Supplementary Online Content

Magaziner J, Mangione KK, Orwig D, et al. Effect of a multicomponent home-based physical therapy intervention on community ambulation after hip fracture in older adults: the CAP randomized clinical trial [published September 10, 2019]. *JAMA*. doi:10.1001/jama.2019.12964

This supplementary material has been provided by the authors to give readers additional information about their work.

eAppendix. Supplementary Analyses

This supplement provides results of analyses outlined in the Detailed Statistical Analysis Plan that is appended to the CAP study record in Clinicaltrials.gov (https://clinicaltrials.gov/ProvidedDocs/04/NCT01783704/SAP_001.pdf) and as an online supplement to the manuscript entitled "Effect of a multicomponent home-based physical therapy intervention on community ambulation after hip fracture in older adults: the CAP randomized clinical trial". This supplement also contains the results of post-hoc sensitivity analyses (sections 1.6 to 1.10).

All p-values and confidence intervals reported in this document are two-sided, with the exception of the p-values and confidence intervals for the primary outcome (community ambulation 16 weeks after randomization) which are reported as one-sided consistent with our pre-specified one-sided hypothesis test for this outcome.

1.1 Analysis of primary outcome

1.1.1 Primary analysis

Of the 210 patients randomized in the study, 23 (14 in the Training group and nine in the active control group) were not assessed for the primary outcome (community ambulation at the 16-week assessment) and were submitted for adjudication. Of these, 10 (five in Training and five in active control) were adjudicated to be non-community ambulators; the remaining 13 had an indeterminate outcome. Adding the adjudicated outcomes to those obtained by assessment with the SMWT, we obtained the following results.

In Training, 22/96 (23%) were community ambulators.
In active control, 18/101 (18%) were community ambulators.
Z-score=0.89. This does not exceed the threshold for null hypothesis rejection.
One-sided p-value=0.19
Estimated difference (Training minus control) (one-sided 97.5% CI): 5.1% (−∞, 16.3%)

1.1.2 Revised definition of primary outcome and adjustment for imbalances between groups

The analysis plan calls for a secondary analysis of the community ambulation outcome excluding (i.e., considering indeterminate) those who were adjudicated as non-ambulators solely due to self-report or proxy report. There were four such participants (two in each group). Changing these participants' outcome to indeterminate and excluding them from analysis resulted in the following findings:

In Training, 22/94 (23%) were community ambulators.
In active control, 18/99 (18%) were community ambulators.
Z-score=0.89. This does not exceed the threshold for null hypothesis rejection
One-sided p-value=0.19
Estimated difference (Training minus control) (one-sided 97.5% CI): 5.2% (−∞, 16.7%)

Weighted analysis adjusting for imbalances between the groups with respect to important variables

The analysis plan also calls for a weighted analysis of the alternative definition of the primary outcome (i.e., considering as indeterminate the outcomes of those who were adjudicated as non-ambulators solely due to self-report or proxy report) to address the possibility that the groups were imbalanced with respect to important predictors of ambulation. To choose the predictors, we used the Beach-Meier approach which involved multiplying Z-scores relating covariates to group assignment

by Z-scores relating covariates to community ambulation and ranking the variables based on the value of the product.[1] Based on the fact that there are 40 community ambulators in the study, and assuming that no more than one variable per 10 events should be included, the plan calls for using the top four ranked variables in an adjusted analysis.

The top four variables (all measured at baseline) were gait speed to walk 50 feet, gait speed to walk four meters, Modified Mini-Mental State (3MS) examination score, and Modified Physical Performance Test (mPPT) score. Adjusting for these variables, the following results were obtained:

In Training, proportion of community ambulators: 21.8%
In active control, proportion of community ambulators: 19.8%
One-sided p-value=0.36
Estimated difference (Training minus control): 2.1% (one-sided 97.5% CI $-\infty$, 13.6%)

Sensitivity to bias due to missing data

The findings described above are unbiased if those who had indeterminate outcomes are similar to those who did not have indeterminate outcomes with respect to community ambulation (i.e., the outcomes are missing at random). In this section, we assess the degree to which our findings would be affected by various degrees of departure from the assumption that the outcomes are missing at random. To do this, we quantified the degree of departure from the assumption of missing at random with a parameter called the response probability ratio (RPR).[2] RPR is the probability of a non-missing outcome among those who were community ambulators divided by the probability of a non-missing outcome among those who were not community ambulators. If RPR=1 then the data are missing at random. The further RPR is from 1, the greater the departure from the missing at random assumption. Though it is impossible to fully estimate the RPR from the data without knowing the community ambulation status of those with missing data, there is a minimum RPR and a maximum RPR consistent with the data. The minimum RPR corresponds to the RPR that would have been observed if every participant who was indeterminate were a community ambulator. Similarly, the maximum RPR consistent with the data would be that which would have been observed if every participant who was indeterminate was not a community ambulator. Table S1 below shows estimated differences between the groups in 16-week community ambulation given various assumptions about RPR based on the primary definition of community ambulation.

eTable 1. Estimated differences in community ambulation and p-values given various assumptions about RPR in each group

Assumptions	Difference (Training minus control) (97.5% CI)	One-sided p-value
$RPR_{TRAINING}=RPR_{CONTROL}=1.0$	5.1% ($-\infty$, 16.3%)	0.19
$RPR_{TRAINING}=RPR_{CONTROL}=0.82$ (min RPR consistent with data)	5.7% ($-\infty$, 18.3%)	0.19
$RPR_{TRAINING}=RPR_{CONTROL}=1.05$ (max RPR consistent with data)	5.0% ($-\infty$, 15.9%)	0.19
$RPR_{TRAINING}=0.71$, $RPR_{CONTROL}=1.05$	12.4% ($-\infty$, 24.6%)	0.024
$RPR_{TRAINING}=1.12$, $RPR_{CONTROL}=0.82$	0.0% ($-\infty$, 11.5%)	0.50

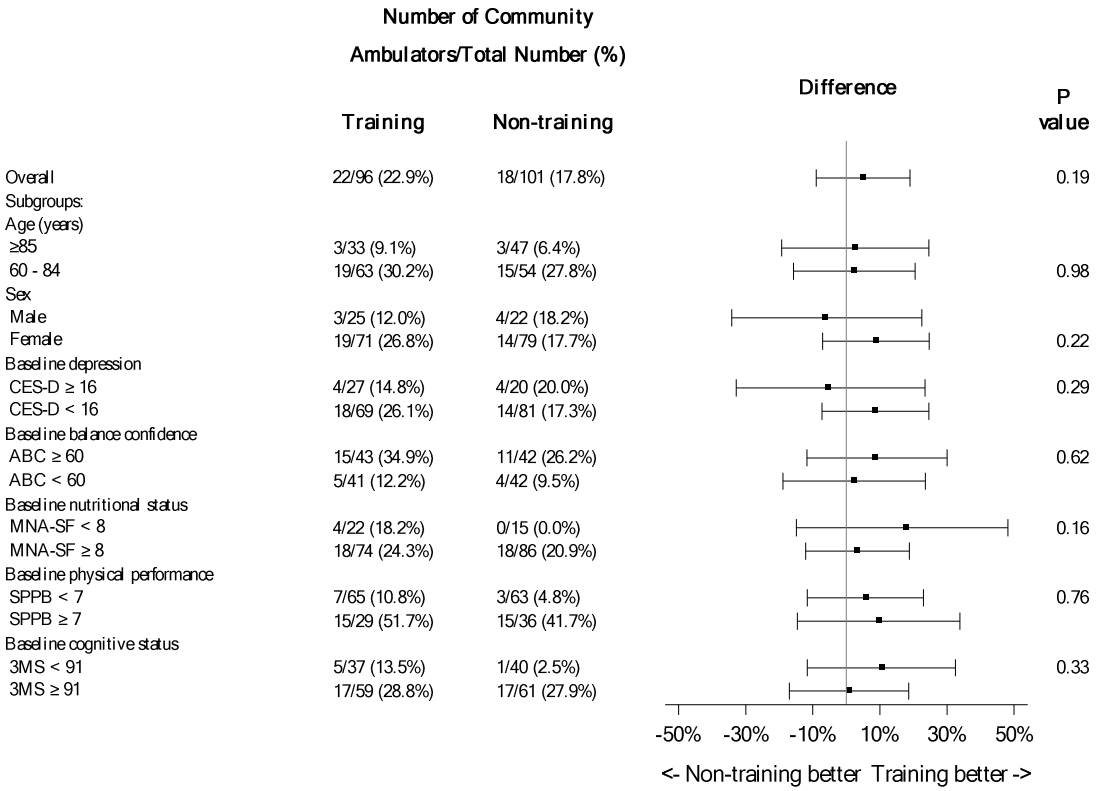
Note that when RPR is assumed to be the same in each group, the results of the analysis are essentially unchanged. Under the extreme assumption that in the Training group, all those who were indeterminate were truly community ambulators, whereas in the active control group none of those who were indeterminate were truly community ambulators, then there is some evidence of a better outcome among those in the Training group (p=0.024).

1.1.3 Assessment of effect modification by patient characteristics, clinical site, and protocol version

Analyses of effect modification by patient characteristics

A forest plot was generated for the difference between Training and active control with respect to the proportion of community ambulators at 16-week follow-up, overall and within certain subgroups, including exact 95% confidence intervals. A p-value for the overall Training-control difference and p-values for the interaction between the Training-control difference and the subgroup variables were calculated using a chi-squared test within a generalized linear model. Due to a zero value, the p-value for the MNA-SF interaction was calculated using a two-sample Z-test instead. The results are shown in Figure S1.

eFigure 1. Difference in proportion of community ambulators at 16-week follow-up (Training minus control), by subgroup



Abbreviations: CES-D=Center for Epidemiologic Studies-Depression, ABC=Activities-Specific Balance Confidence, MNA-SF=Mini Nutritional Assessment-Short Form, SPPB=Short Physical Performance Battery, 3MS=Modified Mini-Mental State

Note: The figure displays the differences in the proportion of community ambulators at 16-week follow-up (Training minus control) (black squares), 95% confidence intervals (horizontal lines), and p-values for the interaction between the Training-control difference and the subgroup variables. All subgroup analyses were prespecified. For ABC, the median score for both groups combined was used as the cutoff to define subgroups.

Analyses of differences among sites with respect to treatment effect

According to the detailed analysis plan, “Study site will be investigated as a modifier of the effect of the intervention by testing a site-by-intervention interaction term on the difference scale based on a binary regression model with an identity link. If there is evidence that study site is an effect modifier (i.e., $p<0.1$ for the interaction term), we will report site-specific treatment effects.” Performing that analysis resulted in the findings in Table S2 (below).

eTable 2. Primary treatment comparison, by clinical site

Clinical site	Treatment group	Proportion (%) achieving community ambulation	Difference (Training minus control) (one-sided 97.5% CI)	One-sided P-value	P-value for interaction by site
1	Training	4/32 (12.5%)	-14.8% (–∞, 4.3%)	0.07	0.051
	Control	9/33 (27.3%)			
2	Training	6/31 (19.4%)	10.8% (–∞, 27.5%)	0.10	
	Control	3/35 (8.6%)			
3	Training	12/33 (36.4%)	18.2% (–∞, 39.2%)	0.049	
	Control	6/33 (18.2%)			

A series of additional analyses was done to determine whether the differences in the Training-control comparison by clinical site might have been caused by data or analysis errors and, if not, to explore possible explanations (such as differences in the way the interventions were provided) for the observed site by group interaction. These analyses revealed no evidence of data or analysis errors or of systematic differences among the sites that could explain the differences in treatment effect. We concluded that chance was the most likely explanation of the site by group interaction.

Analysis of differences in protocol versions with respect to treatment effect

In the fall of 2014, a decision was made to modify the frequency of intervention visits during the first eight weeks of study participation from three per week to two per week. Analyzing the Training-control comparison separately in the two periods resulted in the findings in Table S3 (below).

eTable 3. Primary treatment comparison based on whether the protocol required two or three intervention visits per week during the first eight weeks

Protocol version	Treatment group	Proportion (%) achieving community ambulation	Difference (Training minus control) (one-sided 97.5% CI)	One-sided P-value	P-value for interaction by protocol version
Early version (3 visits/week)	Training	8/37 (21.6%)	6.5% (–∞, 24.5%)	0.24	0.87
	Control	5/33 (15.2%)			
Later version (2 visits/week)	Training	14/59 (23.7%)	4.6% (–∞,18.9%)	0.26	
	Control	13/68 (19.1%)			

1.2 Secondary objectives

1.2.1 Delayed and sustained effects on primary outcome

Analysis of 40-week community ambulation status

Of the 210 patients randomized in the study, 58 (29 in each group) were not assessed for the 40-week SMWT because of the protocol change in version 10. Of the remaining 152 patients, community ambulation was measured for 118 using the SMWT and 34 were submitted for adjudication (17 in each group). Of the 34 adjudicated, 16 were determined not to be community ambulators (10 and 6 in Training and active control, respectively). Adding these 16 to the 118, we have community ambulation results on 134 patients. This resulted in the following findings:

In Training, 15/69 (22%) were community ambulators.
In active control, 14/65 (22%) were community ambulators.
Z-score=0.03
P-value=0.98
Estimated difference (Training minus control) (95% CI): 0.2% (-13.8%, 14.2)

Estimating delayed impact and sustainability using a 3x3 table

Tables S4a, S4b and S4c (below) show 16- and 40-week community ambulation results for all participants and separately for each treatment group.

eTable 4a. Community ambulation at 40 weeks in strata defined by community ambulation at 16 weeks (all participants)

		40-week community ambulation			
		No	**Yes**	**Indeterminate**	**Total**
16-week community ambulation	No	97 (87%)	5 (4%)	10 (9%)	112
	Yes	7 (23%)	24 (77%)	0 (0%)	31
	Indeterminate	1 (11%)	0 (0%)	8 (89%)	9

eTable 4b. Community ambulation at 40 weeks in strata defined by community ambulation at 16 weeks (Training participants)

		40-week community ambulation			
		No	**Yes**	**Indeterminate**	**Total**
16-week community ambulation	No	48 (92%)	2 (4%)	2 (4%)	52
	Yes	5 (28%)	13 (72%)	0 (0%)	18
	Indeterminate	1 (17%)	0 (0%)	5 (83%)	6

eTable 4c. Community ambulation at 40 weeks in strata defined by community ambulation at 16 weeks (active control participants)

		40-week community ambulation			
		No	**Yes**	**Indeterminate**	**Total**
16-week community ambulation	No	49 (82%)	3 (5%)	8 (13%)	60
	Yes	2 (15%)	11 (85%)	0 (0%)	13
	Indeterminate	0 (0%)	0 (0%)	3 (100%)	3

In the pooled analysis (Table S4a), among the 102 participants who were not community ambulators at 16 weeks and who had a determination at 40 weeks, only 5 (5%) were community ambulators at 40 weeks. Thus, it appears that delayed response is rare. In addition, there was some lack of sustainability as among the 31 participants who were community ambulators at 16 weeks, 7 (23%) were not community ambulators at 40 weeks.

Of those in the Training group who were not community ambulators at 16 weeks and who were not indeterminate at 40 weeks, 2/50 (4%) exhibited a delayed response and became community ambulators. In the active control group, the comparable proportion was 3/52 (6%). This difference between the groups in delayed response was not statistically significant (p=1.0 by a two-sided Fisher's Exact Test).

Of those in the Training group who were community ambulators at 16 weeks and who were not indeterminate at 40 weeks, 13/18 (72%) sustained their community ambulation. In the active control group, the comparable proportion was 11/13 (85%). This difference between the groups in sustained response was not statistically significant (p=0.67 by a two-sided Fisher's Exact Test).

There was no difference in the association between 16-week and 40-week community ambulation in the Training group (Table S4b) vs. the active control group (Table S4c) based on a Breslow-Day test (p=0.78)

Estimating 16-week and 40-week differences between groups using a longitudinal regression model

The longitudinal regression model specified in the detailed analysis plan results in estimates of community ambulation in each group at each time point. The model can be used to test the differences between the groups at each time point and differences between the time points in each group. Table S5 below gives the estimates for each of those differences.

eTable 5. Differences in community ambulation for various comparisons based on the longitudinal regression model

Contrast	Difference (Training minus control) in the proportion of community ambulators (95% CI)	P-value
Training at 16 weeks vs. control at 16 weeks	5.2% (-13.5%, 24.0%)	0.58
Training at 40 weeks vs. control at 40 weeks	2.3% (-20.5%, 25.1%)	0.84
Training at 16 weeks vs. Training at 40 weeks	0.7% (-16.7%, 18.2%)	0.93
Control at 16 weeks vs. control at 40 weeks	-2.1% (-19.9%, 15.6%)	0.81

In addition, as indicated in the protocol, we can test the hypothesis of no difference between the 16-week and 40-week time points with respect to the between-group difference in the proportion of community ambulators. This is the difference between the difference estimates provided in the top two rows of Table S5. This difference of differences is not significantly different from 0 (p=0.82).

Note that the longitudinal analysis does not result in direct estimates of sustainability or delayed response. This is because a difference between 16 and 40 weeks in the proportion of community ambulators might be due to delayed response, lack of sustainability, or both.

1.2.2 Analysis of secondary and tertiary quantitative outcomes

The impact of Training on quantitative measures over baseline, 16 weeks and 40 weeks was assessed using longitudinal regression models fit by restricted maximum likelihood. The models included an unstructured variance/covariance matrix to account for within-person correlation. Tables S6a and S6b provide the results of these analyses. One model was fit for each outcome. (The results for community ambulation are included for completeness.) Note that, because this model assumes that due to randomization, the expected outcome at baseline is equivalent in the two groups, the estimated difference in expected value at 16 weeks can also be interpreted as the difference between the groups with respect to the mean change between baseline and 16 weeks.

eTable 6a. Study outcomes at 16 weeks

	Training (N=105)	Active control (N=105)		
	Number of community ambulators/ number assessed or adjudicated (%)	**Number of community ambulators/ number assessed or adjudicated (%)**	**Difference[1,2] (97.5% CI)**	**P-value**
Primary Outcome: Community ambulation	22/96 (23%)	18/101 (18%)	5.1% (–∞, 16.3%)	0.19
	Mean (SD) (number assessed)	**Mean (SD) (number assessed)**	**Difference[1,3] (95% CI)**	
Other Outcomes: Distance walked in 6 minutes (m)	242.4 (83.7) (91)	233.1 (83.1) (96)	8.7 (-9.5, 26.8)	0.35
Short Physical Performance Battery score[4]	6.7 (2.7) (91)	7.1 (3.0) (95)	-0.3 (-0.9, 0.3)	0.34
Modified Physical Performance Test score	20.6 (7.4) (91)	20.9 (7.8) (97)	-0.8 (-2.1, 0.6)	0.27
NHATS balance score (s)	26.1 (12.3) (91)	26.6 (12.0) (97)	-0.9 (-3.8, 1.9)	0.52
Gait speed, 50-ft fast walk (m/s)	0.81 (0.25) (89)	0.81 (0.27) (93)	-0.02 (-0.08, 0.04)	0.47
Gait speed, 4-m usual walk (m/s)	0.73 (0.23) (89)	0.74 (0.22) (95)	-0.02 (-0.07, 0.03)	0.43
Quadriceps muscle strength (lb of force per lb of body weight)	0.28 (0.11) (67)	0.29 (0.12) (67)	-0.01 (-0.04, 0.01)	0.19
	Number with increase/number assessed (%)	**Number with increase/number assessed (%)**	**Difference[1,2] (95% CI)**	
Increase of ≥50 m in distance walked in 6 minutes relative to baseline[4]	44/91 (48%)	40/96 (42%)	6.7% (-7.6%, 20.9%)	0.36

eTable 6b. Study outcomes at 40 weeks

	Training (N=76[5])	Active control (N=76[5])		
	Number of community ambulators/ number assessed or adjudicated (%)	**Number of community ambulators/ number assessed or adjudicated (%)**	**Difference[1,2] (95% CI)**	**P-value**
Community ambulation	15/69 (22%)	14/65 (22%)	0.2% (-13.8%, 14.2%)	0.98
	Mean (SD) (number assessed)	**Mean (SD) (number assessed)**	**Difference[1,3] (95% CI)**	
Distance walked in 6 minutes (m)	239.2 (93.7) (60)	241.6 (95.6) (59)	-6.3 (-30.0, 17.3)	0.60
Short Physical Performance Battery score[4]	6.7 (3.1) (63)	6.8 (3.1) (60)	-0.0 (-0.8, 0.7)	0.89
Modified Physical Performance Test score	19.7 (8.7) (64)	19.8 (8.7) (61)	-0.1 (-2.1, 1.9)	0.90
NHATS balance score (s)	23.8 (12.7) (65)	25.2 (13.6) (62)	-1.6 (-4.8, 1.7)	0.34
Gait speed, 50-ft fast walk (m/s)	0.82 (0.27) (62)	0.80 (0.29) (59)	-0.01 (-0.08, 0.06)	0.82
Gait speed, 4-m usual walk (m/s)	0.77 (0.28) (62)	0.74 (0.26) (59)	0.02 (-0.04, 0.09)	0.48
Quadriceps muscle strength (lb of force per lb of body weight)	0.27 (0.10) (55)	0.30 (0.14) (57)	-0.03 (-0.06, -0.004)	0.026
	Number with increase/number assessed (%)	**Number with increase/number assessed (%)**	**Difference[1,2] (95% CI)**	
Increase of ≥50 m in distance walked in 6 minutes relative to baseline[4]	30/60 (50%)	29/59 (49%)	0.8% (-17.2%, 18.9%)	0.93

[1] All differences are Training minus control. The p-value for the primary outcome (community ambulation at 16 weeks) is reported as one-sided. All others are reported as two-sided.
[2] Differences, 95% confidence intervals, and p-values for binary variables were based on differences in proportions and large-sample methods (i.e., chi-square tests).
[3] Differences, 95% confidence intervals, and p-values for quantitative variables were based on longitudinal regression models allowing different variances and covariances at and between each time point, and fit by maximum likelihood.
[4] Tertiary outcomes
[5] This sample size reflects the number of participants who were randomized under the original protocol which involved follow-up to 40 weeks.

The same models that produced the results shown in Tables S6a and S6b were also used to generate estimates of changes between baseline and follow-up separately in each treatment group. The results of those analyses are shown in Tables S6c and S6d:

eTable 6c. Estimated changes between baseline and 16 weeks in secondary and tertiary quantitative outcomes, by treatment

	Estimated change between baseline and 16-week measures			
	Training		Active control	
	Estimate (95% CI)	P-value	Estimate (95% CI)	P-value
Distance walked in 6 minutes (m)	54.1 (41.1, 67.1)	<0.0001	45.5 (32.8, 58.1)	<0.0001
Short Physical Performance Battery score[1]	1.18 (0.76, 1.61)	<0.0001	1.47 (1.06, 1.89)	<0.0001
Modified Physical Performance Test score	2.41 (1.44, 3.38)	<0.0001	3.16 (2.22, 4.10)	<0.0001
NHATS balance score (s)	0.98 (-1.16, 3.12)	0.37	1.93 (-0.16, 4.01)	0.070
Gait speed, 50-ft fast walk (m/s)	0.12 (0.08, 0.16)	<0.0001	0.14 (0.10, 0.18)	<0.0001
Gait speed, 4-m usual walk (m/s)	0.12 (0.09, 0.15)	<0.0001	0.14 (0.11, 0.17)	<0.0001
Quadriceps muscle strength (lb of force per lb of body weight)	0.01 (-0.00, 0.03)	0.072	0.03 (0.01, 0.04)	0.0003

eTable 6d. Estimated changes between baseline and 40 weeks in secondary and tertiary quantitative outcomes, by treatment

	Estimated change between baseline and 40-week measures			
	Training		Active control	
	Estimate (95% CI)	P-value	Estimate (95% CI)	P-value
Distance walked in 6 minutes (m)	43.0 (26.2, 59.8)	<0.0001	49.3 (32.6, 66.1)	<0.0001
Short Physical Performance Battery score[1]	1.11 (0.60, 1.62)	<0.0001	1.16 (0.64, 1.67)	<0.0001
Modified Physical Performance Test score	1.54 (0.11, 2.96)	0.035	1.66 (0.22, 3.10)	0.024
NHATS balance score (s)	-1.01 (-3.30, 1.29)	0.39	0.57 (-1.76, 2.89)	0.63
Gait speed, 50-ft fast walk (m/s)	0.12 (0.07, 0.17)	<0.0001	0.13 (0.08, 0.18)	<0.0001
Gait speed, 4-m usual walk (m/s)	0.14 (0.10, 0.19)	<0.0001	0.12 (0.07, 0.17)	<0.0001
Quadriceps muscle strength (lb of force per lb of body weight)	0.00 (-0.02, 0.02)	0.99	0.03 (0.01, 0.05)	0.0016

[1] Tertiary outcome

It can be seen that, with few exceptions, there were significant improvements from baseline to follow-up in quantitative outcomes in each group.

1.3 Cost-effectiveness of interventions

The cost-effectiveness analyses were not performed because neither of the following conditions (pre-specified in the detailed statistical analysis plan) was met:

- There is a statistically significant difference in the primary outcome between groups, OR
- There is a statistically significant and clinically meaningful between-group difference in any of the secondary and tertiary outcomes listed below.

Clinically meaningful differences for secondary and tertiary outcomes at 16- and 40-week follow-up:

Distance walked in 6 minutes (m)	50 m[3]
Short Physical Performance Battery score*	1.0[3]
Gait speed, 50-ft fast walk (m/s)	0.10 m/s[4]
Gait speed, 4-m usual walk (m/s)	0.10m/s[3]

* Tertiary outcome

1.4 Adjusted analysis of secondary and tertiary quantitative outcomes to account for possible imbalances due to missing data

Some of the quantitative outcomes at one or more follow-up times are missing. As a secondary analysis, to correct for imbalances due to chance or differential missing data, we adjusted for covariates in our models. The adjusted model for each outcome variable included those baseline covariates ranked highest by the Beach-Meier criterion.[1] The results are shown in Table S7.

eTable 7. Comparison of treatment groups with respect to quantitative secondary and tertiary outcomes adjusting for baseline covariates selected for imbalance between the groups and influence on the outcomes[1]

Outcome variable	Baseline covariates adjusted for	16-week results		40-week results	
		Mean difference (Training minus control) (95% CI)	P-value	Mean difference (Training minus control) (95% CI)	P-value
Distance walked in six minutes (m)	Gait speed 50-ft walk, gait speed 4-m walk, mPPT, age	7.4 (-10.6, 25.4)	0.42	-6.8 (-30.4, 16.8)	0.57
SPPB[2]	Gait speed 50-ft walk, gait speed 4-m walk, mPPT, NHATS	-0.4 (-0.9, 0.2)	0.23	-0.1 (0.8, 0.6)	0.77
mPPT	Gait speed 50-ft walk, gait speed 4-m walk, age, NHATS	-0.9 (-2.2, 0.5)	0.20	-0.4 (-2.3, 1.6)	0.71
NHATS	Gait speed 50-ft walk, gait speed 4-m walk, age, mPPT	-1.1 (-3.9, 1.7)	0.45	-1.7 (-4.8, 1.5)	0.30
Gait speed 50-ft walk	Gait speed 4-m walk, mPPT, age, distance walked in six minutes	-0.01 (-0.07, 0.04)	0.65	0.00 (-0.06, 0.07)	0.99
Gait speed 4-m walk	Gait speed 50-ft walk, mPPT, age, distanced walked in six minutes	-0.02 (-0.06, 0.02)	0.30	0.02 (-0.04, 0.09)	0.48
Quadriceps strength (lb of force per lb of body weight)	Gait speed 50-ft walk, mPPT, age, gait speed 4-m walk	-0.01 (-0.04, 0.01)	0.17	-0.03 (-0.06, -0.00)	0.021

[1] Differences, 95% confidence intervals, and p-values were estimated based on a longitudinal regression model allowing different variances and covariances at each time point and between each pair of time points, including the specified covariates, and fit by maximum likelihood.
[2] Tertiary outcome

There were no significant differences between the groups with respect to these quantitative outcomes at either 16 weeks or 40 weeks, with one exception: the Training group had significantly lower quadriceps strength than the active control group at 40 weeks.

1.5 Accounting for variability introduced by differences among physical therapists

It is possible that physical therapists differed with respect to outcomes achieved. Since physical therapists were nested within treatment groups, this can introduce variability into the analysis of Training-control comparisons. This variation is not accounted for in any of the analyses above. Thus, to the extent that there is variation among physical therapists, the p-values related to the Training-control comparisons above might be lower than they should be. To account for the variation among physical therapists, we used mixed effects models (including a random effect for physical therapist) to estimate the degree of variation among physical therapists with respect to a) community ambulation, b) 16-week distance walked in six minutes, and c) change in distance walked in six minutes between baseline and 16 weeks. In each case, the estimated variation among physical therapists was 0. In other words, the variation observed among therapists was no more than would be expected if there were no true differences among therapists. These results suggest that variation among physical therapists did not increase the variation in the data and the reported p-values based on the original analysis are valid.

1.6 Accounting for variability between sites using a random site effect

1.6.1 Including a random effect of site

For the primary outcome (community ambulation at 16 weeks), in a post-hoc analysis, we fit a generalized linear mixed model for a binary outcome, including a random intercept for site, and using an identity link. For the analyses of the secondary and tertiary quantitative outcomes, we added a random effect for site to the longitudinal regression models. The results are in Tables S8a and S8b. The point estimates from these analyses are identical to those shown in Tables S6a and S6b. There are minor differences in the confidence intervals and p-values reflecting the inclusion of a random effect for site. There is no difference in the interpretation of results from the analyses with and without inclusion of the site effect.

eTable 8a. Study outcomes at 16 weeks, accounting for random differences between sites

	Training (N=105)	Active control (N=105)		
	Number of community ambulators/ number assessed or adjudicated (%)	**Number of community ambulators/ number assessed or adjudicated (%)**	**Difference[1,2] (97.5% CI)**	**P-value**
Primary Outcome: Community ambulation	22/96 (23%)	18/101 (18%)	5.0% (–∞, 16.3%)	0.19
	Mean (SD) (number assessed)	**Mean (SD) (number assessed)**	**Difference[1,3] (95% CI)**	
Other Outcomes: Distance walked in 6 minutes (m)	242.4 (83.7) (91)	233.1 (83.1) (96)	8.7 (-9.2, 26.7)	0.35
Short Physical Performance Battery score[4]	6.7 (2.7) (91)	7.1 (3.0) (95)	-0.3 (-0.9, 0.3)	0.34
Modified Physical Performance Test score	20.6 (7.4) (91)	20.9 (7.8) (97)	-0.7 (-2.1, 0.6)	0.28
NHATS balance score (s)	26.1 (12.3) (91)	26.6 (12.0) (97)	-1.0 (-3.8, 1.9)	0.51
Gait speed, 50-ft fast walk (m/s)	0.81 (0.25) (89)	0.81 (0.27) (93)	-0.02 (-0.08, 0.04)	0.46
Gait speed, 4-m usual walk (m/s)	0.73 (0.23) (89)	0.74 (0.22) (95)	-0.02 (-0.07, 0.03)	0.43
Quadriceps muscle strength (lb of force per lb of body weight)	0.28 (0.11) (67)	0.29 (0.12) (67)	-0.01 (-0.04, 0.01)	0.19
	Number with increase/number assessed (%)	**Number with increase/number assessed (%)**	**Difference[1,2] (95% CI)**	
Increase of ≥50 m in distance walked in 6 minutes relative to baseline[4]	44/91 (48%)	40/96 (42%)	6.7% (-7.6%, 20.9%)	0.36

eTable 8b. Study outcomes at 40 weeks, accounting for random differences between sites

	Training (N=76[5])	Active control (N=76[5])		
	Number of community ambulators/ number assessed or adjudicated (%)	**Number of community ambulators/ number assessed or adjudicated (%)**	**Difference[1,2] (95% CI)**	**P-value**
Community ambulation	15/69 (22%)	14/65 (22%)	0.1% (-14.0%, 14.2%)	0.98
	Mean (SD) (number assessed)	**Mean (SD) (number assessed)**	**Difference[1,3] (95% CI)**	
Distance walked in 6 minutes (m)	239.2 (93.7) (60)	241.6 (95.6) (59)	-6.5 (-30.1, 17.0)	0.59
Short Physical Performance Battery score[4]	6.7 (3.1) (63)	6.8 (3.1) (60)	-0.0 (-0.8, 0.7)	0.89
Modified Physical Performance Test score	19.7 (8.7) (64)	19.8 (8.7) (61)	-0.1 (-2.1, 1.9)	0.91
NHATS balance score (s)	23.8 (12.7) (65)	25.2 (13.6) (62)	-1.6 (-4.8, 1.6)	0.33
Gait speed, 50-ft fast walk (m/s)	0.82 (0.27) (62)	0.80 (0.29) (59)	-0.01 (-0.08, 0.06)	0.82
Gait speed, 4-m usual walk (m/s)	0.77 (0.28) (62)	0.74 (0.26) (59)	0.02 (-0.04, 0.09)	0.48
Quadriceps muscle strength (lb of force per lb of body weight)	0.27 (0.10) (55)	0.30 (0.14) (57)	-0.03 (-0.06, 0.00)	0.028
	Number with increase/number assessed (%)	**Number with increase/number assessed (%)**	**Difference[1,2] (95% CI)**	
Increase of ≥50 m in distance walked in 6 minutes relative to baseline[4]	30/60 (50%)	29/59 (49%)	0.0% (-17.8%, 18.4%)	0.97

[1] All differences are Training minus control. The p-value for the primary outcome (community ambulation at 16 weeks) is reported as one-sided. All others are reported as two-sided.
[2] Differences, 95% confidence intervals, and p-values for binary variables were based on generalized linear mixed models for a single binary variable and a random effect for site.
[3] Differences, 95% confidence intervals, and p-values for quantitative variables were based on longitudinal regression models allowing different variances and covariances at and between each time point, and fit by maximum likelihood.
[4] Tertiary outcome
[5] This sample size reflects the number of participants who were randomized under the original protocol that involved follow-up to 40 weeks.

1.6.2 Including a random effect of site on intervention impact

We estimated the variation between sites with respect to the treatment effect using a two-stage approach and restricted maximum likelihood estimation. The analysis was performed using the rma function in the R package "metafor". The test for homogeneity resulted in a Q=6.11, P=0.05 and an I^2 = 68.4%. The resulting pooled estimate of the difference in community ambulation was 4.6%, one-sided p-value=0.32, one-sided 97.5% confidence interval (−∞, 24.0%).

1.7 Analysis stratified by BMI

Because there was an imbalance between the groups with respect to BMI at baseline, we performed a post-hoc analysis stratifying on BMI at baseline. The results for the primary outcome are as follows (Table S9):

eTable 9. Primary treatment comparison, stratified by BMI at baseline

Baseline BMI	Treatment group	Proportion (%) achieving community ambulation	Difference (Training minus control) (one-sided 97.5% CI)	P-value	P-value for interaction by BMI group
<20	Training	5/12 (42%)	17% (–∞, 67%)	0.28	0.71
	Control	1/4 (25%)			
20-30	Training	12/70 (17%)	1% (–∞, 13%)	0.44	
	Control	13/80 (16%)			
301	Training	5/14 (36%)	12% (–∞, 44%)	0.23	
	Control	4/17 (24%)			

In addition, when we fit a model adjusting for baseline BMI, the estimated difference between the groups with respect to community ambulation was 3% (one-sided 97.5% CI –∞, 14%, one-sided p=0.30).

1.8 Analysis of the relationship between timing of the 16-week assessment and community ambulation status

This section shows post-hoc analyses that take into consideration the time between randomization and the primary outcome assessment.

This analysis was based on the 187 patients who performed the SMWT. The target window for assessment of the primary outcome was 112 to 126 days (16-18 weeks). The time between randomization and the primary outcome assessment actually ranged from 107 to 229 days. Table S10 below shows how the patients were distributed in that range and how the time related to community ambulation status.

eTable 10. Proportion (%) who were community ambulators at 16-week assessment by timing of the assessment

Time between randomization and the primary outcome assessment	Proportion (%) who were community ambulators
16 to 17 weeks post-randomization	19/73 (26%)
>17 to 18 weeks post-randomization	10/57 (18%)
>18 to 20 weeks post-randomization	4/33 (12%)
>20 weeks post-randomization	7/24 (29%)

As can be seen in Table S10, the relationship between time lag and community ambulation does not appear to be linear.

Table S11 shows the relationship between intervention group and time lag.

eTable 11. Number (%) in categories defined by timing of the 16-week assessment, by treatment group

Time between randomization and the primary outcome assessment	Training (n=91)	Active control (n=96)
16 to 17 weeks post-randomization	39 (43%)	34 (35%)
>17 to18 weeks post-randomization	24 (26%)	33 (34%)
>18 to 20 weeks post-randomization	16 (18%)	17 (18%)
>20 weeks post-randomization	12 (13%)	12 (13%)

We estimated the Training-control comparison in a model that adjusted for timing using the categorical variable for time between randomization and assessment shown in Tables S10 and S11. The results were very similar to those in the unadjusted analysis (estimated difference 4.5%, one-sided 97.5% CI −∞, 16.0%, one-side p=0.22).

1.9 Primary analysis considering those who died or who were too ill to perform the test as indeterminates

In the pre-specified analysis, those who died or were judged to be too sick to walk at least 300 m in 6 minutes were adjudicated as non-community ambulators. We performed an additional post-hoc analysis treating these participants as having indeterminate outcomes. For the 16-week assessment, there were four such participants: two in each intervention group. After classifying these participants as indeterminate, we obtained the following results:

In Training, 22/94 (23%) were community ambulators.
In active control, 18/99 (18%) were community ambulators.
Z-score=0.89. This does not exceed the threshold for null hypothesis rejection.
One-sided p-value=0.19
Estimated difference (Training minus control) (one-sided 97.5% CI): 5.2% (−∞, 16.7%)

These findings are almost identical to those obtained when these participants were treated as non-community ambulators.

1.10 Weighted estimating equations approach to adjust for possible biases due to excluding the indeterminates from the primary analysis

Reclassifying those who died or were too ill to walk at least 300 m in 6 min as indeterminates increases the number of individuals excluded from the primary analysis from 13 to 17. To address concerns that this could lead to bias, in a post-hoc analysis, we used a weighted estimating equations approach. To implement this approach, we first fit a stepwise logistic regression model to estimate the probability that a participant would have a missing outcome given baseline predictors. The candidate variables for this prediction model were those in Table 1 of the main paper. This stepwise model selected the following variables for the model: age, baseline Six-Minute Walk Test, NHATS, gait speed, cardiac disease, treatment assignment, and diabetes. Using this model, we estimated the probability that each participant would provide a non-missing outcome given the participant's covariates. We then estimated the Training-control difference in community ambulation, weighting each observation by the inverse of the probability that an outcome would be non-missing. This appropriately upweights observations that have a higher probability of being missing. This model should provide unbiased estimates if the data are missing at random given the covariates within a baseline principal stratum who would be alive and well enough to walk ≥300 m in 6 minutes regardless of treatment assignment[5]. Based on this weighted model, the estimated difference in community ambulation (Training-control) was 5.4% (one-sided 97.5% CI −∞, 17.0%, one-sided p-value p=0.18), which is almost identical to the result of the primary analysis.

eReferences

1. Beach ML, Meier P. Choosing covariates in the analysis of clinical trials. *Controlled clinical trials.* 1989;10(4 Suppl):161s-175s.
2. Magder LS. Simple approaches to assess the possible impact of missing outcome information on estimates of risk ratios, odds ratios, and risk differences. *Controlled clinical trials.* 2003;24(4):411-421.
3. 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(5):743-749.
4. Palombaro K, Craik R, Mangione K, Tomlinson J. Determining meaningful changes in gait speed after hip fracture. *Physical therapy.* 2006;86(6):809-816.
5. Shardell M, Hicks GE, Ferrucci L. Doubly robust estimation and causal inference in longitudinal studies with dropout and truncation by death. *Biostatistics (Oxford, England).* 2015;16(1):155-168.