



ORIGINAL RESEARCH

Assessment of Cognitive and Neurologic Recovery in Ischemic Stroke Drug Trials: Results from a Randomized, Double-blind, Placebo-controlled Study

FUNDING: The A9541004 study was sponsored and financed by Pfizer.

TRIAL REGISTRATION: Clinicaltrials.gov. Unique identifier: NCT01208233. EudraCT Number: 201002141432

FINANCIAL DISCLOSURES: Dr. Di Cesare provided medical and scientific clinical trial services throughout the period 2011–2014, including being part of the medical oversight team of the stroke trial. Drs. Mancuso, Loudon are employees of Pfizer. Dr. Silver has no conflicts of interest relevant to the content of this article.

ADDRESS CORRESPONDENCE TO:

Dr. Peter Loudon, Pfizer Ltd., Building GP4, Granta Park, Great Abington, Cambridgeshire CB21 6GP UK; Phone: +44.1304.640778; Email: peter.t.loudon@pfizer.com

KEY WORDS: Ischemic stroke, clinical trial, cognitive assessment, memory, stroke recovery, neurorestoration, PDE5 Inhibitor

by **FRANCO DI CESARE, MD; JESSICA MANCUSO, PhD; BRIAN SILVER, MD; and PETER T. LOUDON, PhD**

Dr. Di Cesare is with Leoben Research Limited in Glasgow, UK; Dr. Mancuso is with Research Statistics at Pfizer in Groton, Connecticut, USA; Dr. Silver is with Rhode Island Hospital and Alpert Medical School of Brown University in Providence, Rhode Island, USA; and Dr. Loudon is with the Neuroscience and Pain Research Unit at Pfizer in Cambridge, UK.

Innov Clin Neurosci. 2016;13(9–10):32–43

ABSTRACT

Objective. Ischemic stroke is a serious medical condition with limited therapeutic options. The evaluation of the therapeutic potential of novel pharmacological interventions is carried-out in Phase II trials. The study design, primarily intended to evaluate efficacy and safety, is a balance between utilizing as few patients as possible to minimize safety risk and enrolling sufficient patients to detect unambiguous efficacy signals. We sought to determine whether post-stroke recovery outcomes based on behavioral measures of cognitive and motor impairment yielded additional information beyond that of clinician-based methods.

Design. This was a multicenter, multinational, randomized, parallel group, controlled versus placebo, efficacy, and safety study of PF-03049423 for treatment of acute ischemic stroke.

Settings and participants. Our study subjects were acute ischemic stroke inpatients.

Measurements. Outcome measures were derived from rating scales (Modified Rankin Scale, Barthel Index, and National Institutes of Health Stroke Scale) and behavioral tests (Box and Blocks Test, Hand Grip Strength Test, 10-Meter Walk Test, Repeatable Battery Assessment of Neuropsychological Status Naming and Coding Subtests, Line Cancellation Test, and Recognition Memory Test). Assessments were performed at Days 7, 14, 30, 60, and 90. *Post-hoc* analyses of correlations among the outcome measures at each measurement time point on a cohort of 137 subjects were conducted.

Results. Results support the validity of measures from Box and Blocks Test, Hand Grip Strength Test, 10-Meter Walk Test, and Repeatable Battery Assessment of

Neuropsychological Status Coding Subtests to monitor post-stroke recovery in clinical trial settings. Notably, the Recognition Memory Test did not show a correlation with the Modified Rankin Scale, and, in fact, did not show improvement over time.

Conclusion. The behavioral measures of cognitive and motor functions included in this study may extend the evaluation of the therapeutic potential of new treatments for stroke recovery. The lack of correlation between Recognition Memory Test and the traditional efficacy endpoints, at least in part due to absence of any improvement in recognition memory, suggests that there may be cognitive elements not detected by the Modified Rankin Scale. This is clinically relevant and memory improvement has potential as an endpoint in future trials aiming to improve certain aspects of cognition.

INTRODUCTION

Ischemic stroke is a serious life-threatening condition frequently resulting in severe residual disability.¹ Currently, there are limited therapeutic options, creating a compelling need for more efficacious and safer treatment strategies.² The evaluation of the therapeutic potential of novel pharmacological interventions is carried out in Phase II trials. In this particular clinical research setting, the study design is a balance between selecting as small a sample size as possible in order to minimize the unknown safety risks and the need to detect unambiguous signals of efficacy. Traditionally, recovery up to 90 days has been accepted as an adequate follow-up period to evaluate response. The Modified Rankin Scale (mRS)³ or the Barthel Index (BI)⁴ have been the predominant primary efficacy outcome measures used to evaluate treatment response and for sample size determination. Other scales for stroke recovery are used to evaluate other aspects of impairment (e.g.,

neurological), usually measured by the National Institutes of Health Stroke Scale (NIHSS).⁵ Evaluations of post-stroke mood alteration and risk for suicide behavior as well as cognitive functioning are routinely included in Phase II study protocols. Alternative approaches of assessment methodology have been debated. The use of “modality-specific outcome measures” as primary endpoints for clinical trials was proposed by Cramer et al.⁶ Primary outcome measures in clinical trials might separately address cognitive recovery or more specific functions (i.e., upper extremity function or gait) and “major treatment-induced improvements on modality specific outcome measures should be mirrored by improved quality of life.”⁶ Moreover, these measures may allow for reduced sample size to evaluate the therapeutic potential, thereby reducing safety risks for study subjects as well as time and costs of drug development.

This article presents the study results and methodological considerations on the validity of modality specific outcome measures of cognitive and motor recovery that may be helpful to design future clinical trials for stroke recovery-promoting treatments. We sought to determine whether post-stroke recovery assessment based on different domains of functional impairment (cognition and motor function) yielded additional information beyond that of traditional outcome measures of ischemic stroke studies.

METHODS

Clinician-based and modality specific behavior-based assessment strategies were combined in the study design of a recently completed phase II stroke trial.⁷ This was a multinational, multicenter, randomized, double-blind, placebo-controlled study in acute ischemic stroke to evaluate the therapeutic potential in stroke recovery of PF-03049423,⁷ a PDE5 inhibitor. The

study was approved by the competent Regulatory Authorities and Independent Ethics Committees/Review Boards (IECs/IRBs). Main clinical criteria for eligibility included the following: age of 18 to 85 years, supratentorial, ischemic, nonhemorrhagic infarct involving the cortex (strokes involving more than one area were allowed as long as there was documented ischemic cortical involvement); baseline NIHSS of 6 to 20 inclusive; new onset of upper extremity paresis or paralysis on the affected side (NIHSS Item 5, Score 1–4 inclusive); and absence of clinically significant depressive symptoms as documented by Patient Health Questionnaire 8-Item (PHQ-8).⁸ Initial dose of study drug was administered 24 to 78 hours from stroke onset. Subjects who had received thrombolytic therapy were enrolled as long as they had been stable to improving post-thrombolytic treatment. Subjects had to participate in a rehabilitation program. Further details on study objectives and design, subject eligibility, blinding and randomization, outcome measures, enrollment, and participating sites as well as summary of study results can be found at ClinicalTrials.gov identifier: NCT01208233. In the “Proof-of-Concept” part of the trial, subjects were randomly assigned 1:1 to receive PF-03049423, 6mg once a day or placebo once a day for 90 consecutive days. Post-baseline evaluations were carried-out at Days 7, 14, 30, 60, and 90. Assessment methods included mRS, BI, and NIHSS as well as the behavioral testing procedures described herein to monitor post-stroke recovery.

All clinical assessments were administered by raters with appropriate clinical training and certification. The study design used a number of measures to minimize the effect of rater variability on the reliability and validity of functional and behavioral outcome measures. For example, sites were asked to ensure that the same rater

administered assessments to the same patient throughout the study; in addition, the protocol set the requirement for study-specific rater training and qualification. The validity of outcomes was then systematically assessed throughout the study by a dual level data review involving a contract research organisation specialising in cognitive and behavioral assessments (provider of rater training, qualification, and psychometric monitoring services) and Pfizer medical monitoring oversight team, including a reviewer with extensive experience in neuropsychology assessment methodologies, rater training, and reliability assessment (one of the authors: FDC). Interrater reliability was evaluated for clinical (i.e., NIHSS, Barthel), functional (i.e., Box-Block), and cognitive measures. Methods included the expert clinical review as well as the application of statistical analysis, such as standard statistical measures (i.e., kappa statistics and correlation coefficients). No formal report of these analyses was issued. Notably, in no cases did analyses highlight findings suggesting a significant quality issue related to clinical data integrity resulting in corrective actions. For all neuropsychological assessments where a patient was unable to complete a test (e.g., a patient with speech or comprehension difficulties unable to complete a verbal assessment), the result was recorded as “not done” rather than assigning an arbitrary value.

Repeated Battery Assessment of Neuropsychological Status (RBANS)^{10,11} Naming Subtest.

This test requires naming correctly 10 objects drawn in ink. The subject is asked to identify each picture and has 20 seconds to respond to each picture presented. The score range is 0 to 10, with 10 being the highest level of performance.

RBANS Coding Subtest. This test consists of a simple substitution task. The subject is asked to correctly pair specific digits with

given geometric figures using a reference key and within a time limit of 90 seconds. Responses can be written or oral. The outcome measure is the total number of correct responses; higher scores denote better performance.

Line Cancellation Test (LCT).

The subject is presented with a paper sheet that has 28 lines placed across, equally distributed, 14 on the left side and 14 on the right side. The subject is required to cross out all the 28 lines (targets) using his or her non-paretic hand. Two outcome measures are derived: 1) marker of attention is the number of crossed lines as a percentage of the total number of targets $[(L+R)/28] \times 100\%$, where L = number of lines crossed on the left side of the sheet and R = number of lines crossed on the right side of the sheet; 2) marker of hemi-inattention is determined as $[(L/14) \times 100\%]$. For both outcome measures, the score ranges from 0 (worst performance) to 100.

Recognition Memory Test (RMT). This test consists of a delayed nonverbal memory recognition task. The subject is presented with a series of pictures, a subset of which are the objects presented in the Naming Subtest. After each picture is presented, the subject is given five seconds to indicate whether the picture was seen previously. The outcome measure is the total number of pictures correctly recognized; the score range is 0 to 10, with 10 being the best performance.

Box and Blocks Test (B&BT).¹²

The Box and Blocks apparatus consists of a box divided into two sections and one-inch hardwood blocks. The blocks begin in the compartment of the test box to the dominant side of the subject. The subject is asked to transfer the blocks one at a time to the other side of the box as fast as possible in 60 seconds using the non-paretic hand. The box is then turned so all the blocks are in the same side as the paretic hand. The subject is then asked to complete the task with his

or her paretic hand. The outcome measure is the number of blocks moved. Two measures are reported: blocks moved by paretic limb and by non-paretic limb.

Hand Grip Strength Test (HGST).¹³ The HGST measures the maximum isometric strength of the hand and forearm muscles. The subject is asked to squeeze the dynamometer (a Jamar hydraulic hand dynamometer was used in our study) with maximum isometric effort while sitting with shoulder adducted and neutrally rotated, elbow flexed at 90 degrees and the forearm in neutral position, and wrist between 0 to 30 degrees dorsiflexion and a 0 to 15 degrees ulnar deviation. The subject completes this task three times with each hand, starting with the non-paretic hand. The outcome measure is the average score in pounds (lbs) of pressure exerted. Two measures are reported: strength by paretic limb and by non-paretic limb.

10-Meter Walk Test.¹⁴ The 10-Meter Walk Test (10-m-WT) requires a 20m straight path, with 5m for acceleration, 10m for steady-state walking, and 5m for deceleration. Markers are placed at the 5m and 15m positions along the path. The subject begins to walk at a comfortable pace at one end of the path, and continues walking until he or she reaches the other end. A stopwatch is used to determine how much time it takes to traverse the 10m center of the path, starting the stopwatch as soon as the subject's limb crosses the first marker and stopping the stopwatch as soon as the subject's limb crosses the second marker.

Statistical considerations and analysis. Based on the pre-specified futility criteria assessed at an interim analysis, the Sponsor made the decision to immediately stop the study. In the absence of any therapeutic effect of PF-03049423 all analyses reported here were on the combined set of patients who received PF-03049423 and placebo. A number of additional *post-hoc*

analyses were carried out to further assess the performance of the cognitive and motor endpoints that were included in the study design. Results presented in the following refer to the “Inferential Full Analysis Set” (I-FAS) of 137 subjects, combining those randomized to PF-03049423 6mg (N=70) or placebo (N=67) in Cohort 3 who took any study medication. Descriptive statistics (*n*, mean, median, standard deviation, coefficient of variation, min, and max) were provided for each of the outcome measures. “Change from Baseline (Day 1, predose)” across time points was analyzed as a continuous endpoint using mixed model repeated measures (MMRM). The MMRM analysis included treatment, t-PA usage for current stroke, geographical region [investigational sites were clustered based on geographical regions: 1=Asia (contributing with 38 randomized subjects), 2=America (22 subjects), 3=Eastern Europe (51 subjects), 4=Western Europe (26 subjects)], time to study treatment initiation from onset of stroke symptoms, visit and treatment-by-visit interaction as fixed effects, baseline NIHSS score and age as continuous covariates, and subject within treatment and a within subject residual error as random effects. Statistical contrasts were used to make treatment comparisons at Day 90 and/or other time-points. The estimated treatment difference and the corresponding two-sided 80-percent confidence interval were constructed. The validity of motor and cognitive outcomes was evaluated by correlation analysis (Pearson’s coefficient of correlation *r*). The strength of correlation was categorized as weak (*r* range: 0.10–0.39), moderate (*r* range: 0.40–0.69), and strong (*r*≥0.70).¹⁵ It should be noted that the study was solely powered to compare the treatment effect of PF-03049423 6 mg versus placebo based on the primary efficacy endpoint (mRS responder rate at Day 90), and it was

not powered for other purposes such as testing effects of covariates included in the statistical models. For all other statistical analyses that were performed, we used a conventional 0.05 level of statistical significance to report the results, and no multiple comparison adjustment was made. For these reasons, one should not interpret any non-significant results found in these additional analyses as strong evidence of “no effect,” and due to the issue of multiplicity one should be cautious when interpreting *p*-values less than 0.05.

RESULTS

The main demographic and clinical characteristics at screening and baseline are summarized in Table 1.

PF-03049423 did not show any significant clinical benefit on activity limitation or functional impairment recovery as assessed by the primary efficacy endpoint, mRS responder rate (Score 0–2 at Day 90). The primary efficacy analysis of mRS using logistic regression showed no statistically significant difference between PF-03049423 6mg and placebo (responder rate at Day 90 of 42.6% and 46.2%, respectively).⁷ Similarly, the logistic regression analysis on the responder rate as defined by BI≥95 at Day 90 did not show any significant difference between PF-03049423 6mg and placebo (responder rate at Day 90 of 47.1% and 40.0%, respectively). Overall, no statistically significant benefit was observed for PF-03049423 6mg versus placebo at any time throughout the treatment period, as indicated by statistical analyses of outcome measures derived from mRS, NIHSS, BI, B&BT, HGST, 10-m-WT, RBANS Coding and Naming Subtests, LCT, and RMT.

Behavioral functional recovery. A summary of mRS scores across time points is reported in Table 2.

Modified Rankin Scale score decrease reflected a pattern of progressively increased overall ability

TABLE 1. Summary of characteristics at screening and baseline

Number of subjects	137
Age in years, mean (SD)	64.9 (12.2)
Gender (M/F, n)	83/54
Race/Ethnicity n (%)	
Asian	39 (28.5)
Black	9 (6.6)
White	84 (61.3)
Other	5 (3.6)
Stroke onset to treatment start in hours, Mean (SD)	60.5 (12.4)
Subjects received thrombolytic therapy, n (%)	40 (29.2)
NIHSS score, Mean, (SD), Median, Range	10.8 (3.9), 10, 6–20
PHQ-8 score, Mean (SD), Range	2.7 (3.0), 0–11

SD: standard deviation; NIHSS: National Institutes of Health Stroke Scale; PHQ-8: Patient Health Questionnaire

to perform everyday life activities following stroke onset (Figure 1).

The pattern was post-stroke time dependant when the mRS score was analyzed as a continuous endpoint with MMRM (*p*<0.0001). Based on the logistic regression analysis, functional recovery as measured by the mRS responder rate was dependent on baseline NIHSS score (*p*<0.0001) and the geographical region (*p*=0.014). None of the other factors were statistically significant at 0.05 level [gender (*p*=0.08), time to study treatment initiation from onset of stroke symptoms (*p*>0.10), t-PA usage for current stroke (*p*>0.10), age (*p*>0.10)].

Cognitive recovery. A summary of cognitive test scores at different post-stroke time points are reported in Figure 1 and Table 2. RBANS

TABLE 2. Summary of disability and cognitive scores [mean (SD), CV (%), n] across time points

MEASURES	BASELINE	DAY 7	DAY 14	DAY 30	DAY 60	DAY 90
mRS						
Mean (SD)	*	3.6 (1.3)	3.3 (1.4)	3.1 (1.5)	2.8 (1.6)	2.6 (1.7)
CV (%), n	n/a	36.4, 132	41.2, 123	48.6, 119	56.1, 114	65.7, 103
RBANS Naming						
Mean (SD)	7.7 (2.5)	8.2 (2.3)	8.2 (2.3)	8.4 (2.1)	8.7 (1.9)	8.7 (2.0)
CV (%), n	32.4, 109	28.5, 113	27.4, 106	24.7, 106	21.6, 97	22.7, 91
RBANS Coding						
Mean (SD)	11.0 (8.1)	14.5 (9.8)	16.2 (10.8)	18.9 (11.4)	20.5 (11.4)	20.9 (11.5)
CV (%), n	73.6, 86	67.7, 93	66.5, 92	60.2, 92	55.8, 89	54.82, 84
LCT Attention						
Mean (SD)	82.9 (28.5)	87.0 (26.8)	93.3 (19.1)	94.7 (16.6)	96.2 (12.7)	98.0 (9.4)
CV (%), n	34.4, 101	30.8, 111	20.5, 103	17.5, 102	13.2, 100	9.6, 93
LCT/H						
Mean (SD)	79.9 (33.2)	83.5 (32.2)	91.7 (24.0)	93.1 (21.3)	95.3 (16.9)	97.4 (12.7)
CV (%), n	41.5, 101	38.6, 111	26.2, 103	22.8, 102	17.7, 100	13.1, 93
RMT						
Mean (SD)	10.1 (2.4)	9.5 (2.1)	9.8 (1.2)	9.9 (1.9)	9.7 (1.7)	9.8 (1.0)
CV (%), n	23.9, 112	21.7, 117	12.3, 107	19.0, 106	17.8, 100	10.3, 91

(*) The study protocol did not require mRS to be carried out at baseline.

n/a: not applicable; n: number of subjects used in the analysis; SD; Standard Deviation; CV: Coefficient of Variation; RBANS: Repeatable Battery for the Assessment of Neuropsychological Status; LCT: Line Cancellation Test; LCT/H: Line Cancellation Test, Hemi-inattention; RMT: Recognition Memory Test; mRS: modified Rankin Scale

Naming Subtest score increase relative to baseline described a pattern of increased overall naming ability following stroke onset, although the increase in score was less evident than in other cognitive outcomes. Based on the MMRM analysis, the pattern was post-stroke time dependant ($p=0.006$). None of the other factors in the MMRM model were statistically significant at the 0.05 level [treatment ($p>0.10$), t-PA usage for current stroke ($p>0.10$), geographical region ($p=0.07$), gender ($p=0.07$), time to study treatment initiation from onset

of stroke symptoms ($p>0.10$), baseline NIHSS score ($p>0.10$), and age ($p>0.10$)]. RBANS Coding Subtest score increase versus baseline described a pattern of increased overall ability to perform the task following stroke onset. Based on the MMRM analysis, the pattern was post-stroke time dependant ($p<0.0001$). Recovery was dependent upon geographical region ($p=0.038$) and age ($p<0.0001$). Other factors in the model were not statistically significant at the 0.05 level (all of them with $p>0.10$). LCT Hemi-

inattention score increase relative to baseline described a pattern of increased overall ability to perform this task following stroke onset. Based on the MMRM analysis, the measure of hemi-inattention showed time dependency ($p=0.038$) and was influenced by baseline neurologic impairment (NIHSS total score, $p=0.025$); other factors in the model were not statistically significant at the 0.05 level (all of them with $p>0.10$). LCT Attention was dependent on the NIHSS baseline score ($p<0.001$); other factors in the model were not statistically

significant at the 0.05 level [time ($p=0.054$), geographical region ($p=0.092$), all others ($p>0.10$)]. LCT Attention and Hemi-inattention scores showed a strong correlation with $r>0.90$ with $p<0.0001$ across study time-points. RMT score did not provide evidence of recovery of memory ability following stroke onset. Delayed memory recognition was dependent on treatment ($p=0.015$) with evidence of lower performance in the PF-03049523 group (Table 3).

Other factors in the model did not appear to significantly influence memory recovery as evaluated by this non-verbal delayed recognition task [tPA ($p=0.065$), all others ($p>0.10$)].

Motor recovery. A summary of motor measures (B&BT, HGST, 10-m-WT) across different post-stroke time points is reported in Figure 1 and in Table 4.

B&BT paretic and non-paretic limb scores increased relative to baseline, and described a pattern of increased overall ability to perform the tasks following stroke onset. Analysis indicated that these patterns were post-stroke time dependant (paretic: $p<0.0001$; non-paretic: $p<0.0001$). Recovery of dexterity of the paretic upper limb was influenced by t-PA usage for current stroke ($p=0.009$); other factors were not statistically significant (all of them with $p>0.10$). In contrast, the recovery of dexterity of the non-paretic upper limb was dependent on neurologic impairment (NIHSS total score) at baseline ($p=0.042$); other factors were not statistically significant at the 0.05 level (t-PA usage for current stroke, time to study treatment initiation from onset of stroke symptoms, gender, treatment, age, and geographical region: all of them with $p>0.10$). HGST paretic and non-paretic limb scores described a pattern of increased ability to perform the task following stroke onset. These patterns were post-stroke time dependant (paretic: $p<0.0001$; non-paretic: $p<0.0001$);

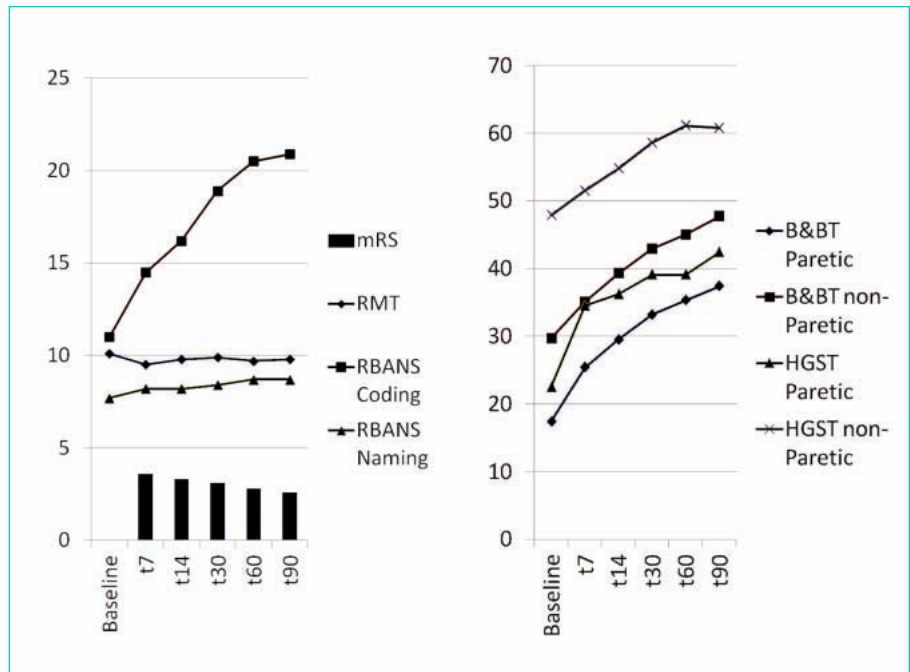


FIGURE 1. Clinical domains of post-stroke recovery

RBANS: Repeatable Battery for the Assessment of Neuropsychological Status; RMT: Recognition Memory Test; B&BT: Box and Blocks Test; HGST: Hand Grip Strength Test, exerted pressure in lbs; 10-m-WT: 10-Meter Walk Test; mRS: modified Rankin Scale

TABLE 3. Summary of analysis MMRM of change from baseline in Recognition Memory Test score across time points

VISIT	TREATMENT	N	DIFFERENCE FROM PLACEBO			
			LS MEAN	STANDARD ERROR	80% CI	P VALUE
Day 7	PF 03049423	58	-1.141	0.5114	(-1.801, -0.481)	0.0279
	Placebo	48			n/a	
Day 14	PF 03049423	52	-1.317	0.495	(-1.955, -0.678)	0.0092
	Placebo	43			n/a	
Day 30	PF 03049423	53	-0.724	0.5655	(-1.454, 0.005)	0.2032
	Placebo	40			n/a	
Day 60	PF 03049423	51	-1.314	0.5122	(-1.974, -0.653)	0.0118
	Placebo	37			n/a	
Day 90	PF 03049423	44	-1.279	0.5041	(-1.929, -0.629)	0.0128
	Placebo	37			n/a	

CI: confidence interval; LS: least squares; MMRM: mixed model repeated measures; n: number of subjects; n/a: not applicable

TABLE 4. Summary of motor function scores [mean(SD), CV(%), *n*] across time points

MEASURES	BASELINE	DAY 7	DAY 14	DAY 30	DAY 60	DAY 90
B&BT paretic						
Mean (SD)	17.4 (14.2)	25.4 (17.3)	29.5 (17.8)	33.2 (18.1)	35.3 (17.6)	37.4 (17.8)
CV (%), <i>n</i>	81.9, 65	68.2, 73	60.3, 69	54.4, 69	50.0, 72	47.5, 69
B&BT non-paretic						
Mean (SD)	29.7, (14.0)	35.1 (14.3)	39.3 (15.1)	42.9 (14.8)	45.0 (14.9)	47.7 (15.4)
CV (%), <i>n</i>	47.1, 117	40.8, 123	38.4, 113	34.5, 106	33.1, 103	32.3, 95
HGST paretic						
Mean (SD)	22.5 (21.1)	34.5 (27.4)	36.2 (26.7)	39.1 (28.3)	39.1 (28.7)	42.4 (29.6)
CV (%), <i>n</i>	93.6, 76	79.3, 78	73.8, 74	72.4, 74	73.3, 77	69.7, 71
HGST non-paretic						
Mean (SD)	47.9, (27.3)	51.5 (28.4)	54.8 (26.9)	58.6 (28.2)	61.1 (27.0)	60.8 (27.8)
CV (%), <i>n</i>	56.9, 123	55.2, 126	49.1, 115	48.1, 112	44.1, 103	45.7, 95
10-m-WT						
Mean (SD)	*	0.7 (0.5)	0.9 (0.6)	0.9 (0.6)	*	1.0 (0.6)
CV (%), <i>n</i>		67.9, 64	68.2, 64	64.7, 77		54.5, 77

(*) The study protocol did not require 10-m-WT to be carried out at Baseline and Day 60.

n: number of subjects used in the analysis; CV: coefficient of variation; SD: standard deviation; B&BT: Box and Blocks Test; HGST: Hand Grip Strength Test, exerted pressure in lbs; 10-m-WT: 10-Meter Walking Test

other factors did not achieve statistical significance at the 0.05 level [paretic: treatment ($p=0.07$), all others ($p>0.10$); non-paretic: age ($p=0.09$), all others ($p>0.10$)]. In the 10-m-WT, gait velocity increase relative to baseline described a pattern of increased ability to perform the task. The pattern was post-stroke time dependant ($p<0.0001$). Gait function recovery was influenced by the baseline NIHSS score ($p=0.001$) and geographical region ($p=0.002$); other factors did not appear to significantly influence this measure at the 0.05 level (all of them with $p>0.10$).

Analysis of correlations.

Clinician-based measures mRS, BI, and NIHSS were convergent measures of overall disability and functional impairment in subjects with acute ischemic stroke resulting in mild-to-moderate neurological

impairment. This was evident by the consistent pattern of strong correlations (absolute value of $r>0.70$; $p<0.0001$) between mRS and BI as well as mRS and NIHSS across all study time-points (Table 5).

A summary of correlations between cognitive and motor function measures with mRS is reported in Table 5. B&BT paretic (r ranges from -0.72 to -0.58, with all $p<0.0001$), HGST paretic (r ranges from -0.62 to -0.52, with all $p<0.0001$), and 10-m-WT [r ranges from -0.51 ($p<0.0001$) to -0.31 ($p=0.0134$)] all showed statistically significant correlations with mRS. In general, RBANS Coding, RBANS Naming, and LCT measures were weakly correlated to mRS. RMT was found to be weakly correlated ($r= -0.21$, $p=0.032$) to mRS at Day 60 but not at other time points (Table 5). A summary of correlations

between cognitive and motor function measures with NIHSS at Day 7 is reported on Table 6.

B&BT paretic and HGST paretic showed a moderate correlation to NIHSS ($r= -0.62$ with $p<0.0001$ and -0.58 with $p<0.0001$, respectively), whereas the correlation between 10-m-WT and NIHSS was weak ($r= -0.27$ with $p=0.029$).

B&BT paretic, HGST paretic, and 10-m-WT were correlated with r values ranging from 0.45 to 0.67 with $p<0.001$. Notably, RBANS Coding and LCT Hemi-attention were associated with NIHSS ($r= -0.33$ with $p=0.0011$ and $r= -0.39$ with $p<0.0001$, respectively) whereas RMT and RBANS Naming measures were not. Finally, RBANS Coding showed weak correlations with B&BT non-paretic and HGST non-paretic ($r= 0.33$ with $p=0.001$ and $r= 0.34$ with $p=0.0007$, respectively).

TABLE 5. Summary of correlations between cognitive and motor measures and mRS across time points [*n*, Pearson's *r*, (*p* value)]

MEASURES	DAY 7	DAY 30	DAY 60	DAY 90
RBANS Naming	113, -0.14 (ns)	106, -0.23 (0.0186)	97, -0.34 (0.0007)	91, -0.20 (ns)
RBANS Coding	93, -0.41 (<0.0001)	92, -0.31 (0.0024)	89, -0.31 (0.0027)	84, -0.28 (0.0095)
RMT	117, -0.15 (ns)	106, 0.16 (ns)	100, -0.21 (0.0325)	91, -0.04 (ns)
LCT Attention	111, -0.32 (0.0006)	102, -0.23 (0.02030)	100, -0.16 (ns)	93, -0.21 (0.0450)
LCT Hemi-inattention	111, -0.35 (0.0002)	102, -0.26 (0.0096)	100, -0.16 (ns)	93, -0.23 (0.0262)
B&BT paretic	73, -0.72 (<0.0001)	69, -0.67 (<0.0001)	72, -0.60 (<0.0001)	69, -0.58 (<0.0001)
B&BT non-paretic	123, -0.58 (<0.0001)	106, -0.52 (<0.0001)	103, -0.51 (<0.0001)	95, -0.45 (<0.0001)
HGST paretic	78, -0.61 (<0.0001)	74, -0.62 (<0.0001)	77, -0.59 (<0.0001)	71, -0.52 (<0.0001)
HGST non-paretic	126, -0.35 (<0.0001)	112, -0.32 (0.0007)	103, -0.33 (0.0007)	95, -0.36 (.0004)
10-m-WT	64, -0.31 (0.0134)	77, -0.51 (<0.0001)	*	75, -0.50 (<0.0001)
PHQ-8	110, 0.22 (0.0203)	98, 0.25 (0.0120)	93, 0.30 (0.0039)	84, 0.24 (0.0303)
Barthel Index	132, -0.89 (<0.0001)	115, -0.86 (<0.0001)	107, -0.82 (<0.0001)	96, -0.80 (<0.0001)
NIHSS	132, 0.70 (<0.0001)	115, 0.76 (<0.0001)	107, 0.72 (<0.0001)	96, 0.72 (<0.0001)

(*) The study protocol did not require 10-m-WT to be carried out on day 60.

n: number of subjects used in the analysis; RBANS: Repeatable Battery for the Assessment of Neuropsychological Status; RMT: Recognition Memory Test; LCT: Line Cancellation Test; B&BT: Box and Blocks Test; HGST: Hand Grip Strength Test, exerted pressure in lbs; 10-m-WT: 10-Meter Walk Test; ns: not statistically significant at the 5% level.

DISCUSSION

Post-stroke recovery is a complex and multidimensional process evolving for several months to years after the event; a multiplicity of neurobiological and psychological factors, as well as the access to specialized medical care, converge to determine a patient's likelihood of survival and the persistence of severe disability. Therefore, there is a compelling medical need for new treatment strategies to reverse the neurobiological damage caused by stroke as well as to restore motor, cognitive, and behavioral functioning. The application of a fit-for-purpose clinical model of stroke recovery, including a clear definition of treatment benefits, identification of relevant domains, methods of evaluation, and derived outcome measures is a key element in clinical study design. In early phase trials exploring therapeutic potential, the identification of the most appropriate

evaluation strategy in terms of clinical validity, utility, and cost-effectiveness is of paramount importance. Specifically, a clear-cut definition of treatment benefits, of their effect size, and clinical significance is essential for predicting satisfactory treatment response (or futility) in larger late-phase clinical development trials. A multinational, multicenter, randomized, double-blind, placebo-controlled study in acute ischemic stroke to evaluate the therapeutic potential in stroke recovery of PF-03049423, a PDE5 inhibitor, was carried out. Based on the pre-specified futility criteria assessed at an interim analysis, the Sponsor made the decision to immediately stop the study. Overall, no statistically significant benefit was observed for PF-03049423 relative to placebo at any time throughout the treatment period, as indicated by efficacy outcome measures derived from mRS, NIHSS BI, B&BT, HGST,

10-m-WT, RBANS Coding and Naming Sub Tests, LCT, and RMT.

As the logistic regression analysis on the primary endpoint indicated important effects for the covariates geographical region, gender, and baseline NIHSS total score (although with no evidence of treatment effects in these sub-groups), an informal *post-hoc* exploratory analysis was performed. In general, low mRS response rates were observed for subjects in America and for subjects with a moderate-to-severe baseline NIHSS (total score: 16-20 inclusive) [7].

As one of the secondary methodological objectives of our trial was to generate data to extend the validation of cognitive and motor performance-based measures as outcome endpoints in stroke recovery trials, a number of additional *post-hoc* analyses were carried on the cohort of 137 subjects, combining those randomized to

TABLE 6. Summary of correlations between measures of cognitive and motor function, and NIHSS at Day 7 [n, Pearson's r, (p-value)]

MEASURES	NIHSS	RMT	RBANS Coding	RBANS Naming	LCT/H	B&BT paretic	B&BT non-paretic	HGST paretic	HGST non-paretic	10-m-WT
	132									
NIHSS	1.0	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
	n/a									
	117	117								
RMT	-0.00	1.0	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
	(ns)	n/a								
	93	93	93							
RBANS Coding	-0.33	-0.05	1.0	n/a	n/a	n/a	n/a	n/a	n/a	n/a
	(-0.0011)	(ns)	n/a							
	113	113	92	113						
RBANS Naming	-0.13	0.08	0.32	1.0	n/a	n/a	n/a	n/a	n/a	n/a
	(ns)	(ns)	(.0017)	n/a						
	111	107	91	104	111					
LCT/H	-0.39	0.13	0.30	0.07	1.0	n/a	n/a	n/a	n/a	n/a
	(<0.0001)	(ns)	(0.0044)	(ns)	n/a					
	73	70	64	69	71	73				
B&BT paretic	-0.62	0.11	0.18	0.15	0.32	1.0	n/a	n/a	n/a	n/a
	(<0.0001)	(ns)	(ns)	(ns)	(0.0059)	n/a				
	123	116	93	112	111	73	123			
B&BT non-paretic	-0.45	0.20	0.33	0.03	0.55	0.69	1.0	n/a	n/a	n/a
	(<0.0001)	(0.0295)	(0.0011)	(ns)	(<0.0001)	(<0.0001)	n/a			
	78	74	67	73	75	73	77	78		
HGST paretic	-0.58	0.14	0.23	0.13	0.28	0.67	0.42	1.0	n/a	n/a
	(<0.0001)	(ns)	(ns)	(ns)	(0.0163)	(<0.0001)	(0.0001)	n/a		
	126	117	93	113	110	73	122	78	126	
HGST non-paretic	-0.35	0.15	0.34	0.18	0.27	0.38	0.42	0.78	1.0	n/a
	(<0.0001)	(ns)	(0.0007)	(ns)	(0.0050)	(0.0009)	(<0.0001)	(<0.0001)	n/a	
	64	63	59	62	62	57	64	59	64	64
10-m-WT	-0.27	-0.04	0.24	0.13	0.06	0.45	0.24	0.46	0.24	1.0
	(0.0288)	(ns)	(ns)	(ns)	(ns)	(0.0005)	(ns)	(0.0002)	(ns)	n/a

n/a: not applicable; n: number of subjects used in the analysis; RBANS: Repeatable Battery for the Assessment of Neuropsychological Status; RMT: Recognition Memory Test; LCT/H: Line Cancellation Test, Hemi-Inattention; B&BT: Box and Blocks Test; HGST: Hand Grip Strength Test, exerted pressure in lbs; 10-m-WT: 10-Meter Walk Test; ns: not statistically significant at the 0.05 level.

PF-03049423 or placebo. We found that post-stroke cognitive recovery is not homogeneous across different cognitive domains. Most (but not all) of the cognitive measures consistently showed a pattern of improved performance following stroke onset. The pattern was clearly time-dependant for measures derived from RBANS Naming and Coding Subtests as well as for the measure of hemi-inattention on the LCT. Our results support the validity of these group measures in the evaluation of recovery from acute ischemic stroke. In contrast, the memory measure of delayed recognition did not show evidence of consistent post-stroke recovery. The lack of correlation between RMT and the standard efficacy endpoints, at least in part due to absence of any improvement in recognition memory, suggests there may be elements of memory functioning not covered by mRS. These specific components of cognitive recovery may be clinically relevant and have potential as an endpoint in future trials aiming to improve certain aspects of cognition. We found RBANS Coding particularly meaningful. RBANS Coding essentially provides a general measure of sustained attention and working memory; this component of cognitive impairment significantly contributes to the overall activity limitation, as measured by mRS. Analysis indicated that the Coding Subtest was influenced by factors such as age and geographic region (implying nonspecific factors potentially relating to the healthcare delivery system). RBANS Coding showed weak correlations with B&BT non-paretic and HGST non-paretic. This finding is not unexpected as the successful completion of RBANS Coding requires elements of dexterity and sustained muscular strength in addition to preserved attention and working memory; similarly, the two tests of motor function require automatic and deliberate recruitment and deployment of cognitive resources to support sustained attention to satisfy

task demands. RBANS Coding informs on a domain of cognitive impairment of particular clinical relevance. In fact, altered attention, information processing, and working memory are among the defining features of the neuropsychological profile of vascular cognitive impairment.¹⁶ Due to their clinical relevance in the model of stroke recovery, the application of other cognitive measures warrants more scrutiny. Specifically, the inability to detect improvement in the Recognition Memory Test implies that improved methods of memory evaluation, not limited to delayed recognition paradigms, should be further investigated. Tests based on tasks demanding sustained cognitive effort and possibly allowing for the evaluation of recovery of verbal abilities should be considered; the test should inform on the modality and efficiency of cognitive processing as well as the organizational strategies used for information retrieval. In this respect, cognitive tests based on the verbal free recall task with repeated sessions of acquisition and retrieval¹⁷ may be a suitable and cost-effective strategy of evaluation in clinical trials. Clinical methods to evaluate upper extremity function and gait (B&BT, HGST and 10-m-WT) are objective measures and therefore may be less prone to investigator bias. Measures derived from these methods reflect the efficiency of a wide range of underlying sensory-motor processes and systems that are particularly vulnerable to and dependant on the characteristics of cortical lesion or dysfunction, and are therefore of primary clinical relevance for stroke assessment. Each of the upper extremity function and gait function measures consistently showed a pattern of improved motor performance following stroke. Each pattern was clearly time dependant. Therefore, these groups of modality-specific measures of upper extremity strength, upper extremity dexterity, and velocity gait confirmed their validity in evaluating post-stroke

recovery. Furthermore, our results show these outcome measures are differently influenced by diverse factors, such as age, nonspecific factors relating to site distribution by geographic region (implying potential differences in healthcare delivery systems), t-PA after-stroke use, or the level of neurological impairment at stroke onset (NIHSS total score).

We demonstrated how modality-specific measures of motor and cognitive functions can be successfully applied to drug stroke clinical trials in subjects suffering with stroke of moderate severity (i.e., NIHSS score at baseline of 6–20). The first question to address was whether the proposed behavioral tests consistently correlate with traditional post-stroke recovery measures up to three months. Modified Rankin Scale, BI, and NIHSS provided convergent measures of overall disability and functional impairment in subjects with acute ischemic stroke resulting in mild-to-moderate neurological impairment. These findings were expected and largely consistent with currently published data.^{18,19} Interestingly, the B&BT, HGST, 10-m-WT, and RBANS Coding tests measures were found to be weakly or moderately correlated with clinician-based assessments (e.g., mRS). The strength of the correlations with motor endpoints highlights how much measures of impairment of the upper limb function and impairment of gait (10-m-WT) partially overlap, but do not coincide, with the description of overall activity limitation in this cohort of patients, as measured by mRS. The MMRM analysis demonstrated that cognitive and motor assessments were differently influenced compared to the clinician-based assessments by factors at baseline, such as t-PA usage, geographic region, and age, indicating that they are distinct. Overall, performance-based measures expand the scope and contribute to a more complete assessment of important elements of

stroke functional recovery. Did the proposed motor and cognitive measures show a potential for superior sensitivity to detect treatment effects? Our results show that RMT demonstrated statistically significant differences in favor of the placebo group; nevertheless, the clinical relevance of this finding remains undetermined. An in-depth evaluation of responsiveness in terms of ability to detect a clinically meaningful change when it has occurred (sensitivity to change) and of the stability in subjects who have not changed (specificity to change) is indeed needed not only for RMT but for the other performance-based motor and cognitive measures adopted in this trial. Secondly, do behavioral testing measures provide additional knowledge value to the clinical trial design? In stroke trials, mRS is the primary instrument to cost-effectively provide synthetic measures that inform on the therapeutic potential of a new intervention and determines sample size. The mRS score heavily depends on motor impairment and limitation of mobility. Our data show that measures derived from behavioral methods add knowledge to post-stroke evaluation by tapping clinical domains not specifically explored by mRS, such as cognitive impairment. Notably, the RMT did not show a correlation with mRS, and, in fact, did not show improvement over time. This observation suggests that there may be elements of cognitive impairment not detected by the mRS, which may be clinically relevant to evaluate the efficacy and safety of treatments as well as to represent an endpoint in future trials aiming to improve certain aspects of cognition.

Under the assumption that there was no treatment difference between the drug and placebo groups, *post-hoc* analyses were carried out by combining the two groups in order to increase the power of the tests undertaken and the precision of the estimates obtained. We acknowledge that it could be argued that it is most

appropriate to only include the placebo group to exclude the possibility of “treatment bias.” However, based on the results of the pre-planned statistical analyses, we believe any bias due to including the treatment arm will be small, and thus combining the groups to increase the statistical power/precision for the *post-hoc* analyses is merited. The data and information generated from this trial should be helpful for prospective researchers in planning future studies. However, due to the limited sample size, pooling of data from subjects on active and placebo, and the large number of analyses and comparisons of exploratory nature, the results and conclusion drawn from this study should be interpreted with caution. Furthermore, this study selected a defined clinical sub-group of stroke patients, and the relevance of these findings to broader populations of stroke patients needs verification.

Finally, two other limitations should be noted. First, according to eligibility criteria, study subjects were required to participate in a rehabilitation program with an expectation that each subject would receive the full course of post-stroke rehabilitation as per standard of care (i.e., physical, occupational, speech, and cognitive therapy as indicated) and to be available for subsequent follow-up visits. However, the study was not designed to evaluate the effect of rehabilitation on treatment effect even though the basic information (i.e., type, duration, start/end date) on the rehabilitative treatment delivered to the study subject was documented. Second, the study design did not specifically address the effect of standard of care delivered by site on treatment outcomes. Trial designers face the challenge of reaching a consensus on the definition of standard of care in acute stroke trials, as this element of methodology clearly is key to increasing the efficiency of measuring treatment effects. There is still work to be done within the

scientific community to satisfactory resolve this methodological issue.

CONCLUSION

We demonstrated that the cognitive and motor assessments included in our Phase II study add knowledge value to the evaluation of the therapeutic potential of new treatments for stroke recovery. Psychometric measures of cognitive function as well as behavioral measures of dexterity and gait may contribute to a deeper understanding of treatment effects and, sequentially, increase the reliability of the process of investigational drug benefit/risk evaluation in clinical development. Although the mRS is currently the gold standard disability measure recognized by regulatory authorities, including the FDA, by assessing the specific motor and cognitive deficits it is possible to measure the effect of novel therapies on the specific deficits expressed by individual patients, and this may thereby reveal treatment effects that would otherwise not be recognized.

REFERENCES

1. Go AS, Mozaffarian D, Roger VL, et al. on behalf of the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Heart disease and stroke statistics—2014 update: a report from the American Heart Association. *Circulation*. 2014;129:e28–e292.
2. Jauch EC, Saver JL, Adams PA, et al. Guidelines for the early management of patients with acute ischemic stroke. *Stroke*. 2013; 44:870–947.
3. Rankin L. Cerebral vascular accidents in subjects over the age of 60. II. Prognosis. *Scott Med J*. 1957;2:200–215.
4. Mahoney MI, Barthel DW. Functional evaluation: the Barthel Index. *Md State Med J*. 1965; 14:61–65.
5. Lyden P, Raman R, Liu L, et al. NIHSS training and certification using a new digital video disk is reliable. *Stroke*. 2005;36:2446–449.
6. Cramer SC, Koroshetz WJ, Finklestein SP. The case for modality-specific outcome measures

- in clinical trials of stroke recovery-promoting agents. *Stroke*. 2007;38(4):1393–1395.
7. Di Cesare F, Mancuso J, Woodward P, et al. Phosphodiesterase-5 inhibitor PF-03049423 effect on stroke recovery. A double-blind, placebo controlled trial. *J Stroke Cerebrovascular Dis*. 2015 Dec 28. [Epub ahead of print]
 8. Menniti FS, Ren J, Coskran TM, et al. Phosphodiesterase 5A inhibitors improve functional recovery after stroke in rats: optimized dosing regimen with implications for mechanism. *J Pharmacol Exp Ther*. 2009;331:842–850.
 9. Kroenke K, Strine TW, Spitzer RL, et al. The PHQ-8 as a measure of current depression in the general population. *J Affect Disord*. 2009;114:163–73.
 10. Wilde MC. The validity of the repeatable battery of neuropsychological status in acute stroke. *Clin Neuropsychol*. 2006;20(4):702–715.
 11. Larson E, Kirschner K, Bode R, et al. Construct and predictive validity of the repeatable battery for the assessment of neuropsychological status in the evaluation of stroke patients. *J Clin Exp Neuropsychol*. 2005; 27(1):16–32.
 12. Mathiowetz V, Volland G, Kashman N, et al. Adult norms for the Box and Block test for manual dexterity. *Am J Occup Ther*. 1985; 39:386–91.
 13. Russell EW, Starkey RI. Halstead. *Russell Neuropsychological Evaluation System*. Los Angeles, CA: Western Psychological Services; 1993.
 14. Perera S, Mody SH, Woodman RC, Studenski SA. Meaningful change and responsiveness in common physical performance measures in older adults. *J Am Geriatr Soc*. 2006;54:743–749.
 15. Dancey CP, Reidy J. *Statistics Without Maths for Psychology: Using SPSS for Windows*. New York, NY: Prentice Hall; 2004.
 16. Sachdev PS, Brodaty H, Valenzuela MJ, et al. The neuropsychological profile of vascular cognitive impairment in stroke and TIA patients. *Neurology*. 2004;23;62(6):912–919.
 17. Di Cesare F, D'Ilario D, Fioravanti M. Differential characteristics of the aging process and the vascular cognitive impairment in the organization of memory retrieval. *J Neurol Sci*. 2012;322(1–2):148–51.
 18. Kwon S, Hartzema AG, Duncan PW, et al. Disability measures in stroke. relationship among the Barthel Index, the Functional Independence Measure, and the Modified Rankin Scale. *Stroke*. 2004;35:918–923.
 19. Goldie FC, Fulton RL, Frank B, et al. Interdependence of stroke outcome scales: reliable estimates from the Virtual International Stroke Trials Archive (VISTA). *Int J Stroke*. 2014;9(3):328–332. ■