done for predictions made at T_{72h} as well as at $T_{discharge}$.

All analyses were two-tailed with a significance level of 0.05 for rejecting the H_0 hypothesis, using SPSS version 17.

RESULTS

One hundred thirty one patients and 20 therapists were selected for this part of the EPOS study, in which the accuracy of PTs' early prediction of upper limb function was investigated. (Figure 4.1). For these 131, of the 220 included patients, the assessments at T_{72h} and $T_{discharge}$ were performed by the same therapist (N=20 out of 27).

Table 4.2 presents the main patient characteristics, as well as the characteristics of the 20 participating PTs. Assessments at T_{72h} were performed at a mean (SD) of 2.03 (1.20) days after stroke, and those at $T_{discharge}$ at 10.10 (4.36) days. The median ARAT score was 6 points at T_{72h}. At 6 months, 35 (26.7%) patients had not regained any dexterity, 49 patients (37.4%) had regained some dexterity and 47 patients (35.9%) regained full recovery.

PTs' predictions of final outcome

Tables 4.3a and 4.3b present cross-tabulations of the PTs' predictions of final outcome and the observed outcome after 6 months.

Accuracy of prediction

R_s for predicted outcome and achieved outcome after 6 months was 0.63 (P<0.01) at T_{72h} and 0.75 (P<0.01) at $T_{discharge}$.

Direction of the difference between predicted and observed outcome

At T_{72h}, 60% of the PTs made an accurate prediction of the outcome, while 20% were too optimistic and 20% too pessimistic. At $T_{discharge}$, 72% of the PTs made an accurate prediction of the outcome, while 18% were too optimistic and 10% too pessimistic. The Wilcoxon signed-rank test indicated that there was no significant difference between predictions of final outcome that were too optimistic or too pessimistic made at T_{72h} (P=0.89) or $T_{discharge}$ (P=0.14).

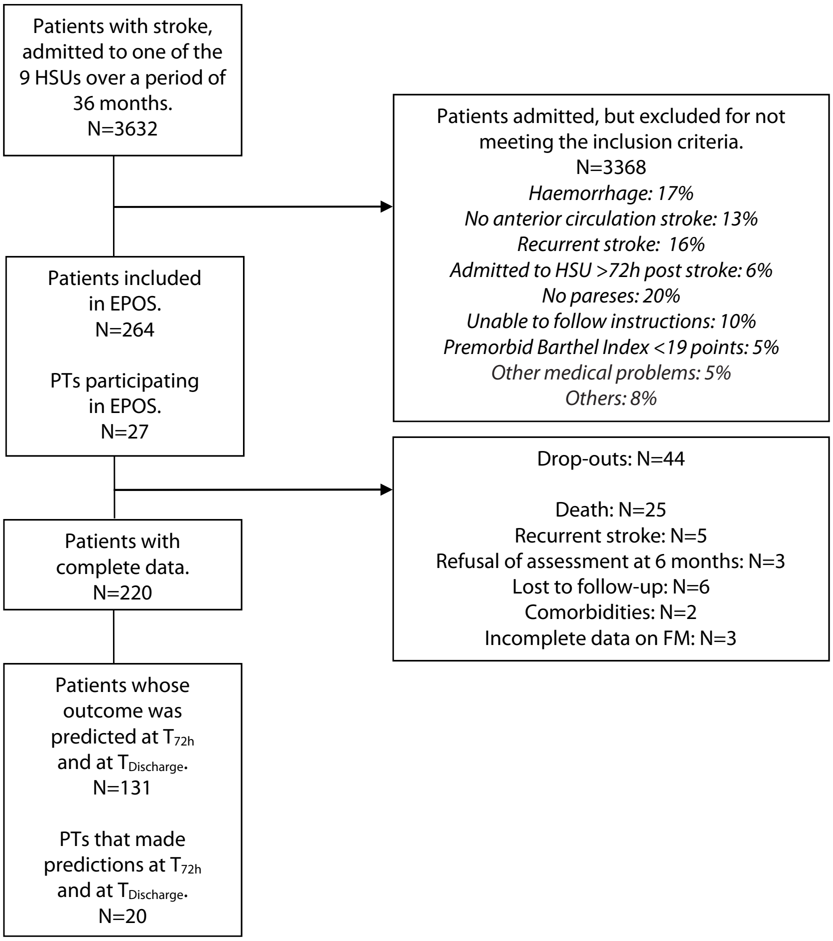

Figure 4.1 Flow-chart for inclusion.

Influence of PTs' experience in stroke rehabilitation

The PTs' ability to predict the outcome was not influenced by their level of skill, expressed in years of experience in stroke rehabilitation, or the number of relevant courses attended.

Computational Prediction Model

Development of the CPM

As shown in Table 4.1 bivariate ordinal logistic regression analysis showed that 11 candidate determinants were significantly related to UL capacity at 6 months post stroke ($P<0.10$): Type of stroke [Bamford classification]; visual inattention [NIHSS, item 11]; deviation con-

Table 4.3a Cross-tabulation for *PTs' predictions* at T_{72h} and final outcome

Prediction <72 h	Final outcome		
	<10	10–56	57
<10	**24**	4	0
10–56	11	**30**	23
57	0	15	**24**

The **bold** numbers represent the correct predictions.

Table 4.3b Cross-tabulation for *PTs' predictions* at $T_{discharge}$ and final outcome

Prediction discharge	Final outcome		
	<10	10–56	57
<10	**26**	1	0
10–56	9	**34**	13
57	0	14	**34**

The **bold** numbers represent the correct predictions.

Table 4.3c Cross-tabulation for *predictions by the CPM* at T_{72h} and final outcome

Prediction <72 h	Final outcome		
	<10	10–56	57
<10	**26**	7	0
10–56	8	**15**	3
57	1	27	**44**

The **bold** numbers represent the correct predictions.

Table 4.3d Cross-tabulation for *prediction by the CPM* at $T_{discharge}$ and final outcome

Prediction discharge	Final outcome		
	<10	10–56	57
<10	**30**	5	1
10–56	4	**17**	2
57	1	27	**44**

The **bold** numbers represent the correct predictions.

jugee [NIHSS, item 2]; hemianopia [NIHSS, item 3]; sensory loss [NIHSS item 8]; urinary incontinence [BI]; shoulder abduction [item from MI]; strength of the paretic leg [MI-Leg]; shoulder elevation [item from FM]; finger extension [item from FM]); sitting balance [TCT]. For deriving the multivariate model, first, we included all significant candidate predictors in an ordinal logistic regression model. Multivariate modelling resulted in finger extension and shoulder abduction as the remaining significant determinants in the final model. Finger extension was measured with an item of the FM-hand score (i.e. extension of the fingers), using an optimal cut-off of 1 point. Shoulder abduction was measured with an item of the MI-arm score (i.e. shoulder abduction) with an optimal cut-off of 9 points. The model at $T_{discharge}$ also included the dichotomized change scores on these determinants from T_{72h} until $T_{discharge}$, using the same cut-off points. Table 4.4 shows the probabilities based on the ordinal logistic regression models. The probability to score 10–56 points on the ARAT at 6 months is for instance estimated at 0.23 for those patients who are not able to show some voluntary finger extension and some shoulder abduction at T_{72h}. When those patients improve from T_{72h} to $T_{discharge}$, so that they show some voluntary finger extension and some shoulder abduction at $T_{discharge}$, their probability to score 10–56 points at 6 months increases from 0.23 to 0.74 at $T_{discharge}$.

4

Accuracy of prediction with CPM

Tables 4.3c and 4.3d present cross-tabulations of the CPMs' predictions of final outcome and the observed outcome after 6 months. R_s between predicted outcome and observed outcome after 6 months were r_s=0.75 (P<0.01) at T_{72h} and r_s=0.76 (P<0.01) at $T_{discharge}$.

Direction of the difference between predicted and observed outcome with CPM

At T_{72h} , 65% of the predictions determined with the CPM were correct, 27% were too optimistic and 8% were too pessimistic. At $T_{discharge}$, 70% of the predictions were correct, 24% were too optimistic and 6% were too pessimistic. The Wilcoxon signed rank test indicated that CPMs' predictions were significantly more often too optimistic than too pessimistic at both points in time (P<0.01).

Comparison between the accuracy of PTs' predictions of final outcome and the CPM

At T_{72h}, r_s for the prediction of final outcome made by PTs (r_s=0.63) was lower than that for the predictions made by the CPM (r_s=0.75). At $T_{discharge}$ the r_s for the PTs' predictions (r_s=0.75) and the predictions made by the CPM (r_s=0.76) were approximately equal.

Table 4.4 CPMs probabilities of scoring 0–9 / 10–56 / 57 on the Action Research Arm test at 6 months post stroke

Model	Finger Extension	Shoulder Abduction			0–9	10–56	57
<72 hours	+	+			0.02	0.36	0.62
	+	-			0.10	0.69	0.20
	-	+			0.34	0.61	0.06
	-	-			0.77	0.23	0.01
	Parameter estimates: Cat 1:-4.021; Cat 2:-0.506; FE(0):-3.342: SA(0):-1.863						
	Finger Extension <72 hours	Shoulder Abduction <72 hours	Improvement in Finger Extension <72 h – Discharge	Improvement in Shoulder Abduction <72 h – Discharge	0–9	10–56	57
Moment of Discharge	+	+	n.a	n.a	0.02	0.36	0.62
	+	-	n.a	+	0.01	0.24	0.75
	+	-	n.a	-	0.10	0.74	0.16
	-	+	+	n.a	0.18	0.73	0.09
	-	+	-	n.a	0.51	0.47	0.02
	-	-	+	+	0.11	0.74	0.15
	-	-	+	-	0.66	0.33	0.01
	-	-	-	+	0.36	0.60	0.04
	-	-	-	-	0.90	0.10	0.00
	Parameter estimates: Cat 1:-8.616; Cat 2:-4.788; FE(0):-4.355: SA (0):-2.172; Improvement FE(0):-1.533; Improvement SA(0):-2.755						

FE, Finger Extension; SA, Shoulder Abduction; n.a., not applicable.

DISCUSSION

The present study suggests that the accuracy of predictions of final outcome in terms of UL capacity made by PTs within 72 hours after stroke is lower than that of the CPM. The CPM was based on the presence of finger extension and shoulder abduction as prognostic determinants.[7] The accuracy of prediction at the moment of discharge was, however, similar for PTs and the CPM.

The present results are in line with a previous study by Kwakkel et al.,[21] which also suggested that predictions made by therapists at two and five weeks after stroke are equally accurate as those by CPMs. However, in that specific study, therapists' predictions of outcome tended to be too pessimistic, whereas the results of the present study indicated that therapists' predictions tended to be too optimistic. This may be due to differences in the patient population studied. Patients included in our cohort can be classified as having mild to moderate stroke, whereas patients in the study by Kwakkel et al.[21] were on average severely or very severely affected at stroke onset. The discrepancy in findings may also result from differences in the timing of measurements after stroke. Therapists may have high expectations of (spontaneous) neurological recovery during the first days after stroke, in which rehabilitation still focuses on restoring function. Conversely, the data from Kwakkel et al.[21] suggest that therapists are more pessimistic in their expectations about recovery at a later stage, in which rehabilitation probably focuses more on the activity or participation level.

Interestingly, the present study suggests that the number of years of experience in treating patients who have had a stroke, and attending specific courses in the field of stroke rehabilitation, does not affect the accuracy of prediction about UL outcome. In line with this, Wainwright et al.[38] suggested that novel PTs (<1 year experience) and experienced PTs (>8 year of experience) often make similar clinical decisions, though they seem to use different thought processes, reflective of their levels of experience, to arrive at their decisions.[38] However, the median (interquartile range) for years of experience of the therapists in the current study was 24 (23–27). This relatively high level of experience might explain our finding that the factor 'years of experience' does not influence the accuracy of prediction

Although the knowledge about therapists' thoughts underlying the predictions they make is still very speculative, it is likely that the PTs in the current study largely made use of an intuitive approach, since they were unable to make use of the previously published CPM.[7] The question remains therefore whether the accuracy of their predictions can be further improved by incorporating knowledge obtained from validated CPMs as well. It is plausible that the use of a prediction model might complement therapist intuition, and improve the accuracy of prediction. This could be particularly true for novice therapists, who often make use of scientific information in making decisions in order to reduce the level of uncertainty with which they are confronted.[38] Future studies are needed to cross-validate our derived CPM to test for generalizability. Subsequently, impact studies are needed to quantify whether the use of these algorithms in routine practice improves decision making and, ultimately, may improve patient outcomes and reduce costs of care.[15, 39]

The increased accuracy of PTs' predictions of final outcome at discharge in the current study suggests that familiarity with patients' actual performance and the observed time-dependent recovery in the first days may contribute to a better appraisal of functional outcome of the upper paretic limb at 6 months post stroke. However, although the computational model for the moment of discharge also takes into account the observed performance and improvement on clinical prognostic determinants, its accuracy did not improve significantly over time. This finding probably reflects the difficulty of capturing non-linear, time-dependent changes in a regression model. The improvement in PTs' predictions about the outcome at 6 months between 72 hours and the moment of discharge might also be explained by a better view on non-assessed outcomes such as mood and motivation to train.

Finally, the current study focused on prediction of UL capacity. It is likely that therapists are less accurate in predicting outcome in terms of ADL[18, 40] or walking ability.[14, 40] The potential for upper limb recovery largely depends on the intactness of at least some fibers of the corticospinal tract system originating in the affected hemisphere.[41] As a consequence, relatively speaking, upper limb recovery may be easier to predict in the first weeks post stroke. Regarding ADL and walking ability, which can also improve as a result of the use of compensation strategies, such as alternative movement patterns or the use of environmental adaptations, accurate prediction of recovery may be much more difficult.

A limitation of the current study was that the PTs who made the predictions also administered the extensive EPOS test battery within 72 hours after stroke onset, as well as every 3 days until the moment of discharge from the hospital stroke unit. Administering such an extended test battery is not common practice in hospital stroke units. Second, although the therapists had no knowledge of the previously published CPM,[7] several courses and conferences in The Netherlands in the last few years have emphasized that early return of synergism[42] and wrist and finger extension[43] are important indicators of favorable UL recovery. As a consequence, the PTs were probably well aware of the predictability of UL capacity. Hence, it is possible that the accuracy of the PTs' predictions was influenced by this information.

Finally, we must be aware that CPMs are developed on group data. However, there are substantial inter-individual differences in recovery patterns. As a consequence, it is a challenge to make accurate predictions for individuals, based on group data. Therefore, future studies should aim to develop models in which data on performance of individual patients can be applied, combined with data from new observations when these data become available.[44] Since non-linearity can be modelled when applying such models, more customized CPMs can be created in the future.

ACKNOWLEDGEMENTS

Sources of support

This study was part of the EPOS research project funded by the 'Wetenschappelijk College Fysiotherapie' (WCF: No. 33368) of the Royal Dutch Society for Physical Therapy (KNGF), The Netherlands. This part of the EPOS project is co-financed by ZON-MW (grant No. 89000001) as a part of the EXPLICIT-stroke program (www.explicit-stroke.nl).

EPOS Investigators

Executive committee: Prof. G. Kwakkel (PhD), principal investigator, Department of Rehabilitation Medicine, Research Institute MOVE, VU University Medical Center, Amsterdam, The Netherlands; and B.C. Harmeling-van der Wel, *co-principal investigator*, Department of Rehabilitation Medicine and Physical Therapy, Erasmus MC University Medical Center Rotterdam, The Netherlands. *Steering committee and data management:* J.M. Veerbeek and R.H.M Nijland, Department of Rehabilitation Medicine, Research Institute MOVE, VU University Medical Centre, Amsterdam. *Revision of the manuscript:* E.E.H. van Wegen (PhD), Department of Rehabilitation Medicine, Research Institute MOVE, VU University Medical Centre, Amsterdam. *Monitoring board:* M.A. van der Beek, UMC Utrecht; W.A.M. Cornelissen, AMC Amsterdam; A.A.G. Goos, Franciscus Hospital, Roosendaal; C.S. Steeg, UMC Sint Radboud, Nijmegen; J.M. Timmermans, LUMC, Leiden; R. Tichelaar, Amphia Hospital, Breda.

REFERENCES

(1) WHO. *World health report 2003*. World Health Organization; 2003.

(2) Dromerick AW, Lang CE, Birkenmeier R, Hahn MG, Sahrmann SA, Edwards DF. Relationships between upper-limb functional limitation and self-reported disability 3 months after stroke. *J Rehabil Res Dev* 2006 May;43(3):401-8.

(3) Franceschini M, La PF, Agosti M, Massucci M. Is health-related-quality of life of stroke patients influenced by neurological impairments at one year after stroke? *Eur J Phys Rehabil Med* 2010 September;46(3):389-99.

(4) Nichols-Larsen DS, Clark PC, Zeringue A, Greenspan A, Blanton S. Factors influencing stroke survivors' quality of life during subacute recovery. *Stroke* 2005 July;36(7):1480-4.

(5) Stinear C. Prediction of recovery of motor function after stroke. *Lancet Neurol* 2010 December;9(12):1228-32.

(6) Summers D, Leonard A, Wentworth D, Saver JL, Simpson J, Spilker JA, Hock N, Miller E, Mitchell PH. Comprehensive overview of nursing and interdisciplinary care of the acute ischemic stroke patient: a scientific statement from the American Heart Association. *Stroke* 2009 August;40(8):2911-44.

(7) Nijland RH, van Wegen EE, 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 April;41(4):745-50.

(8) Nijland R, Kwakkel G, Bakers J, van Wegen E. Constraint Induced Movement Therapy for the upper paretic limb in acute or subacute stroke: A systematic review. *Int J Stroke* 2011 October;6(5):425-33.

(9) van Delden AE, Peper CE, Beek PJ, Kwakkel G. Unilateral versus bilateral upper limb exercise therapy after stroke: a systematic review. *J Rehabil Med* 2012 February;44(2):106-17.

(10) Wolf SL, Winstein CJ, Miller JP, Taub E, Uswatte G, Morris D, Giuliani C, Light KE, Nichols-Larsen D. Effect of constraint-induced movement therapy on upper extremity function 3 to 9 months after stroke: the EXCITE randomized clinical trial. *JAMA* 2006 November 1;296(17):2095-104.

(11) Coupar F, Pollock A, Rowe P, Weir C, Langhorne P. Predictors of upper limb recovery after stroke:a systematic review and meta-analysis. *Clin Rehabil* 2012 April;26(4):291-313.

(12) Murphy TH, Corbett D. Plasticity during stroke recovery: from synapse to behaviour. *Nat Rev Neurosci* 2009 December;10(12):861-72.

(13) Kwakkel G, Veerbeek JM, Harmeling-van der Wel BC, van WE, Kollen BJ. Diagnostic accuracy of the Barthel Index for measuring activities of daily living outcome after ischemic hemispheric stroke: does early poststroke timing of assessment matter? *Stroke* 2011 February;42(2):342-6.

(14) Veerbeek JM, van Wegen EE, Harmeling-van der Wel BC, Kwakkel G. Is accurate prediction of gait in nonambulatory stroke patients possible within 72 hours poststroke?: The EPOS study. *Neurorehabil Neural Repair* 2011 March;25(3):268-74.

(15) Moons KG, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *BMJ* 2009;338:b606.

(16) Croskerry P. A universal model of diagnostic reasoning. *Acad Med* 2009 August;84(8):1022-8.

(17) Counsell C, Dennis M. Systematic review of prognostic models in patients with acute stroke. *Cerebrovasc Dis* 2001;12(3):159-70.

(18) Veerbeek JM, Kwakkel G, van Wegen EE, Ket JC, Heymans MW. Early prediction of outcome of activities of daily living after stroke: a systematic review. *Stroke* 2011 May;42(5):1482-8.

(19) Buckingham CD, Adams A. Classifying clinical decision making: interpreting nursing intuition, heuristics and medical diagnosis. *J Adv Nurs* 2000 October;32(4):990-8.

(20) Kwakkel G, Kollen B, Twisk J. Impact of time on improvement of outcome after stroke. *Stroke* 2006 September;37(9):2348-53.

(21) Kwakkel G, van Dijk GM, Wagenaar RC. Accuracy of physical and occupational therapists' early predictions of recovery after severe middle cerebral artery stroke. *Clin Rehabil* 2000 February;14(1):28-41.

(22) Stroke--1989. Recommendations on stroke prevention, diagnosis, and therapy. Report of the WHO Task Force on Stroke and other Cerebrovascular Disorders. *Stroke* 1989 October;20(10):1407-31.

(23) van Peppen RP, Hendriks HJ, van Meeteren NL, Helders PJ, Kwakkel G. The development of a clinical practice stroke guideline for physiotherapists in The Netherlands: a systematic review of available evidence. *Disabil Rehabil* 2007 May 30;29(10):767-83.

(24) Lyle RC. A performance test for assessment of upper limb function in physical rehabilitation treatment and research. *Int J Rehabil Res* 1981;4(4):483-92.

(25) van der Lee JH, De G, V, Beckerman H, Wagenaar RC, Lankhorst GJ, Bouter LM. The intra- and interrater reliability of the action research arm test: a practical test of upper extremity function in patients with stroke. *Arch Phys Med Rehabil* 2001 January;82(1):14-9.

(26) Yozbatiran N, Der-Yeghiaian L, Cramer SC. A standardized approach to performing the action research arm test. *Neurorehabil Neural Repair* 2008 January;22(1):78-90.

(27) Bamford J, Sandercock P, Dennis M, Burn J, Warlow C. Classification and natural history of clinically identifiable subtypes of cerebral infarction. *Lancet* 1991 June 22;337(8756):1521-6.

(28) de Groot V, Beckerman H, Lankhorst GJ, Bouter LM. How to measure comorbidity. a critical review of available methods. *J Clin Epidemiol* 2003 March;56(3):221-9.

(29) Brott T, Adams HP, Jr., Olinger CP, Marler JR, Barsan WG, Biller J, Spilker J, Holleran R, Eberle R, Hertzberg V, . Measurements of acute cerebral infarction: a clinical examination scale. *Stroke* 1989 July;20(7):864-70.

(30) Collin C, Wade DT, Davies S, Horne V. The Barthel ADL Index: a reliability study. *Int Disabil Stud* 1988;10(2):61-3.

(31) Collin C, Wade D. Assessing motor impairment after stroke: a pilot reliability study. *J Neurol Neurosurg Psychiatry* 1990 July;53(7):576-9.

(32) Sanford J, Moreland J, Swanson LR, Stratford PW, Gowland C. Reliability of the Fugl-Meyer assessment for testing motor performance in patients following stroke. *Phys Ther* 1993 July; 73(7):447-54.

(33) Franchignoni FP, Tesio L, Ricupero C, Martino MT. Trunk control test as an early predictor of stroke rehabilitation outcome. *Stroke* 1997 July;28(7):1382-5.

(34) Beebe JA, Lang CE. Active range of motion predicts upper extremity function 3 months after stroke. *Stroke* 2009 May;40(5):1772-9.

(35) Fritz SL, Light KE, Patterson TS, Behrman AL, Davis SB. Active finger extension predicts outcomes after constraint-induced movement therapy for individuals with hemiparesis after stroke. *Stroke* 2005 June;36(6):1172-7.

(36) Katrak P, Bowring G, Conroy P, Chilvers M, Poulos R, McNeil D. Predicting upper limb recovery after stroke: the place of early shoulder and hand movement. *Arch Phys Med Rehabil* 1998 July;79(7):758-61.

(37) Smania N, Paolucci S, Tinazzi M, Borghero A, Manganotti P, Fiaschi A, Moretto G, Bovi P, Gambarin M. Active finger extension: a simple movement predicting recovery of arm function in patients with acute stroke. *Stroke* 2007 March;38(3):1088-90.

(38) Wainwright SF, Shepard KF, Harman LB, Stephens J. Factors that influence the clinical decision making of novice and experienced physical therapists. *Phys Ther* 2011 January;91(1):87-101.

(39) Reilly BM, Evans AT. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. *Ann Intern Med* 2006 February 7;144(3):201-9.

(40) Kollen B, Kwakkel G, Lindeman E. Functional recovery after stroke: a review of current developments in stroke rehabilitation research. *Rev Recent Clin Trials* 2006 January;1(1):75-80.

(41) Stinear CM, Barber PA, Smale PR, Coxon JP, Fleming MK, Byblow WD. Functional potential in chronic stroke patients depends on corticospinal tract integrity. *Brain* 2007 January;130(Pt 1):170-80.

(42) Kwakkel G, Kollen BJ, van der Grond J, Prevo AJH. Probability of regaining dexterity in the flaccid upper limb: impact of severity of paresis and time since onset in acute stroke. *Stroke* 2003 September;34(9):2181-6.

(43) Kwakkel G, Kollen B. Predicting improvement in the upper paretic limb after stroke: a longitudinal prospective study. *Restor Neurol Neurosci* 2007;25(5-6):453-60.

(44) Rajaraman S, Gribok AV, Wesensten NJ, Balkin TJ, Reifman J. Individualized performance prediction of sleep-deprived individuals with the two-process model. *J Appl Physiol* 2008 February;104(2):459-68.

4