ISSN: 2189-8448

Physical Therapy Research

VOLUME 25, ISSUE 2, 2022

Japanese Society of Physical Therapy

Contents Volume 25 Issue 2 2022

Review

Scientific Research Article

Brief Report

PHYSICAL THERAPY RESEARCH

REVIEW

Interpreting Results from Statistical Hypothesis Testing: Understanding the Appropriate P-value

Eiki TSUSHIMA, PT, MSc, PhD

Graduate School of Health Sciences, Hirosaki University, Japan

ABSTRACT. Clinical research based on epidemiological study designs requires a good understanding of statistical analysis. This paper discusses the common misconceptions of p-values so that researchers and readers of research papers will be able to properly present and understand the results of null hypothesis significance testing (NHST). The p-values calculated by NHST are categorized as three different types: "significant at p <0.05," "significant at p <0.01," or "not significant." If specified, they may be written as p = 0.124. The 95% confidence interval (CI) of the supplementary statistics is presented regardless of the p-value, and the range of the CI is observed and discussed to determine whether the results are clinically valid. The effect size (ES), which is a measure of the magnitude of the effect, is also referenced and discussed. However, the ES should not be overestimated. It is important to examine the actual descriptive statistics and consider them comprehensively as much as possible. A high detection power of 80% or more indicates that NHST with high accuracy was applied. However, even when it falls below 80%, it is important to consider the limitations of the study, because the results are not completely useless.

Key words: Null hypothesis significance testing, P-value, Confidence intervals, Effect size, Power

(Phys Ther Res 25: 49–55, 2022)

Clinical research based on epidemiological study designs requires a good understanding of statistical analysis. This is essential not only for the researcher who actually conducts the study and performs the statistical analysis but also for the clinical specialist who reads the research paper.

We, physical therapists, are not statisticians, and so our understanding of statistics is not sufficient. However, because we are increasingly exposed to clinical research, we must avoid misinterpretations. Interpretations of p-values obtained by null hypothesis significance testing (NHST) is a basic knowledge but is often misunderstood. Although it has been a long time since the American Statistical Association (ASA) reported its statement on p-values[1], it is assumed that the interpretation of p-values is not sufficiently widespread. Andreu et al.[2] surveyed paramedics and respiratory therapists (n = 376) about their knowledge of p-values and reported that 63.1% (237 persons) did not understand them correctly. The lack of training for research and the fact that less than six research papers were read in a year by the subjects were the suggested reasons. In an online survey[3] of 1576 researchers, conducted by *Nature*, more than 90% of the respondents agreed that robust design of experiments, understanding of appropriate statistics, and mentorship are important for reproducible research. Mentorship was also found to be important.

This paper explains the nature of the p-value based on several reports and other statistics that supplement the lack of information. By addressing common misconceptions of the p-value, the paper aims to help researchers, or readers of research papers, to properly present or understand the results of NHST.

Clinical Significance of p-value

What is a p-value?

The p-value is a decision indicator obtained through NHST. In statistics, the p-value is the probability of rejecting the null hypothesis (H_0) when it is true and erroneously adopting the alternative hypothesis (H_1). This probability of rejection is called the level of significance, Type I error, or α

Received: February 16, 2022
Accepted: March 2, 2022
Advance Publication by J-STAGE: May 13, 2022
Correspondence to: Eiki Tsushima. Graduate School of Health Sciences, Hirosaki University, 66-1, Hon-cho, Hirosaki, Aomori 036-8564, Japan
e-mail: pteiki@hirosaki-u.ac.jp
doi: 10.1298/ptr.R0019

Table 1. Determination and errors of hypotheses in NHST

		Judgment	
		Fail to reject H_0	Reject H_0
Actual	H_0 true (H_1 false)	Correct $1-\alpha$	Type I error α (=p-value)
	H_1 true (H_0 false)	Type II error β	Correct $1-\beta$ (power)

H_0, null hypothesis; H_1, alternative hypothesis

NHST, null hypothesis significance testing

(Table 1). We conventionally set the level of significance at $p = 0.05$ or $p = 0.01$. For example, in the case of a test of the difference of means, we reject the null hypothesis that "the means are equal" if the p-value is $p < 0.05$, which is interpreted as "the means are not equal." It is erroneous to interpret this as "the means are equal"[4,5].

The graph in Figure 1 shows the results of the analysis of the difference between the mean of population A ($n_A = \infty$) (population mean μ_A) and the mean of population B ($n_B = \infty$) (μ_B), when they are equal ($\mu_A = \mu_B$) and when they are not equal ($\mu_A \neq \mu_B$). H_0 is the test of the difference of means at the 5% significance level $_B$ for $\mu_A = \mu_B$. An example is given in Figure 1a. It is the 95% ($1-\alpha$) that correctly determines that H_0 is true and there is no difference, and conversely, it is the same 95% ($1-\beta$) (power) that correctly determines that H_1: $\mu_A \neq \mu_B$ is true and there is a difference. This is the result of highly accurate NHST. Because power is unknown for many null hypothesis significance tests, the power in Figure 1b can be as low as 80% (and of course, it can be large). In null hypothesis significance tests with a 5% significance level, even when there is a "significant difference at $p < 0.05$," information about the Type I error is unknown, except for the fact that it is 5%. The above points are common to all null hypothesis significance tests, not only for tests of differences in means but also for tests of correlation, regression analysis, and analysis of variance.

The p-value, which is frequently mentioned in many research reports, is a familiar but unusual number. Even if it is said that "there was a significant difference of less than 5%," it is impossible to understand what the 5% means in the real world.

For example, when the probability that a person does not have a disease after a certain test is as small as $p = 0.03$, the probability that the person does not have a disease is too small, and therefore, the person is considered to have a disease. However, it is not clear how much it means to be free of a disease at $p = 0.03$. Although there have been many reports stating that p-values alone do not provide an accurate picture of the facts[1,4], there are still many reports that make judgments based on null hypothesis significance test p-values alone. The p-value should be used as an initial entry point for judgment, and it is necessary to refer to other statistical values afterward to deepen understanding.

Is there no effect when there is no significant difference ($p > 0.05$)?

I have mentioned above that it is wrong to interpret a p-value ≥ 0.05 as "there is no difference in the means" or "there is no correlation ($r = 0$)." To be precise, it should be interpreted as "there is insufficient evidence for a difference in the means"; i.e., it is unclear whether there is a difference.

The problem of this misunderstanding has been taken up extensively in *Nature*[6], and in 791 papers in five journals, 51% are reported as relevant. Nevertheless, there are still many cases in which a p-value ≥ 0.05 in NHST is described as "no significant difference" and treated as if there were "no effects at all." This should be corrected to "there was no significant difference."

p-value and related statistics

The p-value is a statistic calculated by NHST, but it can be controlled by the data. P-value, effect size (ES), p (α), and power ($1-\beta$) are interrelated, and once three of these are determined, the remaining item can be determined. The relationship between these statistics value is shown in Table 2. Increasing the sample size increases the power and decreases the p-value in NHST, while increasing the ES decreases the required sample size, increases the power, and decreases the p-value. Thus, intentional manipulation of the sample size and ES results in a corresponding change in the p-value, which can be controlled.

Considering this mechanism, we can understand the ambiguous nature of p-values. The p-value becomes smaller just by increasing the sample size, and associating it with the ES causes a major problem[2,7].

Notation of p-values

Most statistical analysis software outputs the p-values of NHST results in detail. By referring to the results, you can enter more accurate p-values, for example, $p = 0.725$ or $p = 0.125$. However, this is erroneous.

In the case of a test of the difference of means, a significant result should be stated only as "there is a significant difference at $p < 0.05$ (or $p < 0.01$)." If there is no significant difference (i.e., $p \geq 0.01$), the difference is judged to be "not significant"[4]. For example, comparing $p = 0.725$ and $p = 0.125$, "$p = 0.725$ is more likely to have no difference" is also an error. In randomized controlled trials, p-values such as $p = 0.852$ are sometimes listed to argue that there is no significant difference in background factors between groups, but this may be misleading due to redundant expressions. However, the redundant expression may mislead readers. In a study that examined 30 main nursing journals[8], it was found that "$p = 0.000$," which is an error, was documented in 93.3% of the journals. Documenting $p = 0.000$ was likely due to the author wanting to strongly emphasize that the results obtained were significant. However, this is misleading to the readers.

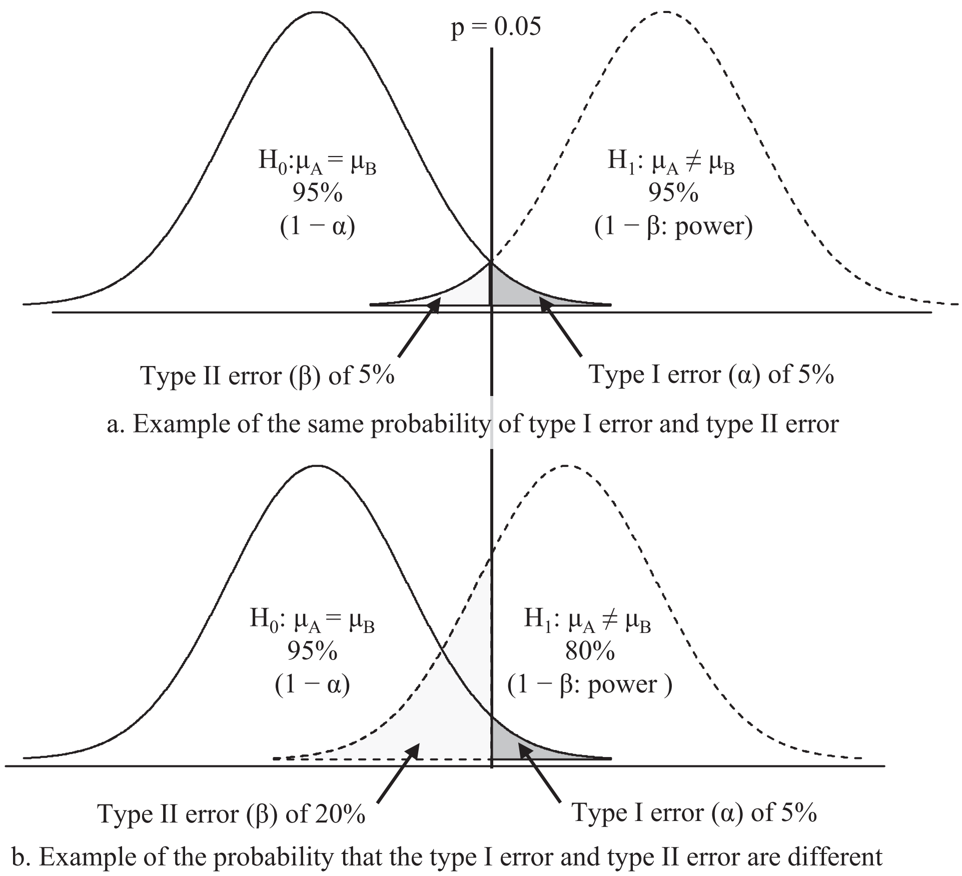


Fig. 1. Probability of "no difference" and "difference" in differences test

Type I error (α) is the probability of mistakenly assuming there is a difference when there is no difference (α = 0.05), and a type II error (β) is the probability of mistakenly assuming there is no difference when there is a difference $_1$(H = $_0$). In the case of a, the probabilities of correctly judging that there is a difference and there is no difference, 1 − α = 0.95 and 1 − β = 0.95, are equal. However, in the actual difference test, the probabilities, 1 − α and 1 − β, are often different, as in b. In this figure, the distributions of H_0 and H_1 have the same shape, but in reality, they are not necessarily the same.

Table 2. Relationship between sample size, power, ES, and p-value

	Sample size	Power	p	ES
Sample size (n)↑		↑	↓	-
Power (1 − β)↑	↑		-†	-
p (α = 0.05)	0.05	0.05		0.05
ES↑	↓	↑	↓	

↑, increase; ↓, decrease; -, unrelated

The orientation of the arrows is also related to the opposite orientation.

† Power is larger when the p-value is reduced, but is shown as -, because in reality, it is constant at p = 0.05.

ES, effect size

In addition, even when the p is close to 0.05, such as p = 0.049 or p = 0.051, it is questionable whether the results should be clearly separated into two[9)]. Depending on the nature of the study, a precise statement such as p = 0.210 may be required. In some cases, the p-value should be presented to three decimal places, although it is judged as "there was a significant difference at $p < 0.05$ (or $p < 0.01$)" or "there is no difference"[10)]. In any case, as long as we do not understand how much difference the p-value causes in reality, it is a never-ending discussion. Being able to interpret the results comprehensively with supplementary statistics would minimize the problem.

Statistical Values That Supplement the p-value

In medical research papers, NHST is performed and many p-values are described. The display of the confidence interval (CI) and result values is recommended as aid in interpreting p-values[1,2,5,7,10,11)], and the display of detection power is also helpful.

Confidence interval (%)

A parameter of characteristic value is a statistical value such as the difference of means, correlation coefficient, or regression coefficient in a population. The CI outputs the inferred value of the range containing the parameter. A

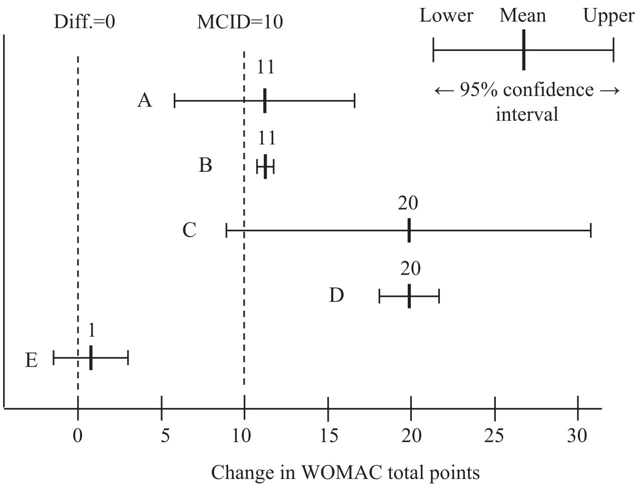


Fig. 2. Example of mean and 95% CI of change (difference) between pre- and post-TKA surgery

The CI uses two values to represent the range: the upper limit (upper) and the lower limit (lower). Five cases of 95% CIs (A–E) were given. Mean is the mean of the difference between preoperative and postoperative changes. For A–D, all the lower values exceeded 0 (diff. = 0 indicates that the difference is 0); thus, there was a significant change (difference) at $p < 0.05$. For E, the lower limit of the 95% CI was below 0, and therefore, a significant difference could not be detected.

CI, confidence interval; TKA, total knee arthroplasty; WOMAC, Western Ontario and McMaster Universities osteoarthritis index

parameter exists with a confidence of 95% in the 95% CI range. The 95% CI was originally compared with the 5% significance level. The 99% CI should be presented in comparison to a significance level of 1%[7]; however, the trend is to present only the 95% CI.

To be precise, it does not mean that the range of the 95% CI contains the parameter with a probability of 95%[4,12,13]. In reality, a parameter is either present (100%) or absent (0%) in the range. The truth of this is a difficult matter to understand, as no conclusion has been reached even in the field of statistics. However, because the range of CI includes a parameter with a high probability, it is possible to estimate the effect more concretely with the CI than with the p-value only.

When the difference is significant at $p < 0.05$, we can estimate the range of the size of the parameter (upper and lower limits) by looking at the CI, and we can consider the extent of the difference by referring to an index such as the minimal clinically important difference (MCID). When the 0.05 difference is not significant, the smaller the upper limit of the CI and the smaller the range of the CI, the smaller the positive effect[12].

To give a simple example, Clement et al.[14] evaluated Western Ontario and McMaster Universities osteoarthritis index (WOMAC) scores 1 year after total knee arthroplasty (TKA) and calculated the MCID before and after surgery. As a result, the MCID of the total WOMAC score was calculated as 10. Suppose that we conducted a study to evaluate the WOMAC score before and after TKA and obtained the results as shown in Figure 2; if the WOMAC score does not improve to a mean value that is higher than 10, there is no clinical efficacy. In cases A to D, a significant change (difference) of $p < 0.05$ was observed in all. A is the mean amount of change (different mean) and is 11, which is certainly higher than the MCID. However, the lower limit of the 95% CI is lower than 10. Moreover, with NHST, even if there is a significant difference, considering the 95% CI, it is difficult to assert that there is a clinically effective difference. In case B, the 95% CI is very narrow and the lower limit exceeds 10. For C, the mean is 20, which is high; however, the lower limit of the 95% CI is below 10, and so the interpretation is the same as that for A. For D, both the mean and lower limit highly exceed 10, and therefore, there is statistical significance and a significant difference even clinically.

E is an example of a non-significant difference in NHST. The mean of the differences is 1, and the lower limit of the 95% CI is below 0, which is a negative value (possibly worsened). Empirically, if the WOMAC does not exceed 5, it is considered to be ineffective, and we may conclude that the difference in case E is not statistically significant and clinically equivalent to no effect.

One disadvantage of these CIs is that many of them can only be calculated using parametric methods (NHST for data that follow a normal distribution); most nonparametric methods, such as the Mann–Whitney and Friedman's tests, cannot be used. The Hodges–Lehmann estimator is used for this purpose, but it is not clear whether it is accurate. Recently, CI of the Mann–Whitney test[15] has been devised. Another disadvantage of the CI is that it is an interval estimate, which makes it difficult to interpret, especially when the sample size is small, because the width of the interval becomes wide. Nevertheless, it should be useful as supplementary information to the p-value.

Effect size

ES is the effect size, and unlike CI, it is not a parameter but the effect of the difference of means, the correlation coefficient or the regression coefficient obtained from the actual sample. N = 10 means the ES of N = 10, N = 100 means the ES of N = 100. Thus, each time you change the sample, you get a different value.

There are two types of ESs, unstandardized effect sizes (U-ESs) and standardized effect size (S-ESs), which have been devised by various researchers. In actuality, 6[15] to 70[16] types have been recognized. In this section, we describe the most frequently used ESs.

The U-ES is the value of the mean or difference of means, regression coefficient, etc. The U-ES varies depending on the size of units and standard deviation of the raw data, making direct comparison with other reports difficult. Therefore, S-ES, which eliminates the effect of units and standard deviation, is often used.

S-ES can be further divided into the d-family and r-family. The d-family includes Cohen's d, Hedges' g, and

Table 3. Observed ORs with 95% CIs when changing data

a. Normative data

		Disease		Odds
		Yes	None	
Testing	Positive +	1,000	500	2.00
	Negative −	500	1,000	0.50
	Total	1,500	1,500	
		OR	95% CI-L†	95% CI-U‡
		4	3.44	4.66

b. Double the number of people without disease in the reference data

		Disease		Odds
		Yes	None	
Testing	Positive +	1,000	1,000	1.00
	Negative −	500	2,000	0.25
	Total	1,500	3,000	
		OR	95% CI-L	95% CI-U
		4	3.51	4.56

c. Five times the number of people without disease in the reference data

		Disease		Odds
		Yes	None	
Testing	Positive +	1,000	2,500	0.40
	Negative −	500	5,000	0.10
	Total	1,500	7,500	
		OR	95% CI-L	95% CI-U
		4	3.56	4.50

d. 10 times the number of people without disease in the reference data

		Disease		Odds
		Yes	None	
Testing	Positive +	1,000	5,000	0.20
	Negative −	500	10,000	0.05
	Total	1,500	15,000	
		OR	95% CI-L	95% CI-U
		4	3.57	4.48

e. 1/10 times the number of people in the reference data

		Disease		Odds
		Yes	None	
Testing	Positive +	100	50	2.00
	Negative −	50	100	0.50
	Total	150	150	
		OR	95% CI-L	95% CI-U
		4	2.47	6.46

(*Continued*)

Table 3. (*Continued*)

f. 1/100 times the number of people in the reference data

		Disease		Odds
		Yes	None	
Testing	Positive +	10	5	2.00
	Negative −	5	10	0.50
	Total	15	15	
		OR	95% CI-L	95% CI-U
		4	0.88	18.26

g. 1/500 times the number of people in the reference data

		Disease		Odds
		Yes	None	
Testing	Positive +	2	1	2.00
	Negative −	1	2	0.50
	Total	3	3	
		OR	95% CI-L	95% CI-U
		4	0.13	119.23

†95% CI-L: lower bounds of the 95% CI

‡95% CI-U: upper bounds of the 95% CI

OR, odds ratio; CI, confidence interval; L, lower; U, upper

Glass's delta. The r-family includes ES r, Pearson's correlation coefficient r, correlation ratio η, phi coefficient, Cramer's V, and so on. The S-ES can be calculated by nonparametric methods.

The d-family expresses the ES as a numerical value with no upper limit 0, whereas the r-family expresses the ES as a range of 0–1, which has the advantage of being easier to interpret.

The criteria for evaluating the size of ESs are also shown. In Cohen's d, ≥0.2 are small effects, ≥0.5 are medium effects, and ≥0.8 are large effects[17)]. Brydges[18)] examined the S-ES based on multiple meta-analyses published in gerontological journals and determined practical values using the quartiles criteria. The actual S-ES was smaller than Cohen's benchmark and was overestimated.

Because there are many calculation methods for ES, it is necessary to clearly state which ES was used. In addition, more detailed information can be obtained by describing both U-ES and S-ES[5)]. This should be considered essential for the odds ratio (OR), which is one of the ESs.

Table 3 shows examples of ORs and 95% CIs of ORs for data with more or fewer people based on the data in a. The ORs are the same for all cases, but some subtables differ greatly when looking at the number of patients. The OR does not change when the number of patients with no disease is increased, as shown in Table 3b–3d. Although the number of people in a and g are very different, the OR remains constant. However, the range of the 95% CI becomes wider as the number of people decreases. In other words, it is necessary to present the raw data or to provide additional information on the 95% CI, especially when presenting ratio values such as OR.

Power (1 − β)

Power is the probability of obtaining a significant result (to judge correctly) when the alternative hypothesis H_1 is true (Table 1). In the example of the difference of means test, it is the probability of correctly judging that there is a significant difference when there is truly a difference in the population mean. The 95% entered in $\mu_A \neq \mu_B$ in Figure 1a and the 80% in Figure 1b represent the power. Even though there is a difference in the population mean, the probability of incorrectly judging that "there is a significant difference" is called a type II error (β).

There is a trade-off between power and p-value, but in reality, the p-value is fixed at p = 0.05. Therefore, it is described as unrelated in Table 2. However, power is not determined only by p-value.

A more accurate null hypothesis significance test also has a higher power, because a higher power (1 − β) means a smaller β. The power of a is higher than that of b in Figure 1, indicating that the null hypothesis significance test is more accurate. Determining the power in this way is useful for evaluating the accuracy of the test.

Although there is theoretical basis for doing so[19], it is customary to set the power to 0.8 (80%) or 0.95 (95%)[7]. Generally, Type II error (β) is taken as 0.2 and the value is 0.87.

There are two types of power analyses, the a priori method and the post hoc procedure. The a priori method determines the power and ES before the study and determines the sample size required for the null hypothesis significance test in advance. Because the power is fixed at 0.8, the sample size can be obtained if the ES is known in advance. Free software such as G*Power[20] is useful for calculating the detection power, and the ES is generally obtained from previous similar studies or preliminary studies. If multiple null hypothesis significance tests are to be performed, choose the largest sample size for each.

In the post hoc procedure, the NHST is actually performed after the study is completed to evaluate the power of detection. If the power is greater than 80% for all the null hypothesis significance tests, a highly accurate null hypothesis significance test has been applied. In some cases, the power may be low because of insufficient sample size due to dropouts or other factors, and in other cases, the power may exceed 80% because of a large ES.

It does not matter if the power is less than 80%. It is important to consider the reasons for the drop and to be aware of the limits of what can be mentioned.

Conclusion

This paper discussed the meaning of p-values and the need for supplementary statistics with reference to various reports. These are summarized in the following section on how to appropriately present the results of null hypothesis significance tests. In the future, we should be aware of the need to be careful not to let misunderstandings lead us in the wrong direction.

1. Judgments based on the p-values calculated by NHST are the following three: "significant at p <0.05," "significant at p <0.01," or "not significant." If specified, "not significant" may be written as p = 0.124.
2. If the statistical software outputs the 95% CI, the value should be added. The 95% CI should be presented regardless of the p-value, and the range should be observed and discussed from a professional perspective to determine whether the results are clinically valid.
3. ES is one indicator of effect size. It is worth considering, but it should not be overestimated. Whenever possible, actual descriptive statistics should also be taken into consideration.
4. In particular, for ESs expressed as ratios such as OR, raw data, descriptive statistics, and 95% CI should be used as additional information to avoid misinterpretation.
5. A high detection power of 80% or more indicates that the NHST with high accuracy was applied. However, it is important to consider the limitations of the study, because the results are not completely useless even when the power is below 80%.

The most important thing is to follow a good research design and to have data that are small in bias and well understood. No amount of good statistical analysis can solve this problem.

Conflict of Interest: The author declares no conflicts of interest.

References

1) Wasserstein RL, Lazar NA, *et al.*: The ASA's statement on p-values: context, process, and purpose. Am Stat. 2016; 70: 129–133.
2) Andreu MF, Ballve LPD, *et al.*: Is the p-value properly interpreted by critical care professionals? online survey. Rev Bras Ter Intensiva. 2021; 33: 88–95.
3) Baker M: 1,500 scientists lift the lid on reproducibility. Nature. 2016; 533: 452–454.
4) Greenland S, Senn SJ, *et al.*: Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. Eur J Epidemiol. 2016; 31: 337–350.
5) Nakagawa S, Cuthill IC: Effect size, confidence interval and statistical significance: a practical guide for biologists. Biol Rev Camb Philos Soc. 2007; 82: 591–605.
6) Amrhein V, Greenland S, *et al.*: Scientists rise up against statistical significance. Nature. 2019; 567: 305–307.
7) Knapp TR, Brown JK: Ten statistics commandments that almost never should be broken. Res Nurs Health. 2014; 37: 347–351.
8) Wu Y, Zhou C, *et al.*: Statistical reporting in nursing research: addressing a common error in the reporting of p values (p = .000). J Nurs Scholarsh. 2020; 52: 688–695.

9) Lang T: Twenty statistical errors even you can find in biomedical research articles. Croat Med J. 2004; 45: 361–370.

10) Sil A, Betkerur J, *et al.*: P-value demystified. Indian Dermatol Online J. 2019; 10: 745–750.

11) Halsey LG: The reign of the p-value is over: what alternative analyses could we employ to fill the power vacuum? Biol Lett. 2019; 15: 20190174.

12) Hawkins AT, Samuels LR: Use of confidence intervals in interpreting nonstatistically significant results. JAMA. 2021; 326: 2068–2069.

13) Morey RD, Hoekstra R, *et al.*: The fallacy of placing confidence in confidence intervals. Psycho Bull Rev. 2016; 23: 103–123.

14) Clement ND, Bardgett M, *et al.*: What is the minimum clinically important difference for the WOMAC index after TKA? Clin Orthop Relat Res. 2018; 476: 2005–2014.

15) Fay MP, Malinovsky Y: Confidence intervals of the Mann-Whitney parameter that are compatible with the Wilcoxon-Mann-Whitney test. Stat Med. 2018; 37: 3991–4006.

16) Schober P, Bossers SM, *et al.*: Statistical significance versus clinical importance of observed effect sizes: what do P values and confidence intervals really represent? Anesth Analg. 2018; 126: 1068–1072.

17) Cohen J: A power primer. Psychol Bull. 1992; 112: 155–159.

18) Brydges CR: Effect size guidelines, sample size calculations, and statistical power in gerontology. Innov Aging. 2019; 3: igz036.

19) Banerjee A, Chitnis UB, *et al.*: Hypothesis testing, type I and type II errors. Ind Psychiatry J. 2009; 18: 127–131.

20) Institut für Experimentelle Psychologie, Heinrich-Heine-Universität [Internet]. Germany: G*Power statistical power analyses for Mac and Windows, Windows [updated 2020 March 17; cited 2022 Feb 5] Available from: https://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower.

PHYSICAL
THERAPY
RESEARCH

SCIENTIFIC RESEARCH ARTICLE

Cost-effectiveness Analysis of Combined Physical and Cognitive Exercises Programs Designed for Preventing Dementia among Community-dwelling Healthy Young-old Adults

Gohei KATO, PT, PhD[1], Takehiko DOI, PT, PhD[2], Hidenori ARAI, MD, PhD[3] and Hiroyuki SHIMADA, PT, PhD[2]

[1)]Department of Physical Therapy, School of Rehabilitation, Tokyo Professional University of Health Sciences, Japan
[2)]Department of Preventive Gerontology, Center for Gerontology and Social Science, Research Institute, National Center for Geriatrics and Gerontology, Japan
[3)]National Center for Geriatrics and Gerontology, Japan

ABSTRACT. Objective: This study aims to estimate the cost-effectiveness of combined physical and cognitive programs designed to prevent community-dwelling healthy young-old adults from developing dementia. Methods: The analysis was conducted from a public healthcare and long-term care payer's perspective. Quality-adjusted life years (QALYs) and expenses for health services and long-term care services were described in terms of effectiveness and cost, respectively. A thousand community-dwelling healthy adults aged 65 years were generated through simulation and analyzed. The incremental cost-effectiveness ratio (ICER) of adults with preventive program intervention compared to those with nonintervention was simulated with a 10-year cycle Markov model. The data sources for the parameters to build the Markov models were selected with priority given to higher levels of evidence. The threshold for assessing cost-effectiveness was set as less than 5,000,000 Japanese yen/QALY. Results: The ICER was estimated as −5,740,083 Japanese yen (US$−57,400)/QALY. Conclusion: A program targeting community-dwelling healthy young-old adults could be cost-effective.
Key words: Cost-effectiveness analysis, Prevention, Dementia, Combined physical and cognitive exercises, Older adults

(Phys Ther Res 25: 56–67, 2022)

The rapid aging of populations is being seen in many parts of the world[1)]. The same trend is also seen in Japan, where in 2018, the population aged 65 years and older was 35.89 million, making up 28.4% of the total population[2)]. Aging induces an increase in the number of frail older adults and the cost of their healthcare and long-term care[3)].

One of the typical and problematic diseases related to aging is dementia, which not only induces deterioration of health but also increases the cost of medical and long-term care and makes it difficult to sustain the health and long-term care systems of Japan[4,5)]; Japanese society's total costs from dementia have been estimated to reach 24.3 trillion Japanese yen by 2060. Therefore, there has been a call for the introduction of preventive programs to avoid the risk of cognitive decline and dementia[6)].

It has been suggested that people who experience multiple-domain mild cognitive impairment (MCI) have a high risk of developing dementia[7)]. MCI has been indicated to be highly reversible to a healthy state if the appropriate intervention is received as early as possible[7–9)]. Exercise has been suggested as an effective intervention to prevent dementia[8,9)], and combined physical and cognitive programs[10,11)] have been suggested to be especially effective if introduced early, when young-old adults are in a healthy or MCI state before developing dementia[7)].

Received: October 23, 2021
Accepted: March 28, 2022
Advance Publication by J-STAGE: June 10, 2022
Correspondence to: Gohei Kato. Department of Physical Therapy, School of Rehabilitation, Tokyo Professional University of Health Sciences, 2-22-10, Shiohama, Koto-ku, Tokyo 135-0043, Japan
e-mail: goheik@tpu.ac.jp
doi: 10.1298/ptr.E10153

To achieve dementia-free communities[7)], introducing a sufficiently cost-effective public health program might be beneficial to society. A Markov model has been suggested to be useful for health-economic analyses because, for example, it can incorporate risks that are continuous over time[12)]. Indeed, it has been applied to health-economic analyses for health-related preventive services in international cases[13)] and in Japan[14)]. In the same way, a health-economic evaluation of primary prevention programs for dementia has also been recommended[15)]. Although there has been a study of health-economic analysis for pharmacological interventions for people who have already developed dementia[16)], to our knowledge, no studies have applied health-economic analysis to primary prevention programs for dementia using a Markov model.

Understanding the value of the cost-effectiveness of a prevention program might aid in making a decision on introducing it into the community.

This study therefore aims to estimate to what extent combined physical and cognitive programs designed to prevent community-dwelling healthy young-old adults from developing dementia would be cost-effective using a Markov modeling analysis.

Methods

The analysis was conducted from a public healthcare and long-term care payer's perspective[17)]. Quality-adjusted life years (QALYs) were described in terms of effectiveness, and expenses in terms of cost, for health services and long-term care services. The target population was Japanese community-dwelling healthy young-old adults; the subjects were aged 65 years. The data sources for the parameters to build the Markov models[17)] were selected with priority given to those with higher levels of evidence, such as systematic reviews, and those developed in Japan, as recommended elsewhere[17)].

The analysis was conducted with TreeAge Pro Healthcare version 2021 R1.1 (TreeAge Software, Williamstown, MA, USA). This study was approved by the Ethics Committee of the Tokyo Professional University of Health Science (TPU-20-002). The sources of the parameters are shown in Table 1[7,17–26)].

The group assumed not to receive the preventive exercise program was defined as the nonintervention group (NIG); the others who participated in the program were defined as the intervention group (IG). The incremental cost-effectiveness ratio (ICER) of the IG to the NIG was analyzed using a 10-year Markov model. One year was set as one cycle. The threshold for assessing cost-effectiveness was set as less than 5,000,000 yen/QALY[27)]. Additionally, the incremental net monetary benefit (INMB)[28)] was calculated by equating 1 QALY to 5,000,000 yen to consider the cost-effectiveness of the preventive exercise program from a restricted societal perspective[17)]. The INMB was used to consider the 10-year productivity loss when the subjects did not receive the intervention. In this study, 100 yen was considered equal to 1 United States dollar (US$) to understand the results from an international perspective.

Structure of the Markov models

Two 10-year cycle Markov models for NIG (Fig. 1A) and IG (Fig. 1B) were built. One thousand healthy community-dwelling adults aged 65 years in the first year of the simulation were selected for the analysis.

The six transition states indicated in a previous study[20)] to categorize people with predementia, MCI, or dementia were applied. The NIG (Fig. 1A) had the states (a) well, (b) mild level of dementia, (c) middle level of dementia, (d) severe level of dementia, (e) dead with dementia, and (f) dead without developing dementia. Similarly, seven transition states for the IG (Fig. 1B) were defined by adding the MCI state to those of the NIG.

The states of death were set as the endpoints of the models. Those who were in a state of dementia were assumed to receive medical treatment. The severity of dementia was classified into each care needs level of the Japanese long-term care system[29)] by a physical therapist who was certified to give a specialist opinion on care needs level; this method, asking for a specialist opinion for the classification, follows a previous cost-effectiveness study for dementia conducted in Japan[16)]. Those who were in the state of a mild level of dementia were assumed to be in support levels 1 or 2, or care need level 1. They were assumed to receive preventive day service for dementia or day service for dementia, and preventive short stay for dementia or short stay for dementia from the long-term care system. Those who were in the middle level of dementia were assumed to be in care need levels 2 or 3. They were assumed to receive day service for dementia and short stay for dementia. Those who were in the state of severe dementia were assumed to be in care need levels 4 or 5. They were assumed to stay in community-based facilities. The possibility of reversion from the MCI state to the well state was only present in the IG, while reversion after dementia was not considered possible. QALY and costs were calculated with a half-cycle correction[13)] and translated to the current value with a 2% per year discount rate[17)].

The individuals in the IG were assumed to receive screening for MCI and dementia once a year, and the program with combined physical and cognitive exercises involved a 90-minute weekly session focused on physical and cognitive activities, which was conducted 40 times in the first year (details are reported elsewhere[11)]). Based on a previous study[11)], 15 individuals were assumed to participate in each session conducted by 2 geriatric physical therapists and 5 instructors in public halls. As a note, the number of participants in each session for this study was smaller than that of the previous study (15 individuals vs. 16–32 individuals). This was because the space of public halls for the intervention in this study was assumed to be narrower than that of the fitness facilities in the previous study. From the

Table 1. Parameters related to the transition probability, effectiveness, costs, and discount rate

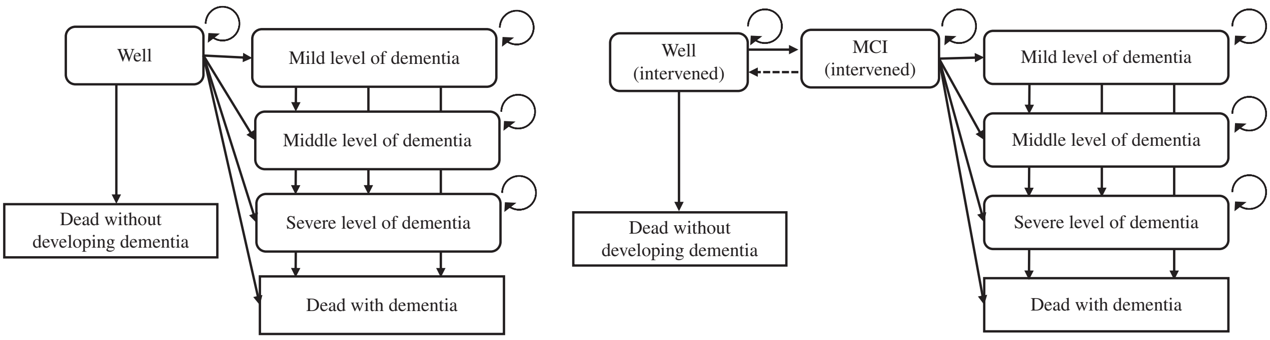
Parameter	Subgroup	Unit	Basic value	Range (lowest – highest)		Type of range	Distribution	Source
Transition probabilities and effectiveness								
Transition probability from state of well to dementia	65–69 years old	%/year	2.18	1.44	14.91	95%CI	Triangular	Ninomiya (2015)[18]
	70–74 years old	%/year	4.84			95%CI	Triangular	Ninomiya (2015)[18]
	75– years old	%/year	10.75			95%CI	Triangular	Ninomiya (2015)[18]
Transition probability from MCI to dementia		%/year	3.93	1.14	5.90	95%CI	Triangular	Shimada (2017)[7]
Probability of severity in dementia	Mild	%	38.50	30.80	46.20	±20%	Dirichlet	Asada (2013)[19]
	Moderate	%	24.10	19.28	28.92	±20%	Dirichlet	Asada (2013)[19]
	Severe	%	37.40	29.92	44.88	±20%	Dirichlet	Asada (2013)[19]
Transition probability from MCI to well		%/year	10.98	5.69	19.02	95%CI	Triangular	Shimada (2017)[7]
Transition probability from well to MCI		%/year	13/15 times of transition probability from state of well to dementia			95%CI	Triangular	Ninomiya (2015)[18] Asada (2013)[19]
Relative risk for death in the state of dementia		%/year	2.80	1.85	4.24	95%CI	Triangular	July (2021)[20]
Effect of combined physical and cognitive exercises to prevent the progression of MCI	All age	rate/year	0.65	0.55	0.76	95 % CI	Triangular	Blondell (2014)[21]
Effect of combined physical and cognitive exercises to prevent the progression of dementia	All age	rate/year	0.86	0.76	0.97	95 % CI	Triangular	Blondell (2014)[21]
Effectiveness	Well	QALY/year	1.00	1.00	1.00	95% CI	Triangular	Estimated by authors
	MCI	QALY/year	0.75	0.16	1.00	95% CI	Triangular	Landeiro (2020)[22]
	Dementia (Mild)	QALY/year	0.61	0.00	1.00	95% CI	Triangular	Landeiro (2020)[22]
	Dementia (Moderate)	QALY/year	0.41	0.00	1.00	95% CI	Triangular	Landeiro (2020)[22]
	Dementia (Severe)	QALY/year	0.21	0.00	0.74	95% CI	Triangular	Landeiro (2020)[22]
	Dead	QALY/year	0.00	0.00	0.00	95% CI	Triangular	Estimated by authors
Costs related to the program								
Cost for screening	Well MCI	yen/year	6,000	4,800	7,200	± 20 %	Triangular	Expert opinion by Shimada (2017)[7]
Number of individuals in a session	Well MCI	n/session	15	12	18	± 20 %	Triangular	Expert opinion by Shimada (2017)[7]

(*Continued*)

Table 1. (*Continued*)

Parameter	Subgroup	Unit	Basic value	Range (lowest – highest)		Type of range	Distribution	Source
Number of physical therapists for a session	Well MCI	n/session	2	1	3	Estimated	Triangular	Shimada (2017)[7]
Number of instructors for a session	Well MCI	n/session	5	4	6	± 20 %	Triangular	Expert opinion by Shimada (2017)[7]
Cost for a physical therapist for a session	Well MCI	yen/session	8,396	6,716	10,075	± 20 %	Triangular	Ministry of Health, Labour and Welfare (2020)[23]
Cost for a instructors for a session	Well MCI	yen/session	3,000	2,400	3,600	± 20 %	Triangular	Expert opinion by Shimada (2017)[7]
Numbers of sessions held in a year	Well MCI (First year)	sessions/year	40	32	48	± 20 %	Triangular	Shimada (2017)[7]
	Well MCI (From 2nd year)	sessions/year	2	2	2.4	± 20 %	Triangular	Estimated by authors
Cost for a public space to conduct the session	Well MCI	yen/session	900	720	1,080	± 20 %	Triangular	Obu city (2021)[24]
Cost for maintaining the program	Well MCI	yen/year	6,000	4,800	7,200	± 20 %	Triangular	Expert opinion by Shimada (2017)[7]
Costs for medical and long-term care								
Cost for day service for dementia	Dementia (Mild/Moderate)	yen/year	1,530,000	1,224,000	1,836,000	± 20 %	Triangular	Ministry of Health, Labour and Welfare (2019)[25]
Cost for preventive day service for dementia	Dementia (Mild)	yen/year	614,400	491,520	737,280	± 20 %	Triangular	Ministry of Health, Labour and Welfare (2019)[25]
Cost for preventive short stay for dementia	Dementia (Mild)	yen/year	471,600	377,280	565,920	± 20 %	Triangular	Ministry of Health, Labour and Welfare (2019)[25]
Cost for short stay use for dementia	Dementia (Mild/Moderate)	yen/year	1,315,200	1,052,160	1,578,240	± 20 %	Triangular	Ministry of Health, Labour and Welfare (2019)[25]
Cost for stay in community-based facility	Dementia (Severe)	yen/year	3,448,800	2,759,040	4,138,560	± 20 %	Triangular	Ministry of Health, Labour and Welfare (2019)[25]
Cost for medical treatment for dementia	Dementia (Mild)	yen/year	527,344	421,875.2	632,812.8	± 20 %	Triangular	Tomata (2014)[26]
	Dementia (Moderate)	yen/year	841,974	673,579.2	1,010,368.8	± 20 %	Triangular	Ministry of Health, Labour and Welfare (2019)[25]
	Dementia (Severe)	yen/year	1,035,384	828,307.2	1,242,460.8	± 20 %	Triangular	Ministry of Health, Labour and Welfare (2019)[25]
Discount rate		%/year	2.00	0.00	4.00	estimated	Triangular	Shiroiwa (2017)[17]

MCI, mild cognitive impairment; CI, confidence interval; QALY, quality-adjusted life year



A: Markov model for the non-intervention group (NIG) B: Markov model for the intervention group (IG)

Fig. 1. Markov models

NIG, non-intervention group; IG, intervention group; MCI, mild cognitive impairment

second year, the individuals who were in the well or MCI states were assumed to receive the screening once a year and the program twice a year.

Parameters

The basic values of the parameters were used to produce a base case of the results from a Markov cohort analysis. The range of values and distribution were applied for sensitivity analysis. The mean values among subgroups, such as sex and type of dementia, were used as needed to produce the base case, while the lowest and highest values among subgroups were used for the sensitivity analysis. Multiple years (n) of transition probabilities were converted to values for 1 year (p) using the following formula as needed[30]:

$$p = 1 - (1 - p)^{(1/n)}.$$

Transition probability

The basic values of the transition probabilities are shown in Table 1. The probability of newly developing dementia in a year was substituted with the estimated prevalence of dementia, following age groups in Japan[18]. The probabilities of newly developing dementia in a year for 66–69-year-old, 70–74-year-old, and 75-year-old adults were set at 2.18%, 4.84%, and 10.75%, respectively[18]. The mean transition probability from the well state to the MCI state was set at 13/15 times the probability of newly developing dementia per year. This was because the prevalence of dementia was 15%, while that of MCI was 13%, according to a study that analyzed 104,785 subjects from 10 areas in Japan[19]. The transition probability from the well state to death was referenced from the Japanese simple life table in 2019[31]. The mean transition probabilities of those 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, and 75 years old were calculated as 0.73%, 0.81%, 0.89%, 0.98%, 1.09%, 1.19%, 1.30%, 1.42%, 1.56%, 1.73%, and 1.92%, respectively.

Effectiveness

The utility values of subjects were used to indicate effectiveness in this study. The mean values of health-related quality of life scaled for predementia Alzheimer's disease, MCI, or dementia with EuroQol-5 Dimension (EQ-5D)[32] reported by Heßmann et al.[27,33] were used for the effectiveness and treated as QALY values (Table 1). The proxy-rated values, not self-rated values, were adapted to the model to maintain validity and take into account the caregivers' productivity loss[27]. The mean values for the QALY values per year for the states of well, MCI, mild severity of dementia, moderate severity of dementia, severe severity of dementia, and dead (with or without dementia) were 1.00, 0.75, 0.61, 0.41, 0.21, and 0.00, respectively. The values could not be below 0.00 or higher than 1.00.

Costs

The cost of one hour for a physical therapist, 2798.5 yen, was referenced from the basic statistical survey of wage structure in Japan in 2020[23]. This was an average of the values for a temporary physical therapist's hourly salary in organizations with 100–999 workers (4104 yen) and in organizations with 10–99 workers (1493 yen). The cost of a physical therapist for a session was calculated as (physical therapist's salary paid per hour × working hours × numbers of physical therapists in a session) ÷ numbers of individuals. The physical therapists were set to contribute three hours each to a session. The working hours included not only conducting exercise but also the preparation of and finishing up for the session.

The screening cost, 6000 yen, and an hourly cost of 1000 yen for an instructor to support the program were set by the authors who conducted the program[11,34]. As a screening test, the Mini-Mental State Examination (MMSE)[35], the Alzheimer's Disease Assessment Scale-Cognitive subscale (ADAS cog)[36], and the Wechsler Memory Scale-Revised[32] were set. The instructors were set to contribute three hours each to a session.

The annual medical cost for treating dementia was calculated based on a study that reported the monthly medical costs for treating dementia, depending on the care needs levels[26]. The mean values of these, depending on the severity of dementia, were used for the models. The medical costs for

Table 2. Base case values simulated by 10 years Markov cohort analysis

	Cumulative effectiveness (QALY)	Incremental effectiveness (QALY)	Cumulative cost (yen)	Incremental cost (yen)	ICER (yen/QALY)	INMB (yen/10 years)
Non-intervention model	8.13		3,022,498			
Intervention model	8.62	0.49	203,666	–2,818,833	–5,736,166	5,275,903

QALY, quality-adjusted life year; ICER, incremental cost-effectiveness ratio; INMB, incremental net monetary benefit

dementia with care support levels 1 and 2 were referenced from the value of noncertified long-term care needs from a previous study[26].

Analysis

The individual cohorts for the analysis were generated via simulation to calculate the base case of the ICER. The sample size of the cohorts for the simulation was set at 1000 individuals, following a previous study[15,38]; the result was said to be unchanged by the sample size[28]. The individual cohorts were assumed to follow the parameters of the basic values shown in Table 1. The base case of the ICER was calculated by inputting the basic values for the model parameters and analyzing them using Markov cohort analysis. The mean cumulative 10-year costs and QALYs were calculated based on a cohort analysis of 1000 individuals. These results were then used to calculate the base case of the ICER.

Next, a one-way sensitivity analysis was conducted to scale the size of the impact of the parameters on the base case and investigate the direction of them by changing the parameters within the ranges between the low values and the high values shown in Table 1. To avoid confusion when the ICER showed negative values, instead of reporting ICERs, INMB[39] was reported to consider the effect of the uncertainty of the parameters on the base case. If the values of the parameters increased with the increased values of the INMBs, the associations were considered to have improved; otherwise, they were considered to have worsened. The proportion of the size of impact (%) was calculated to measure how much of the total uncertainty of the base case of the INMB was represented by the specified parameters by dividing the squared spread value by the total sum of each parameter's squared spread value of INMBs.

Finally, a set of Monte Carlo probabilistic simulations for 1000 individuals and microsimulations with 1000 trials were conducted to check the robustness of the base case. The simulated results were plotted, and a 95% confidence ellipse and the willingness to pay (WTP) line were drawn; the plots on the WTP line were equal to 5,000,000 yen; therefore, the scatter plots and the area of the 95% confidence ellipse under the WTP line were interpreted as cost-effective ICER values.

In conducting the sensitivity analyses, the parameters were estimated to follow the distribution shown in Table 1. The range of the discount rate was set as 0.00%–4.00%. The low value and the high value of transition probability from well to death were set as 0.73% and 1.92%, respectively.

Results

The base case of the ICER

Simulated trends of the numbers in each state among 1000 individuals for 10 years in the NIG and the IG are shown in Fig. 2A and 2B, respectively. The Markov cohort model simulated with the passage of 10 years the numbers in each state in the NIG and the IG, respectively; after 10 years, there were 690/849 in the state of well, 58/2 in the state of a mild level of dementia, 39/2 in the state of the middle level of dementia, 72/3 in the state of severe level of dementia, 37/4 in the state of dead with dementia, and 104/114 in the state of dead without developing dementia. After 10 years, the number in the state of MCI was 26 in the IG.

The cumulative QALYs and costs for 10 years in the NIG/IG were 8.13 QALY/8.62 QALY, and 3,022,498 yen/203,666 yen, respectively. As such, the incremental effectiveness, incremental costs, ICER (base case), and INMB were estimated as 0.49 QALY, –2,818,833 yen, –5,736,166 yen (US$–57400)/QALY, and 5,275,903 yen, respectively (Table 2).

One-way sensitivity analysis

The ICER showed negative values; therefore, the INMB was analyzed. The most impactive parameter on the base case of INMB was the transition probability from well to dementia (risk percentage of 97.44%); the direction of the impact was improved (Table 3). The second-highest parameter was the transition probability from well to MCI (0.97%); the direction of the impact was worsened. The least impactive parameters on the base case ranked from 21 to 29 were the costs of the preventive programs (Table 3).

Probabilistic sensitivity analysis

Figure 3 shows the scatter plots and the 95% confidence ellipse from the results of the Monte Carlo probabilistic simulation for 1000 individuals and microsimulation with 1000 trials. No plots (0%) were located in the upper area of the WTP line; a large area of the 95% confidence ellipse was under the WTP line. The mean (95% confidence interval) of the simulated ICER was –6,874,807 (–14,075,713 and –3,729,394) yen/QALY.

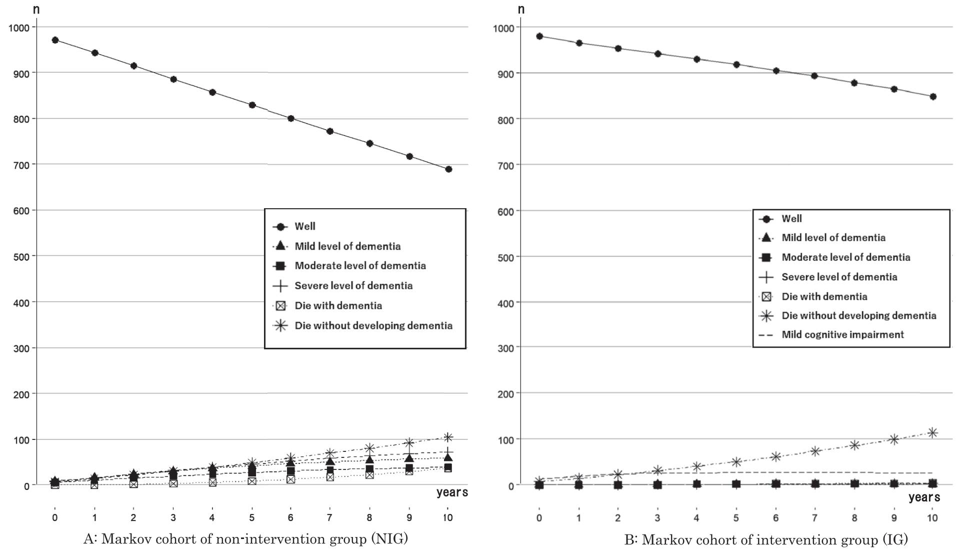


Fig. 2. Ten years simulated trends of the number of states among 1000 individuals
NIG, non-intervention group; IG, intervention group

Discussion

This study attempted to estimate the cost-effectiveness of combined physical and cognitive programs designed to prevent community-dwelling healthy young-old adults from developing dementia using a Markov cohort modeling simulation.

The Markov cohort modeling revealed that the base case of the ICER for 10 years was less than 5,000,000 yen/QALY, and furthermore, the value was negative. This supports the idea that introducing prevention programs might have much better cost-effectiveness than that of nonintervention from a public healthcare and long-term care payer's perspective. Furthermore, the INMB showed a positive value of 5,275,903 yen; this indicates that the prevention program was also cost-effective from a restricted societal perspective.

A previous study revealed that dementia prevention intervention was cost-effective over a short period[40], whereas our study suggested that the program might be cost-effective over longer periods.

Although our study showed good cost-effectiveness, a previous study[41] that investigated the cost-effectiveness of an exercise program for older adults who had already developed mild to moderate dementia reported that the program was not cost-effective. Two reasons might be considered for this. First, adults who develop dementia lose their chance to recover to a healthy state, while adults with MCI still have a chance to revert to a healthy state. Therefore, older adults who receive a preventive exercise program have a better chance of improving their health and saving expenses for their future healthcare and long-term care. Our study results (Fig. 2) showed a trend of the numbers in the well state in the IG, demonstrating a lower negative slope than that of the NIG. Moreover, the trends in the number of states concerning dementia showed horizontal slopes in comparison with the NIG, which had positive slopes. As such, the preventive exercise program for older adults in healthy and MCI states might result in more health-related utility gain and a reduction in expenses concerning the development of dementia than nonintervention. In other words, the suggested introduction of preventive programs to young-old adults in a healthy state or MCI state might not only be effective[7], but also economically beneficial for the health system and in the long term for Japan. Second, an intervention program for adults with dementia might cost more than a preventive program. Therefore, the program for adults with dementia would need to provide an individually tailored frequent exercise program, with 20–30 minutes of exercise at least five times per week[42]; our proposed program could be provided collectively and less frequently, at most once per week. As our results indicate, the costs for the preventive programs showed the lowest impact, 0% of the size of the impacts on the INMB (Table 3). Hence, our study supported the idea that collective exercise for dementia prevention programs represented a lower cost[43,44] than individual exercise programs for adults with dementia[41].

According to the results of the one-way sensitivity analysis, there was no parameter range indicating the effect of ICER values exceeding 5,000,000 yen/QALY. Furthermore, the probabilistic sensitivity analysis showed that no simulated values (0%) were over 5,000,000 yen/QALY; certainly, the mean value and the range were not over 5,000,000 yen/QALY. Combined, the sensitivity analysis suggested that the base case had high robustness.

To consider the most cost-effective way to deliver a preventive program to the community, interpreting the results of the one-way sensitivity analysis might be useful. First, the transition probability from well to dementia is worth paying attention to because of its notable impact on

Table 3. Results of one-sensitivity analysis

Parameters	Unit of parameters	Low value of parameters	High value of parameters	Incremental cost, yen (Low value)	Incremental cost, yen (High value)	Incremental effectiveness, QALY (Low value)	Incremental effectiveness, QALY (High value)	Direction of impact on INMB	INMB, yen (Low value)	INMB, yen (High value)	Spread	Size of impact	Rank
Transition probability from well to dementia	%/year	1.44	14.91	−1,857,147	−15,097,902	0.33	2.53	Improved	3,508,401	27,725,220	24,216,819	97.44%	1
Transition probability from well to MCI	%/year	1.25	9.32	−2,850,307	−2,480,181	0.53	0.12	Worsened	5,482,745	3,070,710	2,412,034	0.97%	2
Utility of mild severity of dementia	QALY/year	0.00	1.00	−2,818,833	−2,818,833	0.68	0.37	Worsened	6,204,850	4,681,985	1,522,865	0.39%	3
Utility of severe severity of dementia	QALY/year	0.00	0.74	−2,818,833	−2,818,833	0.56	0.31	Worsened	5,633,536	4,375,007	1,258,529	0.26%	4
Discount rate	%/year	0.00	4.00	−3,169,289	−2,518,118	0.55	0.44	Worsened	5,936,051	4,710,194	1,225,857	0.25%	5
Probability to be severe level of dementia	%	29.92	44.88	−3,086,784	−2,548,013	0.44	0.54	Improved	4,737,488	5,808,704	1,071,216	0.19%	6
Utility of moderate severity of dementia	QALY/year	0.00	1.00	−2,818,833	−2,818,833	0.57	0.37	Worsened	5,682,240	4,691,174	991,066	0.16%	7
Utility of MCI	QALY/year	0.16	1.00	−2,818,833	−2,818,833	0.39	0.57	Improved	4,786,017	5,649,672	863,655	0.12%	8
Probability to be mild level of dementia	%	30.80	46.20	−2,954,314	−2,681,953	0.46	0.52	Improved	4,988,577	5,560,434	571,857	0.05%	9
Probability to be middle level of dementia	%	19.28	28.92	−2,676,455	−2,960,412	0.46	0.52	Improved	5,001,310	5,549,048	547,738	0.05%	10
Cost for stay in community-based facility for dementia	yen/year	2,759,040	4,138,560	−2,583,898	−3,053,767	0.49	0.49	Improved	5,040,968	5,510,837	469,869	0.04%	11
Transition probability from well to die	%/year	0.73	1.92	−2,917,401	−2,538,425	0.49	0.48	Worsened	5,375,925	4,931,448	444,477	0.03%	12
Transition probability of reversing from MCI to well	%/year	20.90	57.00	−2,793,731	−2,838,896	0.46	0.52	Improved	5,083,098	5,416,020	332,922	0.02%	13
Effect of combined physical and cognitive exercises to prevent the progression of MCI	rate/year	0.55	0.76	−2,833,451	−2,803,828	0.51	0.47	Worsened	5,371,929	5,177,414	194,515	0.01%	14
Transition probability from MCI to dementia	%/year	1.14	5.90	−2,885,101	−2,773,405	0.50	0.49	Worsened	5,386,981	5,199,350	187,631	0.01%	15

(*Continued*)

Table 3. (*Continued*)

Parameters	Unit of parameters	Low value of parameters	High value of parameters	Incremental cost, yen (Low value)	Incremental cost, yen (High value)	Incremental effectiveness, QALY (Low value)	Incremental effectiveness, QALY (High value)	Direction of impact on INMB	INMB, yen (Low value)	INMB, yen (High value)	Spread	Size of impact	Rank
Cost for day service for dementia	yen/year	1,224,000	1,836,000	−2,727,113	−2,910,552	0.49	0.49	Improved	5,184,183	5,367,622	183,439	0.01%	16
Cost for short stay use for dementia	yen/year	1,052,160	1,578,240	−2,739,990	−2,897,675	0.49	0.49	Improved	5,197,060	5,354,746	157,686	0.00%	17
Cost for preventive day service for dementia	yen/year	491,520	737,280	−2,793,882	−2,843,783	0.49	0.49	Improved	5,250,952	5,300,853	49,901	0.00%	18
Cost for preventive short stay use for dementia	yen/year	377,280	565,920	−2,799,681	−2,837,984	0.49	0.49	Improved	5,256,751	5,295,054	38,303	0.00%	19
Effect of combined physical and cognitive exercises to prevent the development of dementia	rate/year	0.76	0.97	−2,829,313	−2,807,434	0.49	0.49	Worsened	5,293,468	5,256,792	36,675	0.00%	20
Cost for screening for cognitive impairment	yen	4,800	7,200	−2,829,254	−2,808,411	0.49	0.49	Worsened	5,286,324	5,265,481	20,842	0.00%	21
Cost for maintaining the program	yen/year	4,800	7,200	−2,828,077	−2,809,588	0.49	0.49	Worsened	5,285,147	5,266,658	18,489	0.00%	22
Number of individuals in a session	n/session	12	18	−2,814,810	−2,821,231	0.49	0.49	Improved	5,271,880	5,278,301	6,421	0.00%	23
Number of physical therapists (PTs) for a session	n/session	1	3	−2,820,871	−2,816,794	0.49	0.49	Worsened	5,277,941	5,273,864	4,076	0.00%	24
Cost of a physical therapist for a session	yen/session	6,716	10,075	−2,819,648	−2,818,017	0.49	0.49	Worsened	5,276,718	5,275,087	1,631	0.00%	25
Cost for a public space to conduct the session	yen/session	720	1,080	−2,819,488	−2,818,177	0.49	0.49	Worsened	5,276,558	5,275,247	1,311	0.00%	26
Cost for instructors for a session	yen/session	2,400	3,600	−2,818,833	−2,818,833	0.49	0.49	Improved	5,275,903	5,275,903	0	0.00%	27
Number of instructors for a session	instructors/session	4	6	−2,818,833	−2,818,833	0.49	0.49	Improved	5,275,903	5,275,903	0	0.00%	28
Numbers of sessions held in a year	sessions/year	32	48	−2,818,833	−2,818,833	0.49	0.49	Improved	5,275,903	5,275,903	0	0.00%	29

PTs, physical therapists; MCI, mild cognitive impairment; QALY, quality-adjusted life years; INMB, incremental net monetary benefit

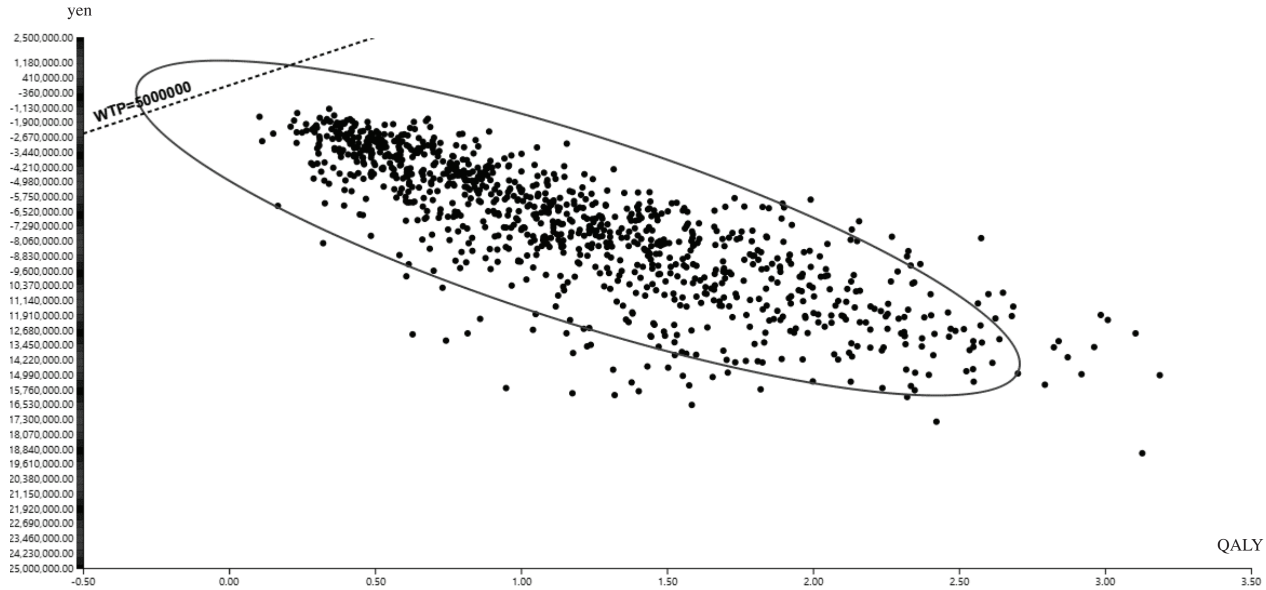


Fig. 3. Scatter plots and 95% confidence eclipse from the results of the Monte Carlo probabilistic simulation for 1000 individuals and microsimulation with 1000 trials

QALY, quality adjusted life years; WTP, willingness to pay

INMB. The results indicate that a higher transition probability from well to dementia improved INMB. This suggests that early effective screening to find community-dwelling young-old adults who might have a high risk of developing dementia might contribute to the impactive improvement of the cost-effectiveness of the preventive exercise program. Second, the transition probability from well to MCI might be worth paying attention to because the results indicate that a higher transition probability from well to MCI worsened INMB. Thus far, there has been little evidence that older adults can be prevented from developing dementia without diet changes, exercise, and cognitive stimulation[8]. Hence, our study suggested that early intervention, including exercise programs, such as the combined physical and cognitive exercises programs designed for preventing dementia in community-dwelling healthy young-old adults before they progress to the MCI state, might increase the cost-effectiveness in the community from a public healthcare and long-term care perspective.

Limitations

This study has three main limitations. First, we classified the severity of dementia into care needs levels of long-term care with specialist opinion. Since the classification system has been revised, it was not possible to make a simple comparison between a previous study[16] and our study; however, our classification had the possibility of overestimating the cost of long-term care services. The one-way sensitivity analysis confirmed that when the costs related to long-term care services decreased, the INMB value became small, which meant that the cost-effectiveness worsened. However, even if the cost for a stay in a community-based facility for dementia service, which was the most expensive and had the largest impact on INMB value among long-term care services, decreased in the range of 20%, the INMB value was still 5,040,968 yen, and the impact was as small as 0.04% (Table 3). Our conclusion therefore might not be significantly changed. Further research is needed to clarify this.

Second, proxy-rated utility instead of self-rated utility was used to define the effectiveness values. The self-rated values among people with dementia were reported to be higher than the proxy-rated values[27]. According to our one-way sensitivity analysis, parameters related to utilities of the state of dementia induced a worsening impact on INMB, which indicate that the cost-effectiveness of the preventive program might be worsened. However, we found that the size of the impact was small, ranging from 0.16% to 0.39%; hence, our conclusion might not be changed even if the self-rated utility was used in the model.

Finally, this study used the data from Japan as a priority and was conducted with assumptions about the background of Japanese healthcare and the long-term care system. Other international perspectives might be necessary to adapt our results to other countries.

Conclusion

The present results indicate that the combined physical and cognitive programs designed to prevent community-dwelling healthy young-old adults from developing dementia might have good cost-effectiveness with high robustness.

Acknowledgments: The work was funded by Health, Labour and Welfare Policy Research Grants (20CA20) by Ministry of Health, Labour and Welfare of Japan.

Conflict of Interest: The author reports no conflicts of interest in this work.

References

1) Beard JR, Officer AM, *et al.*: The World Report on Ageing and Health. Gerontologist. 2016; 56: S163–S166.
2) Cabinet Office, Government of Japan [Internet]. Annual Report on the Ageing Society [Summary] FY 2020 [cited 2021 Oct 1]. Available from: https://www8.cao.go.jp/kourei/english/annualreport/2020/pdf/2020.pdf (in Japanese)
3) Makizako H, Shimada H, *et al.*: Physical frailty and future costs of long-term care in older adults: results from the NCGG-SGS. Gerontology. 2021; 67: 695–704.
4) Okamoto Y: Health care for the elderly in Japan: medicine and welfare in an aging society facing a crisis in long term care. BMJ. 1992; 305: 403–405.
5) Sado M, Ninomiya A, *et al.*: The estimated cost of dementia in Japan, the most aged society in the world. PLoS One. 2018; 13: e0206508.
6) Toman J, Klímová B, *et al.*: Multidomain lifestyle intervention strategies for the delay of cognitive impairment in healthy aging. Nutrients. 2018; 10: 1560.
7) Shimada H, Makizako H, *et al.*: Conversion and reversion rates in Japanese older people with mild cognitive impairment. J Am Med Dir Assoc. 2017; 18: 808.e1–808.e6.
8) Sanford AM: Mild cognitive impairment. Clin Geriatr Med. 2017; 33: 325–337.
9) Jongsiriyanyong S, Limpawattana P: Mild cognitive impairment in clinical practice: a review article. Am J Alzheimers Dis Other Demen. 2018; 33: 500–507.
10) Suzuki T, Shimada H, *et al.*: A randomized controlled trial of multicomponent exercise in older adults with mild cognitive impairment. PloS One. 2013; 8: e61483.
11) Shimada H, Makizako H, *et al.*: Effects of combined physical and cognitive exercises on cognition and mobility in patients with mild cognitive impairment: a randomized clinical trial. J Am Med Dir Assoc. 2018; 19: 584–591.
12) Sonnenberg FA, Beck JR: Markov models in medical decision making: a practical guide. Med Decis Making. 1993; 13: 322–338.
13) Stollenwerk B, Bartmus T, *et al.*: Cost-effectiveness of hip protector use on a geriatric ward in Germany: a Markov model. Osteoporos Int. 2015; 26: 1367–1379.
14) Kato G, Kurachi Y: Cost-effective analysis of exercise programs designed for fall prevention among healthy younger old community-dwelling adults. Physical Therapy Japan. 2020; 47: 420–430. (in Japanese)
15) Handels R, Wimo A: Challenges and recommendations for the health-economic evaluation of primary prevention programmes for dementia. Aging Ment Health. 2019; 23: 53–59.
16) Ikeda S, Yamada Y, *et al.*: Economic evaluation of Donepezil treatment for Alzheimer's disease in Japan. Iryo To Shakai. 2000; 10: 27–38. (in Japanese)
17) Shiroiwa T, Fukuda T, *et al.*: Development of an official guideline for the economic evaluation of drugs/medical devices in Japan. Value Health. 2017; 20: 372–378.
18) Ninomiya, T[Internet]. Research on future estimation of the elderly population with dementia in Japan 2014 Summary / Shared Research Report (Nihonniokeru ninchishono koreishajinkono shoraisuikeinikansurukenkyu heisei26nenndo soukatu • bunntankenkyu hokokusho). [cited 2021 Jul 5]. Available from: https://mhlw-grants.niph.go.jp/system/files/2014/141031/201405037A/201405037A0001.pdf (in Japanese)
19) Asada T[Internet]. Prevalence of dementia in urban areas and coping with dementia living dysfunction (Toshibuniokeru ninnchishoyubyorituto ninnchishono seikatukinoshogaiheno taio). [cited 2021 Jul 5]. Available from: http://www.tsukuba-psychiatry.com/wp-content/uploads/2013/06/H24Report_Part1.pdf (in Japanese)
20) July J, Pranata R: Prevalence of dementia and its impact on mortality in patients with coronavirus disease 2019: a systematic review and meta-analysis. Geriatr Gerontol Int. 2021; 21: 172–177.
21) Blondell SJ, Hammersley-Mather R, *et al.*: Does physical activity prevent cognitive decline and dementia? a systematic review and meta-analysis of longitudinal studies. BMC Public Health. 2014; 14: 510.
22) Landeiro F, Mughal S, *et al.*: Health-related quality of life in people with predementia Alzheimer's disease, mild cognitive impairment or dementia measured with preference-based instruments: a systematic literature review. Alzheimers Res Ther. 2020; 12: 154.
23) Ministry of Health, Labour and Welfare[Internet]. Basic statistical survey of wage structure in FY2020 (Reiwagannnenn chinnginn kozo kihonn tokei chosa). [cited 2021 Jul 5]. Available from: https://www.e-stat.go.jp/stat-search/file-download?statInfId=000031919933&fileKind=0 (in Japanese)
24) Obu city [Internet]. Cost for a public space of Obu (Obu kominkan shiyo ryo) [cited 2021 July 5]. Available from: https://www.city.obu.aichi.jp/bunka/manabi/kouminkan_shisetsu/1006956/1006958.html (in Japanese)
25) Ministry of Health, Labour and Welfare [Internet]. Overview of actual statistics of long-term care benefit costs in FY2018 (Heisei30nenndo kaigokyufu tojittai tokei no gaikyo) [cited 2021 July 5]. Available from: https://www.mhlw.go.jp/toukei/saikin/hw/kaigo/kyufu/18/dl/03.pdf (in Japanese)
26) Tomata Y, Tsuji I, *et al.*: Prediction of future cost savings in long-term care and medical care if Japan achieves the health expectancy target of Health Japan 21 (second term). Nihon Koshu Eisei Zasshi (JAPANESE JOURNAL OF PUBLIC HEALTH). 2014; 61: 679–685. (in Japanese)
27) Shiroiwa T, Sung Y-K, *et al.*: International survey on willingness-to-pay (WTP) for one additional QALY gained: what is the threshold of cost effectiveness? Health Econ. 2010; 19: 422–437.
28) Gray AM, Clarke PM, *et al.*: Applied Methods of Cost-effectiveness Analysis in Healthcare. Oxford University Press, Oxford, 2010, pp. 282–283.
29) Iwagami M, Tamiya N: The long-term care insurance system in Japan: past, present, and future. JMA J. 2019; 2: 67–69.
30) Gidwani R, Russell LB: Estimating transition probabilities from published evidence: a tutorial for decision modelers. Pharmacoeconomics. 2020; 38: 1153–1164.
31) Ministry of Health, Labour and Welfare[Internet]. Overview of the simple life table for the first year of Reiwa (Ryō wa gan'nen kan'i seimei-hyō no gaikyō) [cited 2021 Jul 5]. Available from: https://www.mhlw.go.jp/toukei/saikin/hw/life/life19/index.html (in Japanese)
32) Sullivan PW, Slejko JF, *et al.*: Catalogue of EQ-5D scores for the United Kingdom. Med Decis Making. 2011; 31: 800–804.

33) Heßmann P, Seeberg G, *et al.*: Health-related quality of life in patients with Alzheimer's disease in different German Health Care Settings. J Alzheimers Dis. 2016; 51: 545–561.
34) Doi T, Shimada H, *et al.*: Multicomponent exercise to improve cognition in older adults with mild cognitive impairment. Research in Exercise Epidemiology. 2017; 19: 102–109. (in Japanese)
35) Folstein MF, Folstein SE, *et al.*: "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res. 1975; 12: 189–198.
36) Rosen WG, Mohs RC, *et al.*: A new rating scale for Alzheimer's disease. Am J Psychiatry. 1984; 141: 1356–1364.
37) Elwood RW: The Wechsler Memory Scale-Revised: psychometric characteristics and clinical application. Neuropsychol Rev. 1991; 2: 179–201.
38) Gulliford M, Charlton J, *et al.*: Social and material deprivation and the cost-effectiveness of an intervention to promote physical activity: cohort study and Markov model. J Public Health (Oxf). 2014; 36: 674–683.
39) Vreman RA, Geenen JW, *et al.*: The application and implications of novel deterministic sensitivity analysis methods. Pharmaco Economics. 2021; 39: 1–17.
40) McRae I, Zheng L, *et al.*: Cost-effectiveness of dementia prevention interventions. J Prev Alzheimers Dis. 2021; 8: 210–217.
41) Khan I, Petrou S, *et al.*: Does structured exercise improve cognitive impairment in people with mild to moderate Dementia? a cost-effectiveness analysis from a confirmatory randomised controlled trial: the Dementia and Physical Activity (DAPA) trial. Pharmacoecon Open. 2019; 3: 215–227.
42) D'Amico F, Rehill A, *et al.*: Cost-effectiveness of exercise as a therapy for behavioural and psychological symptoms of dementia within the EVIDEM-E randomised controlled trial. Int J Geriatr Psychiatry. 2016; 31: 656–665.
43) Blanchet S, Chikhi S, *et al.*: The benefits of physical activities on cognitive and mental health in healthy and pathological aging. Geriatr Psychol Neuropsychiatr Vieil. 2018; 16: 197–205.
44) Kramer A: An overview of the beneficial effects of exercise on health and performance. Adv Exp Med Biol. 2020; 1228: 3–22.

PHYSICAL THERAPY RESEARCH

SCIENTIFIC RESEARCH ARTICLE

Relationship between Muscle Flexibility and Characteristics of Muscle Contraction in Healthy Women during Different Menstrual Phases

Hiraku NAGAHORI, PT, MSc[1] and Nami SHIDA, PT, PhD[2]

[1)]*Department of Rehabilitation, Mitaka Clinic, Japan*
[2)]*Department of Physical Therapy, Tokyo Metropolitan University, Japan*

ABSTRACT. Objective: Skeletal muscle function is vital for preventing injury during exercise. It has been reported that skeletal muscle function fluctuates with the menstrual cycle and is considered one of the causes of injury. This study aimed to clarify the relationship between muscle flexibility and muscle contraction characteristics and their changes with the menstrual cycle. Methods: The subjects were healthy women who voluntarily participated in the study through recruitment posters. Muscle flexibility was measured with the passive knee extension (PKE) test, isokinetic knee flexor strength, and the maximum muscle strength exertion angle under two conditions of 60°/s and 120°/s in dominant hamstrings. Additionally, their correlations were analyzed and compared between the menstrual and ovulatory phases. Results: Sixteen subjects (mean age: 20.56 ± 0.73 years; body mass index: 20.21 ± 1.60) participated in the study. Correlation analysis showed a significant negative correlation between PKE and the maximum muscle strength exertion angle under the condition of 60°/s during the menstrual phase (r = –0.54; p = 0.03). No significant difference was observed in the two-group comparison of the variables measured during the menstrual and ovulatory phases. Conclusion: This study confirmed that the more flexible muscles generate the maximum strength at a more contracted position during the menstrual phase in women. In the future, it is necessary to examine the relationship between the results of this study and exercise performance and injury occurrence.

Key words: Muscle flexibility, Passive knee extension test, Muscle strength, Menstrual cycle

(Phys Ther Res 25: 68–74, 2022)

Skeletal muscle function and exercise performance in women are affected by the menstrual cycle. Anaerobic exercise performance may be reduced in the early follicular and late luteal phases than other phases of the menstrual cycles[1)]. Conversely, one study suggests that while exercise performance may be reduced in the early follicular phase, the overall effect is inconsequential[2)]. However, injury occurrence, in terms of the incidence of anterior cruciate ligament (ACL) injuries, was reportedly higher in the preovulatory[3–5)] and ovulatory phases[3–6)]. Additionally, muscle and tendon injuries are more common in the late follicular phase than in the early follicular or luteal phase. A cause of the changes in exercise performance and injury occurrence between menstrual cycles is attributed to the assumption that the skeletal muscle function varies during the menstrual cycle[1,7,8)].

Skeletal muscle is a crucial component of the motor system, and muscle flexibility and strength are essential for injury prevention[9–11)]. Previous studies suggest that muscle flexibility does not affect the maximal muscle strength but affects the maximal muscle strength exertion angle, one of the risk factors for muscle injury[12,13)]. However, these studies[12,13)] were conducted without separating men and women. It has been suggested that female's motor function can be affected by fluctuations in female hormones associated with the menstrual cycle[1,2)], and it is necessary to evaluate muscle function separately for each gender.

In other words, changes in skeletal muscle function during the menstrual cycle have been shown to affect female's exercise performance and injury occurrence. However, the

Received: January 26, 2022
Accepted: May 25, 2022
Advance Publication by J-STAGE: July 5, 2022
Correspondence to: Hiraku Nagahori. Department of Rehabilitation, Mitaka Clinic, 3-21-12, Shinkawa, Mitaka, Tokyo 181-0004, Japan
e-mail: hiraku0311@gmail.com
doi: 10.1298/ptr.E10173

relationship between muscle flexibility and muscle strength in women and how the menstrual cycle affects muscle flexibility and strength are not clear. While a previous study on muscle flexibility[14] reported that muscle flexibility is higher during ovulation than during menstruation. In contrast, another study said that muscle flexibility is not affected by the menstrual cycle or changes in female hormones[15]. Regarding muscle strength, it has been reported that the maximum hamstring muscle strength in the late luteal phase was significantly greater than that in the menstrual phase[16]. However, there are also reports that hamstring and quadriceps muscle strength did not differ between menstrual cycles[17,18].

Nevertheless, there is no consensus on the effects of the menstrual cycle and female hormones on skeletal muscle function. Only a few studies have reported on the relationship between muscle flexibility and muscle contractile properties, such as the maximal muscle strength and the angle of maximal muscle exertion in females. In this regard, we conducted this novel study to investigate the correlation between muscle flexibility and the maximum muscle strength and the angle of maximum muscle exertion in women. We also investigated whether muscle flexibility and contraction characteristics alter with the menstrual cycle, and if so, then how. We hypothesized that higher muscle flexibility would result in a smaller angle of maximum muscle exertion but that there would be no correlation with the maximum muscle strength. We also hypothesized that the menstrual cycle would produce significant differences in muscle flexibility and the maximal muscle strength. For this study, the "maximal muscle strength" and "joint angle at which the maximal muscle strength was exerted" (hereafter, maximal muscle strength exertion angle) were defined as the "muscle contraction characteristics."

Methods

Study design and participant eligibility

This was a repeated-measures design study conducted between September 2019 and August 2020. The inclusion criteria included (1) regular menstrual cycle; (2) no associated menstrual symptoms or gynecological diseases that would affect daily life; (3) no pregnancy; (4) no history of neurological diseases, acute trauma, pain at the measurement site, or previous knee joint injury on the measurement side; and (5) no intake of hormones or oral contraceptives in the past 6 months. All participants provided written informed consent for participation in this study. The experiment was conducted with three or more people in the same room to care for the subject during the measurement. This study was approved by the Research Ethics Committee of Tokyo Metropolitan University Arakawa Campus (Approval No. 19043).

Muscle flexibility (Fig. 1)

The flexibility of the hamstrings in the dominant leg was evaluated using the passive knee extension (PKE) test[19]. The dominant leg was defined as the leg to kick a ball maximally. The subject was placed in a supine position, the hip and knee joints on the measurement side were in 90° flexion, and the knee joint was extended altruistically until the angle at which resistance was first felt or the lumbar spine was no longer parallel to the bed. The angle between the extension from the femur and the lower leg was measured (Fig. 1). Thus, the smaller the angle, the greater the flexibility of the hamstrings. The measurements were taken twice, and the average value was adopted. The measurements were taken using an electronic angle meter (Ito easyangle; Ito Co., Saitama, Japan).

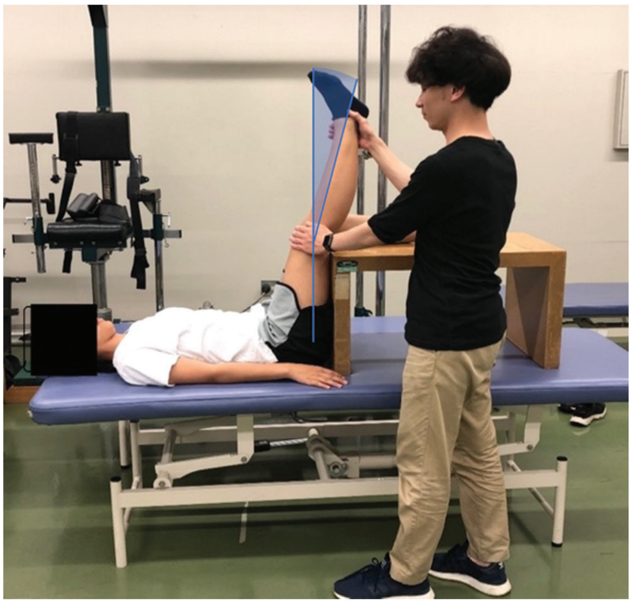


Fig. 1. The PKE test

The subject was positioned supine with the hip and knee joints on the measurement side in 90° flexion. The measurer extended the subject's knee joint altruistically until resistance was felt or the lumbar spine was no longer parallel to the bed. The angle between the extension from the femur and the lower leg was measured.

PKE, passive knee extension

Feature of muscle contraction

The dominant knee flexor strength was evaluated by calculating the torque-to-body weight ratios (Nm/kg). Additionally, the angle at which peak torque was obtained was measured as the maximal muscle strength exertion angle. These variables were assessed using Cybex NORM (Medica Inc., Osaka, Japan). For the measurement, the trunk was fixed to the seat with a belt, the upper limbs were gripped, the hip joint was set at 80° of flexion, and the range of motion (ROM) of the knee joint was set from 10° to 90° of flexion. The axis of motion was aligned with the height of the tibiofemoral joint, and the subjects were instructed to dorsiflex the ankle joint. Before the measurement, the subjects were made to pedal a bicycle ergometer (5 W, 5 min, 50 rpm) as a warm-up, followed promptly by muscle strength measurement. For muscle strength measurement, the subjects performed isokinetic knee joint flexion at 60°/s and 120°/s, and the maximum muscle strength and maximum muscle strength exertion angle were measured. The measurements were performed twice in each condition, and the average value was used for analysis.

Joint hypermobility

It has been suggested that ligament laxity fluctuates with the menstrual cycle[20], and joint hypermobility was assessed because it may affect the results of muscle flexibility tests. In this study, the joint hypermobility was assessed using the Beighton hypermobility score (BHS), which evaluates joint hypermobility in nine areas, including both knee joints, both elbow joints, both mother fingers, both little fingers, and the spinal column. We defined the subject with a score of ≥4 as positive following a previous study[21].

All the aforementioned measurements were performed by a single physical therapist, with information about the subject's menstrual phase withheld.

The measure of the menstrual cycle

The subjects were asked to undergo basal body temperature measurement, and their menstrual cycles were monitored 1–2 months before the measurement. Those whose menstrual cycle was between 25 and 36 days were considered to have a regular menstrual cycle, and those who did not fall into this category were excluded from the measurement. In a previous study[22], it has been recommended that the menstrual cycle be validated retrospectively when using the basal body temperature method. In another study[23] investigating changes in exercise performance with the menstrual cycle, the basal body temperature method was used to assess the subject's menstrual cycle. Therefore, in this study, we confirmed the validity of the menstrual cycle by retrospectively checking the basal body temperature of the participants after all measurements were completed.

Measurement phase

The measurements were taken during two phases: the menstrual phase (within 3 days of the start of menstruation) and the ovulatory phase (the last 2–3 days before the low-temperature phase and the 1st–2nd day of the high-temperature phase). The reasons for selecting these phases were as follows: (1) both phases are easy to recognize; (2) the secretion of estrogen and progesterone is low during the menstrual phase, indicating little influence of female hormones; and (3) there are reports that muscle strength is higher during the ovulatory phase than that during other menstrual phases[24] and hamstring flexibility is higher[17]. The subjects were asked to refrain from strenuous exercise that they do not usually do for several days before the measurement and were asked not to do any exercise that might affect muscle strength or flexibility, such as stretching, before the measurement.

Statistical analysis

The Shapiro–Wilk's normality test was performed for each measurement item's basic attribute. To evaluate the relationship between muscle flexibility and muscle contraction characteristics, Pearson's correlation coefficient or Spearman's rank correlation coefficient was performed for PKE value, the maximum knee flexor strength, and the maximal muscle strength exertion angle. Additionally, to evaluate the difference in the maximal maximum knee flexor strength and muscle flexibility according to the menstrual phase, a paired t-test or the Wilcoxon's signed-rank test was performed on the maximum knee flexor strength, PKE value, and BHS during the menstrual and ovulatory phases. The significance level was set at 5%, and the statistical analysis was performed using Statistical Package for the Social Sciences version 26 (IBM Corp., Armonk, NY, USA).

Table 1. Participants' basic characteristics (n = 16)

Variable	Mean ± SD
Age (year)	20.56 ± 0.73
Height (cm)	161.31 ± 4.79
Weight (kg)	52.75 ± 3.53
Measured leg (Rt/Lt)	15/1

SD, standard deviation; Rt, right; Lt, left

Results

Participants' basic characteristics

Eighteen female university students participated in the study, and two were excluded because of irregular menstruation confirmed using basal body temperature. Therefore, 16 participants (mean age, 20.56 ± 0.73 years; height, 161.31 ± 4.79 cm; and weight, 52.75 ± 3.53 kg) were included in the final study. The characteristics of the participants are listed in Table 1.

The comparison between the menstrual phases

The comparison between menstrual and ovulatory phases in terms of maximum knee flexor strength, PKE values, and the BHS showed no significant difference (Table 2).

The relationship between muscle flexibility and feature of muscle contraction (Table 3)

1) Muscle flexibility and maximal strength exertion angle
   A significant negative correlation was observed between PKE value and the maximal muscle strength exertion angle under the condition of 60°/s during menstruation ($r = -0.54$; $p = 0.03$), and no significant correlation was observed in other conditions (Table 3).
2) Muscle flexibility and maximum knee flexor strength
   No significant correlation was observed in all conditions (Table 3).

A previous study[14] reported that the effect size (Cohen's d) for significant differences in ROM between menstrual cycles was 0.89. Thus, referring to the study[14], we calculated power ($1 - \beta$) with d set at 0.89, an alpha level at 0.05, and a

Table 2. Results of passive knee extension, the maximum muscle strength exertion angle, knee flexor strength, and BHS by menstrual phases

		Menstrual cycle phases		p-value
		MP	OP	
PKE value (°)		32 ± 13$^{\alpha}$	32 ± 12$^{\alpha}$	0.99
MMSEA (°)	60°/s	53.53 ± 8.00$^{\alpha}$	48.88 ± 10.39$^{\alpha}$	–
	120°/s	51.16 ± 12.15$^{\alpha}$	46.00 (42.76–49.91)$^{\beta}$	–
KFS (Nm/kg)	60°/s	1.02 ± 0.16$^{\alpha}$	1.04 ± 0.18$^{\alpha}$	0.60
	120°/s	0.96 (0.75–0.98)$^{\beta}$	0.89 (0.75–0.99)$^{\beta}$	0.61
BHS		2.26 (2.00–3.25)$^{\beta}$	2.00 (1.75–3.00)$^{\beta}$	0.31

$^{\alpha}$Mean ± SD, $^{\beta}$Median (IQR)

PKE, passive knee extension; MMSEA, maximum muscle strength exertion angle; KFS, knee flexor strength; BHS, Beighton hypermobility score; MP, menstrual phase; OP, ovulatory phase; SD, standard deviation; IQR, interquartile range

Table 3. Relationship between muscle flexibility and feature of muscle contraction

			MMSEA (°)		KFS (Nm/kg)	
			r	p-value	r	p-value
PKE value	MP	60°/s	–0.54	0.03	0.28	0.28
		120°/s	0.21	0.44	0.32	0.23
	OP	60°/s	–0.09	0.75	–0.14	0.66
		120°/s	–0.06	0.81	0.10	0.72

PKE, passive knee extension; MP, menstrual phase; OP, ovulatory phase; MMSEA, maximum muscle strength exertion angle; KFS, knee flexor strength

sample size at 16. We also used G*Power (version 3.1.9.6 for macOS; Heine-Universität, Düsseldorf, Germany) to determine the power ($1 - \beta$).

Discussion

We investigated the correlation between muscle flexibility and muscle contraction characteristics in the young female hamstring. We also investigated whether there were significant differences in muscle flexibility and the maximal muscle strength between the menstrual and ovulatory phases of the menstrual cycle. We hypothesized that muscle flexibility is not correlated with the maximal muscle strength and is negatively correlated with the maximal muscle strength exertion angle. We also proposed that muscle flexibility and the maximal muscle strength will be higher during the ovulatory phase than during the menstrual phase. Based on our experiments, we found that a positive correlation existed only between muscle flexibility and the maximum muscle strength angle during the menstrual phase. Furthermore, there were no significant differences in muscle flexibility and the maximal muscle strength between the menstrual and ovulatory phases.

In this study, there was no significant difference between menstrual and ovulatory phases in terms of the maximum knee flexor strength, PKE values, and the BHS. Similar to this study, there have been reports of no significant differences in the maximal muscle strength between the menstrual and ovulatory phases[25,26]. Factors that have been reported to affect muscle strength include muscle activity[27], the relationship between muscle length and tension[28], and stiffness[29]. Estrogen, one of the female hormones secreted during ovulation, can decrease stiffness and reduce muscle strength[8,30]. On the other hand, it has been reported that estrogen can directly affect contractile properties, such as cross-bridging in muscle cells[29,31,32]. Estrogen is also thought to have a neuroexcitatory effect, which has been reported to reduce inhibition and increase spontaneous activation[33]. The neuroexcitatory effect results in improved muscle contractility[8]. Based on these findings, we hypothesize that there was no significant difference in the maximal muscle strength between the menstrual and ovulatory phases due to the opposing phenomenon of estrogen suppressing muscle weakness.

The female gonadocorticoid, estrogen, affects soft tissue collagen concentration, which is significantly higher in tendons than in the muscle tissue, and may be attributed to greater muscle flexibility in females[34]. Thus, estrogen is assumed to affect tendons more than muscles. PKE is used to evaluate muscle tightness, flexibility, or stiffness[19].

Accordingly, the PKE (which assesses muscle flexibility) is unlikely to be affected by estrogen levels. Thus, as hypothesized, estrogen secretion during ovulation has little effect on muscle flexibility, and there was no statistically significant difference in PKE between the menstrual and ovulatory phases.

Similarly, there was no significant change in joint laxity between menstrual cycles in previous studies[35–37]. Another study reported that the change in BHS scores between the menstrual and ovulatory phases was rather small[38]. However, since the BHS is a criterion score rather than a continuous variable, significant changes are only visible if the baseline ROM is near the criterion value[38]. In this study, we obtained comparable results regarding joint laxity, which did not change significantly between menstrual cycles, probably due to the characteristics of the measurement method.

While there were no significant differences in the maximal muscle strength, PKE, and BHS between menstrual and ovulatory cycles in this study, some studies have shown significant differences in muscle strength, muscle flexibility, and joint laxity between menstrual cycles[7,14,16,25,38,39]. Therefore, we believe that further studies are needed.

A small PKE value means flexible muscle. Therefore, in this study, there was a significant positive correlation between muscle flexibility and the maximal muscle strength exertion angle under the condition of 60°/s during menstruation. It has been reported that muscle flexibility is negatively correlated with muscle stiffness, which is essential for strength production in early muscle contraction[40–42]. Therefore, we believe that the muscle with higher flexibility, which is thought to have lower muscle stiffness, showed more maximum muscle strength in the contracted position. Additionally, this study results were different from those of previous reports[12,13], which showed that the maximal muscle strength exertion angle was smaller with higher flexibility. The cutoff value of PKE was 19.2[43], but the mean PKE value of the subjects in this study was 32.0° in both menstrual and ovulatory phases. However, previous reports showed that the hamstring flexibility of the subjects was higher than that of our subjects. In addition, both studies were tested using two-group comparison rather than correlation analysis. We suppose that the differences in the hamstring flexibility of the subjects and the statistical methods led to the different results.

No significant correlation was observed between muscle flexibility and the maximal muscle strength in both menstrual and ovulatory phases. This result was similar to that of previous reports[12,13]. In Agopyan's[44] study on hamstrings of female dancers, there was no relationship between muscle flexibility and the maximal muscle strength. Wan et al.[13] stated that while muscle flexibility contributes to preventing muscle damage during transverse stretching, muscle strength increases muscle stiffness and contributes to the prevention of muscle damage during centrifugal contraction. This suggests that muscle flexibility and the maximal muscle strength are independent skeletal muscle functions that need to be assessed and intervened individually in physical therapy and training situations.

It has been shown that the occurrence of musculoskeletal injuries differs according to the menstrual cycle[3–6,45,46]. Furthermore, it has been reported that the risk of ACL injury in women changes with the menstrual cycle[18,47,48]. Additionally, musculoskeletal injuries, such as ACL injuries and hamstring injuries, have been reported, regardless of the level of competition[49,50]. To conclude, we found that there were no significant differences in muscle flexibility and the maximum muscle strength between the menstrual and ovulatory phases in females. However, the correlation between muscle flexibility and the maximum muscle strength angle showed different results between the menstrual and ovulatory phases. This result indicates that the muscle contraction characteristics in women vary depending on the menstrual cycle, which suggests the need for considering the effects of the menstrual cycle when assessing exercise function from the perspective of injury prevention and exercise performance.

Limitations

There are several limitations to this study. First, selection bias may exist in this study because we recruited students from a specific university. For that reason, we should be careful about generalizing the result. Second, sex hormones were not measured at each measurement phase. Therefore, it is not possible to examine whether hormones affect skeletal muscle function. However, because the measurements were taken during the menstrual cycle as determined by the basal body temperature method, which is easy to utilize, the study is highly versatile and has clinical significance. Finally, we did not investigate the relationship between muscle function and the incidence of injury. The results suggest that there may be changes in skeletal muscle function between menstrual cycles. The study results may provide a basis for explaining the differences in injury that occur between menstrual cycles.

Conclusion

This study investigated the relationship between muscle flexibility and muscle contraction characteristics in the general female population and the effect of the menstrual cycle on these characteristics. The study results revealed that there is a positive correlation between muscle flexibility and the maximal muscle exertion angle during menstruation. The results suggest that skeletal muscle function in women may be affected by the menstrual cycle, in terms of not only muscle strength and flexibility but also the exertion angle.

Acknowledgments: The authors would like to thank K. Asano, M. Iwahori, and C. Kutsuma for their help in this study. They are also grateful to H. Kuruma for giving some advice to facilitate the investigation.

Conflict of Interest: There is no conflict of interest to be disclosed in this study.

References

1) Carmichael MA, Thomson RL, *et al.*: The impact of menstrual cycle phase on athletes' performance: a narrative review. Int J Environ Res Public Health. 2021; 18: 1667.
2) McNulty KL, Elliott-Sale KJ, *et al.*: The effects of menstrual cycle phase on exercise performance in eumenorrheic women: a systematic review and meta-analysis. Sports Med. 2020; 50: 1813–1827.
3) Adachi N, Nawata K, *et al.*: Relationship of the menstrual cycle phase to anterior cruciate ligament injuries in teenaged female athletes. Arch Orthop Trauma Surg. 2008; 128: 473–478.
4) Ruedl G, Ploner P, *et al.*: Are oral contraceptive use and menstrual cycle phase related to anterior cruciate ligament injury risk in female recreational skiers? Knee Surg Sports Traumatol Arthrosc. 2009; 17: 1065–1069.
5) Hewett TE, Zazulak BT, *et al.*: Effects of the menstrual cycle on anterior cruciate ligament injury risk: a systematic review. Am J Sports Med. 2007; 35: 659–668.
6) Wojtys EM, Huston LJ, *et al.*: The effect of the menstrual cycle on anterior cruciate ligament injuries in women as determined by hormone levels. Am J Sports Med. 2002; 30: 182–188.
7) Herzberg SD, Motu'apuaka ML, *et al.*: The effect of menstrual cycle and contraceptives on ACL injuries and laxity: a systematic review and meta-analysis. Orthop J Sports Med. 2017; 5: 2325967117718781.
8) Chidi-Ogbolu N, Baar K: Effect of estrogen on musculoskeletal performance and injury risk. Front Physiol. 2019; 9: 1834.
9) Ahmad CS, Redler LH, *et al.*: Evaluation and management of hamstring injuries. Am J Sports Med. 2013; 41: 2933–2947.
10) Padua DA, DiStefano LJ, *et al.*: National athletic trainers' association position statement: prevention of anterior cruciate ligament injury. J Athl Train. 2018; 53: 5–19.
11) Olsen OE, Myklebust G, *et al.*: Exercises to prevent lower limb injuries in youth sports: cluster randomised controlled trial. BMJ. 2005; 330: 449.
12) Alonso J, McHugh MP, *et al.*: Effect of hamstring flexibility on isometric knee flexion angle-torque relationship. Scand J Med Sci Sports. 2009; 19: 252–256.
13) Wan X, Qu F, *et al.*: Relationships among hamstring muscle optimal length and hamstring flexibility and strength. J Sport Health Sci. 2017; 6: 275–282.
14) Bell DR, Myrick MP, *et al.*: The effect of menstrual-cycle phase on hamstring extensibility and muscle stiffness. J Sport Rehabil. 2009; 18: 553–563.
15) Melegario SM, Simão R, *et al.*: The influence of the menstrual cycle on the flexibility in practitioners of gymnastics at fitness centers. Rev Bras Med Esporte. 2006; 12: 125–128. (in Portuguese)
16) Gordon D, Hughes F, *et al.*: The effects of menstrual cycle phase on the development of peak torque under isokinetic conditions. Isokinet Exerc Sci. 2013; 21: 285–291.
17) Janse de Jonge XAK, Boot CRL, *et al.*: The influence of menstrual cycle phase on skeletal muscle contractile characteristics in humans. J Physiol. 2001; 530: 161–166.
18) Watanabe K, Kouzaki M, *et al.*: Regional neuromuscular regulation within human rectus femoris muscle during gait in young and elderly men. J Biomech. 2016; 49: 19–25.
19) Gnat R, Kuszewski M, *et al.*: Reliability of the passive knee flexion and extension tests in healthy subjects. J Manipulative Physiol Ther. 2010; 33: 659–665.
20) Slauterbeck JR, Fuzie SF, *et al.*: The Menstrual Cycle, Sex Hormones, and Anterior Cruciate Ligament Injury. J Athl Train. 2002; 37: 275–278.
21) Beighton P, Solomon L, *et al.*: Articular mobility in an African population. Ann Rheum Dis. 1973; 32: 413–418.
22) Schmalenberger KM, Tauseef HA, *et al.*: How to study the menstrual cycle: practical tools and recommendations. Psychoneuroendocrinology. 2021; 123: 104895.
23) Šimić N, Ravlić A: Changes in basal body temperature and simple reaction times during the menstrual cycle. Arh Hig Rada Toksikol. 2013; 64: 99–106.
24) Sarwar R, Niclos BB, *et al.*: Changes in muscle strength, relaxation rate and fatiguability during the human menstrual cycle. J Physiol. 1996; 493: 267–272.
25) Miyazaki M, Maeda S: Changes in hamstring flexibility and muscle strength during the menstrual cycle in healthy young females. J Phys Ther Sci. 2022; 34: 92–98.
26) Fridén C, Hirschberg AL, *et al.*: Muscle strength and endurance do not significantly vary across 3 phases of the menstrual cycle in moderately active premenopausal women. Clin J Sport Med. 2003; 13: 238–241.
27) Fowles JR, Sale DG, *et al.*: Reduced strength after passive stretch of the human plantarflexors. J Appl Physiol. 2000; 89: 1179–1188.
28) Avela J, Kyröläinen H, *et al.*: Altered reflex sensitivity after repeated and prolonged passive muscle stretching. J Appl Physiol. 1999; 86: 1283–1291.
29) Ryan ED, Beck TW, *et al.*: Do practical durations of stretching alter muscle strength? a dose-response study. Med Sci Sports Exerc. 2008; 40: 1529–1537.
30) Bell DR, Blackburn JT, *et al.*: Estrogen and muscle stiffness have a negative relationship in females. Knee Surg Sports Traumatol Arthrosc. 2012; 20: 361–367.
31) Wattanapermpool J, Reiser PJ: Differential effects of ovariectomy on calcium activation of cardiac and soleus myofilaments. Am J Physiol. 1999; 277: H467–H473.
32) Lowe DA, Baltgalvis KA, *et al.*: Mechanisms behind estrogen's beneficial effect on muscle strength in females. Exerc Sport Sci Rev. 2010; 38: 61–67.
33) Ansdell P, Brownstein CG, *et al.*: Menstrual cycle-associated modulations in neuromuscular function and fatigability of the knee extensors in eumenorrheic women. J Appl Physiol. 2019; 126: 1701–1712.
34) Taş S, Aktaş D, *et al.*: Menstrual cycle does not affect the mechanical properties of muscle and tendon. Muscles Ligaments Tendons J. 2020; 10: 11–16.
35) Karageanes SJ, Blackburn K, *et al.*: The association of the menstrual cycle with the laxity of the anterior cruciate ligament in adolescent female athletes. Clin J Sport Med. 2000; 10: 162–168.
36) Hertel J, Williams NI, *et al.*: Neuromuscular performance and knee laxity do not change across the menstrual cycle in female athletes. Knee Surg Sports Traumatol Arthrosc. 2006; 14: 817–822.

37) Maruyama S, Yamazaki T, *et al.*: Relationship between anterior knee laxity and general joint laxity during the menstrual cycle. Orthop J Sports Med. 2021; 9: 2325967121993045.

38) Shultz SJ, Levine BJ, *et al.*: A comparison of cyclic variations in anterior knee laxity, genu recurvatum, and general joint laxity across the menstrual cycle. J Orthop Res. 2010; 28: 1411–1417.

39) Tenan MS, Peng YL, *et al.*: Menstrual cycle mediates vastus medialis and vastus medialis oblique muscle activity. Med Sci Sports Exerc. 2013; 45: 2151–2157.

40) Wilson GJ, Wood GA, *et al.*: The relationship between stiffness of the musculature and static flexibility: an alternative explanation for the occurrence of muscular injury. Int J Sports Med. 1991; 12: 403–407.

41) de Aquino CF, Gonçalves GGP, *et al.*: Analysis of the relation between flexibility and passive stiffness of the hamstrings. Rev Bras Med Esporte. 2006; 12: 195–200. (in Portuguese)

42) Blackburn JT, Padua DA, *et al.*: The relationships between active extensibility, and passive and active stiffness of the knee flexors. J Electromyogr Kinesiol. 2004; 14: 683–691.

43) Yıldırım MŞ, Tuna F, *et al.*: The cut-off values for the diagnosis of hamstring shortness and related factors. Balkan Med J. 2018; 35: 388–393.

44) Agopyan A: Hamstring flexibility is not correlated with isokinetic knee muscle strength in female modern dancers. Med Probl Perform Art. 2018; 33: 95–101.

45) Beynnon BD, Johnson RJ, *et al.*: The relationship between menstrual cycle phase and anterior cruciate ligament injury: a case-control study of recreational alpine skiers. Am J Sports Med. 2006; 34: 757–764.

46) Martin D, Timmins K, *et al.*: Injury incidence across the menstrual cycle in international footballers. Front Sports Act Living. 2021; 3: 616999.

47) Balachandar V, Marciniak JL, *et al.*: Effects of the menstrual cycle on lower-limb biomechanics, neuromuscular control, and anterior cruciate ligament injury risk: a systematic review. Muscles Ligaments Tendons J. 2017; 7: 136–146.

48) Montalvo AM, Schneider DK, *et al.*: Anterior cruciate ligament injury risk in sport: a systematic review and meta-analysis of injury incidence by sex and sport classification. J Athl Train. 2019; 54: 472–482.

49) Kuske B, Hamilton DF, *et al.*: Patterns of hamstring muscle tears in the general population: a systematic review. PLoS One. 2016; 11: e0152855.

50) Mei Y, Ao YF, *et al.*: Clinical characteristics of 4355 patients with anterior cruciate ligament injury. Chin Med J (Engl). 2013; 126: 4487–4492.

PHYSICAL THERAPY RESEARCH

SCIENTIFIC RESEARCH ARTICLE

Effect of Ward-dedicated Physical Therapy Staffing on Outcomes among General Medical Patients in an Acute Hospital: A Difference-in-difference Analysis

Kosho OHTA, PT[1], Kazuaki UDA, PT, MPH, PhD[2,3], Asumu FUJII, PT[1], Shingo MURANAGA, PT, PhD[1], So NAKAJI, MD[4] and Koichi MIYAKOSHI, MD[1]

[1)] *Department of Rehabilitation Medicine, Kameda Medical Center, Japan*
[2)] *Clinical Research Support Division, Kameda Institute for Health Science, Kameda University of Health Sciences, Japan*
[3)] *Department of Health Services Research, Faculty of Medicine, University of Tsukuba, Japan*
[4)] *Department of Gastroenterology, Kameda Medical Center, Japan*

ABSTRACT. Objective: The study aimed to examine the effect of dedicated physical therapy (PT) staffing on the outcomes of patients admitted to a general medical ward with acute cholangitis. Methods: This retrospective observational study was conducted in an 865-bed tertiary-care hospital in Japan. Patients with acute cholangitis between September 2015 and August 2017 were enrolled. Patients admitted to a ward with dedicated PT staffing were included in the dedicated group, while those admitted to a ward without dedicated PT staffing were included in the non-dedicated group. Each group was further divided into pre-dedicated and post-dedicated period based on September 1, 2016, at which PT staffing was implemented. The primary outcome was absolute functional gain (AFG), which was defined as the difference between Barthel index at discharge and that at admission. A difference-in-difference analysis was conducted to examine the changes in AFG associated with ward-dedicated PT staffing. Results: We identified 456 patients with acute cholangitis. Complete case analysis was applied, resulting in 252 patients in the final analysis. Patients were assigned to the dedicated group in the pre-dedicated period (n = 66) and post-dedicated period (n = 52), and to the non-dedicated group in the pre-dedicated period (n = 60) and post-dedicated period (n = 74). The adjusted difference-in-difference estimator was 17.1 (95% confidence interval: 5.6 to 28.5, p = 0.003) for AFG. Conclusion: Ward-dedicated PT staffing may improve the AFG of general medical patients in an acute hospital. Ward-dedicated PT staffing should be among the strategies utilized in the acute care process.

Key words: Health services administration, Rehabilitation, Activities of daily living

(Phys Ther Res 25: 75–83, 2022)

Patients in an acute hospital are at risk of disability, and hospitalization-associated disability occurs in approximately one-third of older patients[1,2)]. Thus, health professionals should evaluate and identify the rehabilitation needs of patients at a high risk for hospitalization-associated disability.

Previous studies have demonstrated the effectiveness of early mobilization programs[3–6)] and special care procedures (e.g., comprehensive geriatric assessment[7–10)] and acute care for elders[11,12)]) as ward-based interventions. In a retrospective pre-post study, Johnson et al.[13)] demonstrated that additional cardiovascular intensive care unit (CVICU)-dedicated physical therapy (PT) staffing results in increased PT treatment and reduced CVICU and post-CVICU length of stay (LOS). In a retrospective observational study, Engel et al.[14)] demonstrated that assigning physical therapists to the intensive care unit (ICU) increased the number of patients who received PT, shortened the time required prior to the initiation of PT, and decreased the ICU LOS. Livingstone et al.[15)] also reported that PT/occupational therapy staffing is associated with improved outcomes and overall quality of life among long-stay residents. Moreover, in a randomized control trial, Landefeld et al.[8)] reported that comprehensive geriatric

Received: January 11, 2022
Accepted: June 25, 2022
Advance Publication by J-STAGE: August 12, 2022
Correspondence to: Kosho Ohta. Department of Rehabilitation Medicine, Kameda Medical Center, 929, Higashi-cho, Kamogawa, Chiba 296-8602, Japan
e-mail: ota.kosho@kameda.jp
doi: 10.1298/ptr.E10168

assessment can improve basic functioning during activities of daily living (ADLs) and reduce the frequency of discharge to long-term care institutions.

Several studies have investigated the effectiveness and safety of early mobilization programs that involve special care procedures and dedicated PT staffing[3,5,7,16–19]. However, the effect of early mobilization programs and dedicated PT staffing on patient outcomes has been mainly investigated among ICU patients[13,14] and nursing home residents[15], while few studies have focused on patients in an acute general hospital. Therefore, the present study aimed to examine the effect of dedicated PT staffing among patients in a general medical ward at an acute hospital.

Methods

Ethical approval

The study protocol was reviewed and approved by Research Ethics Committee, Kameda Medical Center (No. 18-212). Informed consent was obtained in the form of opt-out on the web site because the data were obtained via a retrospective review.

Health insurance reimbursement program for maintaining or improving ADLs

In 2014, the Japanese Ministry of Health, Labor and Welfare established a new health insurance reimbursement program for maintaining or improving ADLs, which involves ward-dedicated rehabilitation therapy staffing. This program allows for preventive interventions for functional decline. The following conditions are required to receive reimbursement:

- The ward needs dedicated rehabilitation therapists.
- The dedicated therapists perform ADLs assessment, provide instructions to maintain or improve ADLs function to aid in the prevention of falls and pressure ulcers, provide information regarding patient status to family, and assist with early discharge planning.
- The dedicated therapists also share information related to the following with medical staff: current ADLs function, post-discharge risks, functional prognosis, activity level, and level of participation preferred by the patient.

In the hospital where the study was conducted, this insurance program was introduced in one general medical ward on September 1, 2016, but not in another general medical ward. Therefore, we had the opportunity to verify the effectiveness of the ward-dedicated PT staffing by using the difference-in-difference method with patients data from before and after the introduction of this insurance program in general medical wards with or without the ward-dedicated PT staffing.

Setting

This retrospective observational study was conducted in an 865-bed tertiary-care hospital in Japan. On September 1, 2016, two full-time, dedicated PT staff were assigned to one of the general medical wards where gastrointestinal and oncology patients were admitted (45-bed capacity) to work for 8 hrs/day and 7 days/week. Thus, one to two therapists were present in this ward each day of the week, including weekends.

We defined the patients in the general medical ward with dedicated PT staffing as the dedicated group and those in another general medical ward without dedicated PT staffing as the non-dedicated group. Patients in the non-dedicated group were gastrointestinal and oncology patients as in the dedicated group. Moreover, we defined the period before September 1, 2016 (the date when the dedicated PT staffing was implemented in our hospital) as the pre-dedicated period (year 1) and the period after September 1, 2016, as the post-dedicated period (year 2).

As a result, patients in the general medical wards where gastrointestinal and oncology patients were admitted were divided into 4 groups: those admitted to the general medical ward with dedicated PT staffing before the start of the staffing (the dedicated group in year 1), those admitted to the general medical ward with dedicated PT staffing after the start of the staffing(the dedicated group in year 2), those admitted to another general medical ward without dedicated PT staffing before the start of the staffing (the non-dedicated group in year 1), and those admitted to another general medical ward without dedicated PT staffing after the start of the staffing (the non-dedicated group in year 2).

The reason why we staffed ward-dedicated PT to the wards where gastrointestinal and oncology patients were admitted was because these patients were mainly older and sometimes orders of rehabilitation were missed despite the need for rehabilitation. Among them, we especially focused on patients with acute cholangitis, who are often older and have a period of rest due to initial treatment. Therefore, we included patients with acute cholangitis as participants.

Participants

We enrolled patients with acute cholangitis who were admitted to the general medical wards from September 2015 to August 2017. Patients with acute cholangitis were diagnosed and assessed for severity in accordance with the Tokyo guidelines[20]. Data were retrospectively obtained from April 2019 to July 2019 by two physical therapists. This study used the following criteria (Fig. 1):

Inclusion criteria

- Patients in a general medical ward with acute cholangitis.

Exclusion criteria

- Transferred from another ward after 14 days of hospitalization.
- Patients admitted in both pre-dedicated period (year 1) and post-dedicated period (year 2).
- Patients who had missing Barthel index (BI) data.

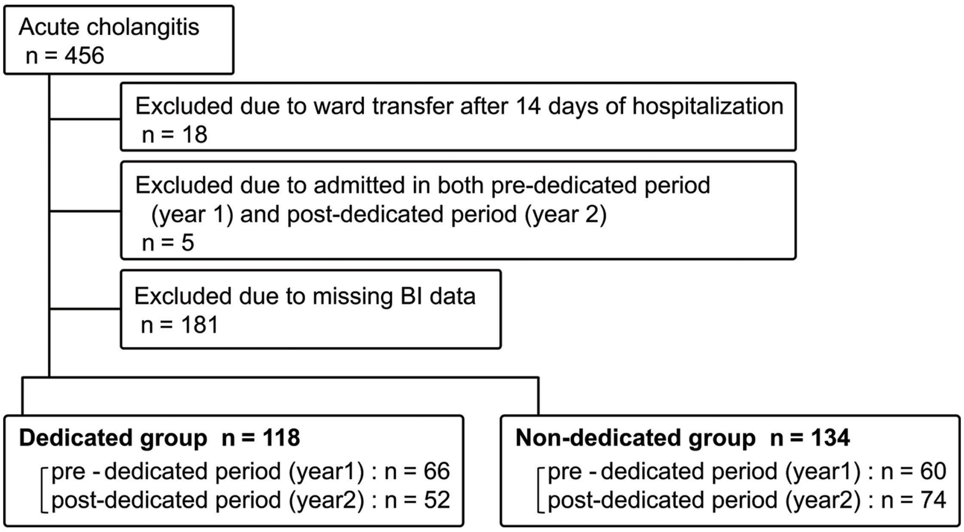


Fig. 1. Study inclusion and exclusion criteria

We excluded the patients admitted in both pre-dedicated period (year 1) and post-dedicated period (year 2) because the effect of the periods could not be isolated.

Study groups

The patients were divided into 4 groups: dedicated group in year 1, dedicated group in year 2, non-dedicated group in year 1, and non-dedicated group in year 2. The patients were assigned to either the dedicated or non-dedicated group using the hospital's bed management system. The decision was made independent of disease severity, etiology, and ADLs dependency.

Regular rehabilitation was conducted in four groups, and dedicated physical therapists were involved only in the dedicated group in year 2. A team of nine physical therapists were involved in the 4 groups, while two dedicated physical therapists were added and involved before regular rehabilitation in the dedicated group in year 2. The details on rehabilitation are described in the "Regular rehabilitation" and "Ward-dedicated PT staffing" sections.

Regular rehabilitation

Regular rehabilitation was delivered in all groups. Patients received regular rehabilitation if the gastroenterologist judged it necessary. The gastroenterologist's decision was based on ability, including the decline in ADLs. Therapists compared the patient's current status with the pre-admission status and provided rehabilitation to the patients to bridge the gap. As there was no clear protocol, rehabilitation was planned for each individual patient. For example, if the patient was able to walk before admission, the goal of rehabilitation was to restore the ability to walk. If we determined that the patient lacked muscle strength to achieve the goal, we provided strength exercises. Regarding the cancellation criteria for rehabilitation, the recommendations provided by The Japanese Association of Rehabilitation Medicine were used as a guide[21].

Ward-dedicated PT staffing

The ward-dedicated PT was implemented for the dedicated group in year 2. The dedicated physical therapists evaluated the patients' social information, grip strength, walking speed, Fried's frailty phenotype[22], and Short Physical Performance Battery scores[23]. The dedicated physical therapists assessed each patient's abilities and provided instruction on preventive self-exercise to reduce bed rest, maintain and improve ADLs function, and maintain appropriate physical activity. The dedicated physical therapists also delivered the following interventions: walking exercises, stair climbing exercises, muscle strength exercises, and sit-to-stand training for the aim of aerobic and resistance training. In addition, we practiced the requirements of the facility standards as described in the "Health insurance reimbursement program for maintaining or improving ADLs" section.

When the dedicated physical therapists intervention was not enough, the ward-dedicated physical therapists requested additional order for regular rehabilitation to the gastroenterologist. The criteria for requesting a regular rehabilitation were as follows and the decision was made comprehensively: decreased physical function, BI of 85 points or less (criteria for calculation of disuse syndrome rehabilitation fee), history of immobility, possibility of improvement in physical function, and disease trajectory.

Study variables

Data on patient characteristics were obtained from the electronic medical records, which included the following information: principal diagnosis, age, sex, body mass index (BMI), Charlson comorbidity index (CCI), etiology and severity of cholangitis, endoscopic retrograde cholangiopancreatography (ERCP)-related complications, numeric pain rating scale (NPRS) scores at admission, BI values, and the time prior to the initiation of rehabilitation. A gastroenterologist evaluated the BI data at admission and discharge.

Outcome measures and study covariates

The primary outcome was absolute functional gain (AFG), which was defined as the difference between the total BI at discharge and that at admission. The secondary outcomes were LOS and time prior to the initiation of rehabilitation, which was defined as the period between patient admission and the first clinical assessment by a rehabilitation therapist. Factors that were considered potential confounders included age, sex, BMI, CCI, etiology and severity of cholangitis, ERCP-related complications, and NPRS at admission.

Statistical analysis

Data analysis

A complete case analysis was applied, which resulted in 252 patients in the final analysis. Student's t-test (continuous variables) and Fisher's exact test (categorical variables) were used to perform pre/post comparisons in each group.

Difference-in-difference approach

We used the difference-in-difference method, which is a quasi-experimental study design[24–27], to investigate the effect of dedicated PT staffing on patient outcomes. This method can distinguish background trends in patient outcomes and the initiation of dedicated PT staffing. The quasi-experimental study design prevents the effects of all confounders on the dedicated and non-dedicated groups without the need to adjust for each confounder. Thus, the difference-in-difference method can reduce potential biases from unmeasured variables.

No specific changes were noted in the dedicated or non-dedicated group within the pre- or post-dedicated period. Hence, we compared effects before and after the implementation of dedicated PT staffing, assuming parallel trends and common shocks. Moreover, we identified the effect of dedicated PT staffing on patient outcomes based on the difference between pre-post time differences in the dedicated group and pre-post time differences in the non-dedicated group. We developed the following model for the difference-in-difference method based on linear regression:

$$y(\text{AFG}) = \varepsilon + \beta 1 \text{ dedicated group} + \beta 2 \text{ post-dedicated period} + \beta 3 \text{ dedicated group} * \text{post-dedicated period} + \beta X$$

where X refers to age, sex, BMI, CCI, etiology of cholangitis, severity of cholangitis, ERCP-related complications, or NPRS at admission. The β parameters are the regression coefficients associated with X, respectively, and ε is the random error component reflecting the difference between the observed and fitted linear relationship.

The effect of dedicated PT staffing on patient outcomes was denoted as the coefficient of interaction term during the pre- and post-dedicated periods ($\beta 3$ in the model). The level of statistical significance was set at $p < 0.05$. Anonymized data were analyzed using R version 3.6.0[28].

Results

Our analysis of electronic medical records revealed that 456 patients with acute cholangitis were admitted to the general medical wards. Those who were transferred from another ward after 14 days of hospitalization (n = 18), those who were admitted in both pre-dedicated period (year 1) and post-dedicated period (year 2) (n = 5), and those with missing BI data (n = 181) were excluded. Thus, a total of 252 patients were included in the final analysis (Fig. 1). Figure 2 shows the scheme of the study design. The dedicated and non-dedicated groups were divided into pre- and post-dedicated periods (n = 118 [pre: 66, post: 52] and n = 134 [pre: 60, post: 74], respectively). Table 1 shows a comparison of year 1 and year 2 within the dedicated and non-dedicated groups regarding patient demographic and clinical characteristics. There were no significant differences in age, sex, BMI, CCI, severity of cholangitis, etiology of cholangitis, ERCP-related complications, and NPRS between year 1 and year 2 within the dedicated and non-dedicated groups.

The adjusted outcome obtained using the difference-in-difference method is shown in Table 2 and Figure 3. Analysis of the interaction revealed that AFG had significantly improved in the Dedicated group (difference-in-differences estimates, 95% confidence interval [CI]: 17.1, 5.6 to 28.5) relative to the non-Dedicated group. However, we observed no significant difference in the time prior to the initiation of rehabilitation (difference-in-differences estimates, 95% CI: −0.4, −2.3 to 1.5) and LOS (difference-in-differences estimates, 95% CI: 1.7, −2.1 to 5.6) in the difference-in-difference analysis.

Discussion

The present study investigated the effects of ward-dedicated PT staffing on outcomes among general medical patients in an acute hospital, with adjustments for demographic and clinical characteristics. Our findings indicated that AFG increased in patients with acute cholangitis after the implementation of ward-dedicated PT. A subsequent difference-in-difference analysis to remove measured or unmeasured confounders (a limitation of previous studies) also indicated that ward-dedicated PT staffing was associated with an increase in AFG. Moreover, the increase in AFG occurred independent of the time at which rehabilitation was initiated. However, in the BI at the time of admission of the dedicated group, a decrease from 80.4 in the pre-dedicated period to 59.9 in the post-dedicated period was observed. This result indicates that the dedicated group may have been a patient population with more potential for improvement in AFG than the other groups in the second year.

Previous studies have reported that dedicated PT staffing reduces the length of time prior to the initiation of PT, increases PT treatment, and reduces ICU and post-ICU LOS[13,14]. Moreover, specific care procedures in the ward

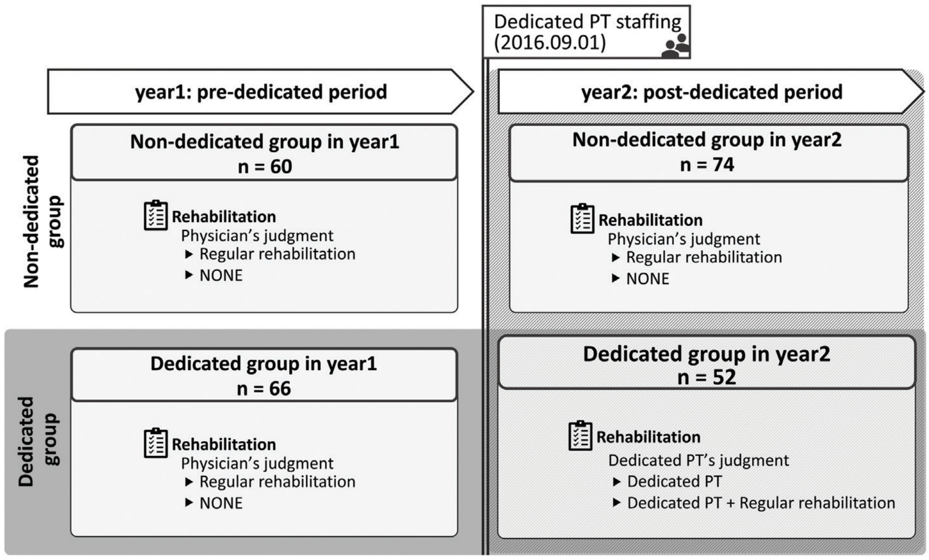


Fig. 2. Scheme of a study design to evaluate the effect of ward-dedicated PT staffing implementation
PT, physical therapy

Table 1. Patient demographic and clinical characteristics

Variable	Dedicated group (n = 118)			Non-dedicated group (n = 134)		
	Year 1 Pre-dedicated period (n = 66)	Year 2 Post-dedicated period (n = 52)	p value	Year 1 Pre-dedicated period (n = 60)	Year 2 Post-dedicated period (n = 74)	p value
Age, mean (SD), yr	75.2 (11.8)	77.0 (12.7)	0.42	74.0 (12.1)	75.9 (11.5)	0.56
Sex, male (%)	36 (54.5)	29 (55.8)	1.00	36 (60.0)	45 (60.8)	1.00
BMI, mean (SD), kg/m^2	22.2 (4.1)	22.4 (4.4)	0.80	23.2 (4.5)	23.2 (4.8)	0.90
CCI, n (%)			0.35			0.41
Score : 0–1	30 (45.5)	17 (32.7)		16 (26.7)	28 (37.8)	
Score : 2–3	20 (30.3)	21 (40.4)		29 (48.3)	31 (41.9)	
Score : ≥4	16 (24.2)	14 (26.9)		15 (25.0)	15 (20.3)	
Severity of cholangitis, (%)			0.92			0.74
Mild	40 (60.6)	30 (57.7)		39 (65.0)	43 (58.1)	
Moderate	22 (33.3)	18 (34.6)		18 (30.0)	26 (35.1)	
Severe	4 (6.1)	4 (7.7)		3 (5.0)	5 (6.8)	
Etiology of cholangitis, (%)			0.85			0.98
CBD stone	40 (60.6)	35 (67.3)		41 (68.3)	53 (71.6)	
Malignant obstruction	13 (19.7)	9 (17.3)		10 (16.7)	11 (14.9)	
Benign stricture	2 (3.0)	2 (3.8)		2 (3.3)	2 (2.7)	
Other	2 (3.0)	2 (3.8)		6 (10.0)	6 (8.1)	
Unknown	9 (13.6)	4 (7.7)		1 (1.7)	2 (2.7)	
ERCP-related complications, (%)			0.67			0.9
Bleeding after EST	1 (1.5)	2 (3.8)		2 (3.3)	3 (4.1)	
Perforation	0 (0.0)	1 (1.9)		0 (0.0)	1 (1.4)	
Post-ERCP pancreatitis	2 (3.0)	1 (1.9)		1 (1.2)	0 (0.0)	
Mortality	0 (0.0)	0 (0.0)		0 (0.0)	0 (0.0)	
ADLs at admission, mean (SD), BI	80.4 (32.7)	59.9(40.1)	0.003	71.4 (41.8)	68.6 (41.2)	0.69
Pain at admission, mean (SD), NPRS	1.3 (2.2)	1.1 (1.9)	0.72	1.0 (1.8)	1.1 (2.0)	0.8

ADLs, activities of daily living; BI, Barthel index; BMI, body mass index; CBD, common bile duct; CCI, Charlson comorbidity index; ERCP, endoscopic retrograde cholangiopancreatography; EST, endoscopic sphincterotomy; NPRS, numeric pain rating scale; SD, standard deviation

Table 2. Unadjusted and adjusted outcomes obtained using the difference-in-difference approach

Variable	Dedicated group: Year 1 Pre-dedicated period (n = 66)	Dedicated group: Year 2 Post-dedicated period (n = 52)	Non-dedicated group: Year 1 Pre-dedicated period (n = 60)	Non-dedicated group: Year 2 Post-dedicated period (n = 74)	Difference-in-differences estimates (95% CI)*	p value
AFG mean (SD), BI	0.15 (27.1)	17.1 (28.6)	1.9 (18.8)	3.5 (16.0)	17.1 (5.6 to 28.5)	0.003
Rehab initiation (SD), d	3.6 (3.0)	1.8 (1.2)	4.4 (3.4)	2.9 (2.2)	−0.4 (−2.3 to 1.5)	0.68
LOS mean (SD), d	14.1 (6.3)	16.4 (12.1)	13.3 (7.2)	13.6 (5.9)	1.7 (−2.1 to 5.6)	0.38

*These adjusted coefficient estimates were estimated from linear regressions that controlled for age, sex, BMI, CCI, severity of cholangitis, etiology of cholangitis, ERCP-related complication, and NPRS

AFG, absolute functional gain; BI, Barthel index; CI, confidence interval; rehab, rehabilitation; SD, standard deviation; LOS, length of stay; BMI, body mass index; CCI, Charlson comorbidity index; ERCP, endoscopic retrograde cholangiopancreatography

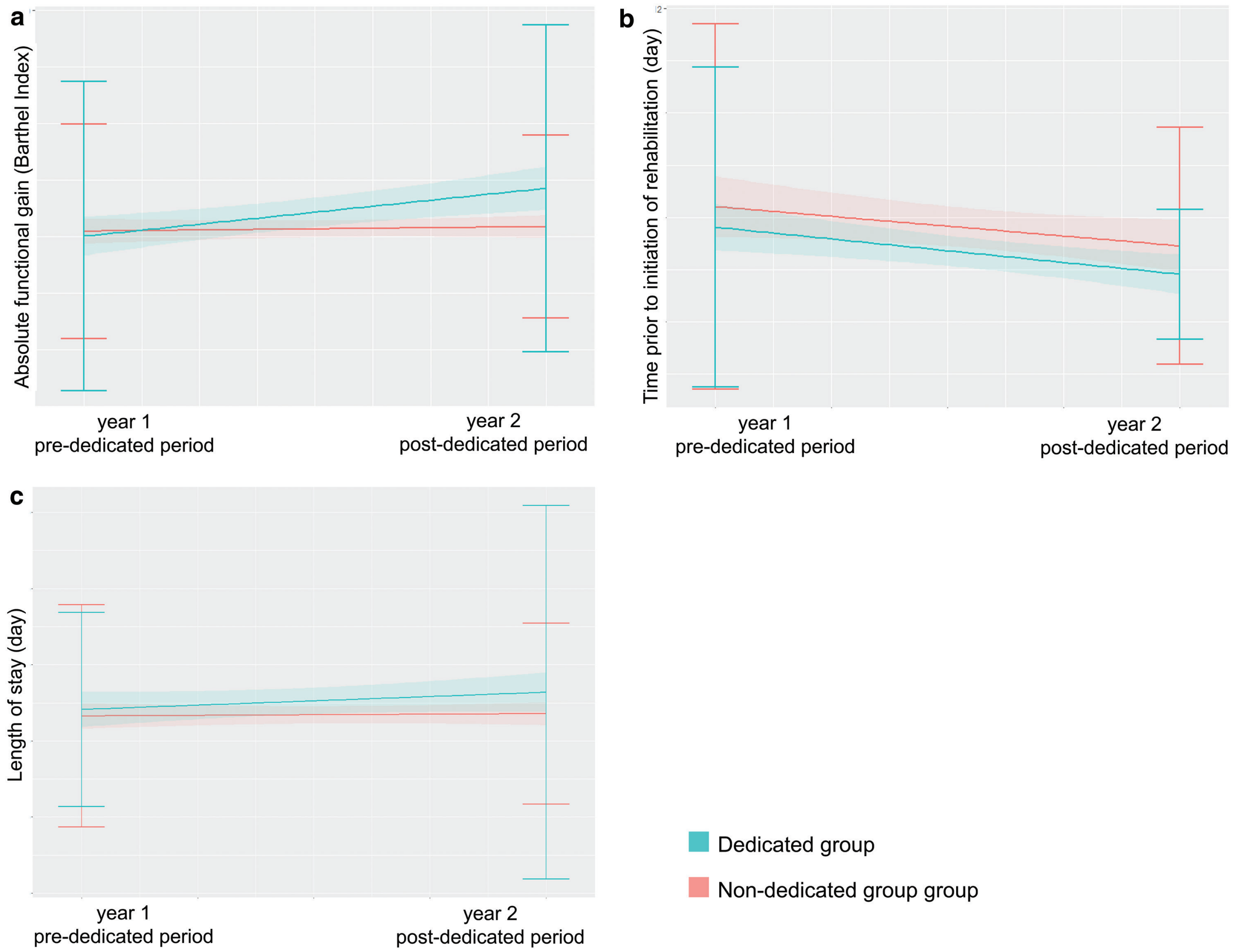


Fig. 3. Changes in outcomes before and after the implementation of ward-dedicated PT staffing in the dedicated and non-dedicated groups

(a) AFG (BI). (b) Time prior to initiation of rehabilitation (days). (c) LOS (days)

PT, physical therapy; AFG, absolute functional gain; BI, Barthel index; LOS, length of stay

may reduce short-term mortality, prevent falls, decrease delirium, and improve physical and cognitive functions[7–12]. However, particularly for patients in an acute general hospital, data to clearly show such benefits are lacking. In a retrospective pre-post study, Johnson et al.[13] demonstrated the effect of increasing the number of physical therapists from two to four in a CVICU setting. Engel et al.[14] also reported a positive effect of a physical therapist-led early mobilization program for ICU patients. However, because of the retrospective nature of the study, effects specific to the implementation of dedicated PT staffing could not be distinguished from the natural temporal trends in patient outcomes. In our study, we were able to exclude the effects of natural temporal trends and those of confounding factors using a quasi-experimental design.

Acute cholangitis is common in older patients, occurring primarily due to stones in the bile ducts and gallbladder[29,30]. Factors associated with the development of cholelithiasis include age, obesity, and physical inactivity[31,32]. In addition, early decreases in function are likely to occur in patients hospitalized with acute cholangitis due to symptoms (intermittent fever accompanied by chills, right upper quadrant pain, and jaundice) and treatments (antibiotic and biliary drainage). Thus, patients with acute cholangitis may exhibit frailty due to physical inactivity and may be at high risk of hospitalization-associated disability.

We hypothesized that the implementation of ward-dedicated PT staffing would decrease the time required prior to the initiation of rehabilitation and improve AFG. Based on unadjusted results, early rehabilitation assessment and AFG significantly improved following the implementation of ward-dedicated PT staffing. However, when the results were adjusted via the difference-in-difference analysis, we observed no significant difference in the time prior to the initiation of rehabilitation and LOS, whereas differences in AFG remained significant.

Our findings also indicated that ward-dedicated PT staffing significantly improved AFG without decreasing the length of the period prior to rehabilitation. However, implementation of dedicated PT staffing did not decrease the length of this period, which was unexpected. Nonetheless, there may be a clinical explanation for this result. The study hospital encouraged early initiation of rehabilitation for acute medical patients. Physicians performed ERCP in patients with moderate to severe cholangitis within 2 days of hospitalization, and rehabilitation was initiated on the day after ERCP. The time to the initiation of rehabilitation decreased in both the dedicated and non-dedicated groups after the implementation of ward-dedicated staffing. In addition, there was no decrease in the LOS reported in previous studies. In comparison to the CVICU population, the following factors may have affected the LOS in the general medical ward staffing in this study: medical needs (antibiotic therapy duration) and social needs (discharge coordination and introduction of long-term care services). Further studies with strict discharge criteria are needed to evaluate the effect of dedicated PT staffing on LOS.

In a prospective cohort study, Brown et al.[33] noted that low mobility and bed rest are common in hospitalized older patients and are important predictors of functional decline. Zisberg et al.[34] demonstrated that in-hospital mobility is an important modifiable factor related to immediate and short-term functional outcomes in older adults. In addition, mobilization programs have been shown to encourage out-of-bed ADLs[35] and prevent functional decline[36]. Lack of health literacy, which is defined as the ability to access, understand, evaluate, and use health information to make reasoned, health-related decisions[37], has been associated with physical inactivity and frailty among patients[38–40]. In our study, a dedicated physical therapist evaluated all patients before the initiation of regular rehabilitation, and inpatients with no functional decline at admission received instructions regarding self-exercise and appropriate physical activity. These care processes may have contributed to improvements in AFG by increasing health literacy.

Implications

In previous studies, acute inpatient rehabilitation typically focused on early mobilization and discharge planning[3,4,6,17–19,41–43], and the impact of dedicated PT staffing was mainly investigated in ICU or nursing home settings[13–15]. Our results suggest that ward-dedicated PT staffing is effective in improving AFG in general medical patients at an acute hospital. However, a 2017 survey revealed that only 2.4% (89 of 3748) of acute hospitals in Japan have ward-dedicated PT staffing. Our findings suggest that ward-dedicated PT staffing should be among the strategies utilized in the acute care process.

Study limitations

This study has several limitations. First, this study included a single-center hospital ward; thus, the generalizability of the methods and findings is limited. Therefore, further multi-center studies are required. Second, the ADLs evaluation was insufficient. BI data were missing, and a difference in BI at admission between each group was observed. In addition, the BI subitems were important in PT but could not be considered due to lack of data. Excluded patients may have had difficulty measuring BI data, and the present results may not be applicable to a population of patients with acute cholangitis for whom BI data would be difficult to measure. Moreover, as the baseline ADLs function (i.e., before admission) was unknown, discriminating between functional decline after hospitalization and baseline functional decline was difficult. However, ADLs function may always be uncertain given that it can be modified by several factors, such as the severity of disease and pain. Third, we could not have the data on daily rehabilitation time or duration of PT in the study because we could not have IDs that can be traced back to electronic medical record of the patients. Therefore, it is difficult to discuss the detail mechanism regarding the results of this study. Further studies are required to establish specific treatment protocols and to verify our findings in a large multi-center prospective cohort.

Conclusion

In the present observational study, we performed a difference-in-difference analysis of medical patients in an acute general hospital to elucidate the effect of ward-dedicated PT staffing. This quasi-experimental design allowed us to exclude the effects of natural temporal trends and confounding factors that limited previous studies. Our findings indicated that ward-dedicated PT staffing can improve AFG independent of the time at which rehabilitation is initiated in general medical patients admitted to an acute hospital.

Acknowledgments: The authors would like to thank all members of the general medical ward. They also thank Hiroki Matsui, MPH, for lending their expertise on the study design and Kimio Yamamoto, PT, for sharing dataset with them. Mihoko Takano, PT, and Yusuke Hirata, PT, performed the PT interventions described in this study and were instrumental to the execution of the ward-dedicated PT staffing. They are also indebted to Nao Hikota, PT, Kazuki Ikeda, PT, Tetsuya Shiraishi, PT, and Kaori Tokeshi, SLT, for their comments and suggestions.

Conflict of Interest: All authors have no competing interests to declare.

References

1) Covinsky KE, Pierluissi E, *et al.*: Hospitalization-associated disability: "she was probably able to ambulate, but I'm not sure". JAMA. 2011; 306: 1782–1793.
2) Sager MA, Rudberg MA, *et al.*: Hospital admission risk profile (HARP): identifying older patients at risk for functional decline following acute medical illness and hospitalization. J Am Geriatr Soc. 1996; 44: 251–257.
3) Schweickert WD, Pohlman MC, *et al.*: Early physical and occupational therapy in mechanically ventilated, critically-ill patients. Lancet. 2009; 373: 1874–1882.
4) Needham DM, Korupolu R, *et al.*: Early physical medicine and rehabilitation for patients with acute respiratory failure. Arch Phys Med Rehabil. 2010; 91: 536–542.
5) Adler J, Malone D: Early mobilization in the intensive care unit. Cardiopulm Phys Ther J. 2012; 23: 5–13.
6) Gruther W, Pieber K, *et al.*: Can early rehabilitation on the general ward after an intensive care unit stay reduce hospital length of stay in survivors of critical illness? Am J Phys Med Rehabil. 2017; 96: 607–615.
7) Stuck AE, Siu AL, *et al.*: Comprehensive geriatric assessment: a meta-analysis of controlled trials. Lancet. 1993; 342: 1032–1036.
8) Landefeld CS, Palmer RM, *et al.*: A randomized trial of care in a hospital medical unit especially designed to improve the functional outcomes of acutely-ill older patients. N Engl J Med. 1995; 332: 1338–1344.
9) Ellis G, Langhorne P: Comprehensive geriatric assessment for older hospital patients. Br Med Bull. 2005; 71: 45–59.
10) Ekerstad N, Dahlin Ivanoff S, *et al.*: Acute care of severely frail elderly patients in a CGA-unit is associated with less functional decline than conventional acute care. Clin Interv Aging. 2017; 12: 1239–1249.
11) Counsell SR, Holder CM, *et al.*: Effects of a multicomponent intervention on functional outcomes and process of care in hospitalized older patients: a randomized controlled trial of acute care for elders (ACE) in a community hospital. J Am Geriatr Soc. 2000; 48: 1572–1581.
12) Fox MT, Persaud M, *et al.*: Effectiveness of acute geriatric unit care using acute care for elders components: a systematic review and meta-analysis. J Am Geriatr Soc. 2012; 60: 2237–2245.
13) Johnson JK, Lohse B, *et al.*: Improving outcomes for critically ill cardiovascular patients through increased physical therapy staffing. Arch Phys Med Rehabil. 2019; 100: 270–277.E1.
14) Engel HJ, Tatebe S, *et al.*: Physical therapist-established intensive care unit early mobilization program: quality improvement project for critical care at the University of California San Francisco Medical Center. Phys Ther. 2013; 93: 975–985.
15) Livingstone I, Hefele J, *et al.*: The relationship between quality of care, physical therapy, and occupational therapy staffing levels in nursing homes in 4 years' follow-up. J Am Med Dir Assoc. 2019; 20: 462–469.
16) Siebens H, Aronow H, *et al.*: A randomized controlled trial of exercise to improve outcomes of acute hospitalization in older adults. J Am Geriatr Soc. 2000; 48: 1545–1546.
17) Morris PE, Goad A, *et al.*: Early intensive care unit mobility therapy in the treatment of acute respiratory failure. Crit Care Med. 2008; 36: 2238–2243.
18) Kress JP: Clinical trials of early mobilization of critically ill patients. Crit Care Med. 2009; 37(Suppl): S442–S447.
19) Hartley PJ, Keevil VL, *et al.*: Earlier physical therapy input is associated with a reduced length of hospital stay and reduced care needs on discharge in frail older inpatients. J Geriatr Phys Ther. 2019; 42: E7–E14.
20) Kiriyama S, Takada T, *et al.*: New diagnostic criteria and severity assessment of acute cholangitis in revised Tokyo guidelines. J Hepatobiliary Pancreat Sci. 2012; 19: 548–556.
21) The Japanese Association of Rehabilitation Medicine Clinical Practice Guidelines Committee: Guidelines for Safety Management and Promotion in Rehabilitation Medicine: 2nd Ed. Shindan To Chiryo Sha, 2018. (in Japanese)
22) Fried LP, Tangen CM, *et al.*: Frailty in older adults: evidence for a phenotype. J Gerontol. 2001; 56: M146–M156.
23) Guralnik JM, Ferrucci L, *et al.*: Lower-extremity function in persons over the age of 70 years as a predictor of subsequent disability. N Engl J Med. 1995; 332: 556–562.
24) Rajaram R, Chung JW, *et al.*: Association of the 2011ACGME resident duty hour reform with general surgery patient outcomes and with resident examination performance. JAMA. 2014; 312: 2374–2384.
25) Patel MS, Volpp KG, *et al.*: Association of the 2011ACGME resident duty hour reforms with mortality and readmissions among hospitalized Medicare patients. JAMA. 2014; 312: 2364–2373.
26) Dimick JB, Ryan AM: Methods for evaluating changes in health care policy: the difference-in-differences approach. JAMA. 2014; 312: 2401–2402.

27) Cataife G, Pagano MB: Difference-in-difference: simple tool, accurate results, causal effects. Transfusion. 2017; 57: 1113–1114.
28) Ihaka R, Gentleman R: R: a language for data analysis and graphics. J Comput Graph Stat. 1996; 5: 299–314.
29) Ahmed M: Acute cholangitis-an update. World J Gastrointest Pathophysiol. 2018; 9: 1–7.
30) Kimura Y, Takada T, *et al.*: Definitions, pathophysiology, and epidemiology of acute cholangitis and cholecystitis: Tokyo-guidelines. J Hepatobiliary Pancreat Surg. 2007; 14: 15–26.
31) Zimmer V, Lammert F: Acute Bacterial Cholangitis. Viszeralmedizin. 2015; 31: 166–172.
32) Lammert F, Gurusamy K, *et al.*: Gallstones. Nat Rev Dis Primers. 2016; 2: 16024.
33) Brown CJ, Friedkin RJ, *et al.*: Prevalence and outcomes of low mobility in hospitalized older patients. J Am Geriatr Soc. 2004; 52: 1263–1270.
34) Zisberg A, Shadmi E, *et al.*: Low mobility during hospitalization and functional decline in older adults. J Am Geriatr Soc. 2011; 59: 266–273.
35) Liu B, Moore JE, *et al.*: Outcomes of mobilisation of vulnerable elders in Ontario (MOVE ON): a multisite interrupted time series evaluation of an implementation intervention to increase patient mobilisation. Age Ageing. 2018; 47: 112–119.
36) Brown CJ, Foley KT, *et al.*: Comparison of posthospitalization function and community mobility in hospital mobility program and usual care patients. JAMA Intern Med. 2016; 176: 921–927.
37) Sørensen K, Van den Broucke S, *et al.*: Health literacy and public health: a systematic review and integration of definitions and models. BMC Public Health. 2012; 12: 80.
38) Friis K, Vind BD, *et al.*: The relationship between health literacy and health behaviour in people with diabetes. J Diabetes Res. 2016; 2016: 7823130.
39) Aaby A, Friis K, *et al.*: Health literacy is associated with health behaviour and self-reported health: a large population-based study in individuals with cardiovascular disease. Eur J Prev Cardiol. 2017; 24: 1880–1888.
40) Huang CH, Lai YC, *et al.*: Impact of health literacy on frailty among community-dwelling seniors. J Clin Med. 2018; 7: 481.
41) Mundy LM, Leet TL, *et al.*: Early mobilization of patients hospitalized with community-acquired pneumonia. Chest. 2003; 124: 883–889.
42) Kosse NM, Dutmer AL, *et al.*: Effectiveness and feasibility of early physical rehabilitation programs for geriatric hospitalized patients. BMC Geriatr. 2013; 13: 107.
43) Hodgson CL, Capell E, *et al.*: Early mobilization of patients in intensive care: organization, communication and safety factors that influence translation into clinical practice. Crit Care. 2018; 22: 77.

PHYSICAL THERAPY RESEARCH

BRIEF REPORT

Outcomes of Physiotherapy on Activities of Daily Living and Discharge to the Community in Psychiatric Long-term Care Ward Patients

Munetsugu KOTA, PT, PhD[1], Sae UEZONO, PT[2], Yusuke ISHIBASHI, PT, MS[3], Shouichi KURAMOCHI, PT[4], Sousuke KITAKAZE, PT[5] and Seiji KAGANOI, PT, MA[6]

[1)]*Department of Health Sciences, Hiroshima Cosmopolitan University, Japan*
[2)]*Department of Rehabilitation, Hirakawa Hospital, Japan*
[3)]*Department of Rehabilitation, Akitsu Kounoike Hospital, Japan*
[4)]*Department of Rehabilitation, Hospital Bando, Japan*
[5)]*Department of Rehabilitation, Maple Hill Hospital, Japan*
[6)]*Department of Rehabilitation, Geisei Hospital, Japan*

ABSTRACT. Objective: The purpose of this study was to investigate the outcomes of physiotherapy on patients in psychiatric long-term care wards in Japan and to identify the characteristics of patients who have been discharged to the community. Methods: The subjects comprised 50 patients who were admitted to the psychiatric long-term care wards at four different hospitals in Japan and prescribed physiotherapy. General physiotherapy for the patients' diseases was provided. The main outcome was whether a patient was discharged to the community (discharged group) or remained hospitalized (hospitalized group) at the end of physiotherapy. Basic subject characteristics, including age, sex, F-code, classification of the diagnosis that led to physiotherapy, length of hospital stay, and length of physiotherapy, were collected from medical records. The Functional Independence Measure (FIM) tool was administered at the initial and final evaluations. Results: At the end of physiotherapy, there were 14 subjects in the discharged group and 36 subjects in the hospitalized group. There were significant differences in the classification of diagnosis, length of stay (LOS), and classification of LOS between the two groups. Two-way analysis of variance showed interactions between the FIM subitems of self-care, transfer, and locomotion. Conclusion: The discharged group had higher FIM scores at the start of physiotherapy and a greater FIM gain.

Key words: Psychiatric long-term care wards, Physiotherapy, ADL, Schizophrenia

(Phys Ther Res 25: 84–91, 2022)

In Japan, the number of psychiatric beds per 100,000 people in 1998 was 287, the highest number in the world[1)]. The average length of stay (LOS) for psychiatric and behavioral treatment is 277.1 days in Japan[2)]. Moreover, long periods of hospitalization are common in psychiatric long-term care wards: 51.1% of inpatients have an LOS exceeding 5 years and 82.2% have an LOS exceeding 1 year[3)].

Received: December 1, 2021
Accepted: March 12, 2022
Advance Publication by J-STAGE: May 13, 2022
Correspondence to: Munetsugu Kota. Department of Health Sciences, Hiroshima Cosmopolitan University, 3-2-1, Otsuka-Higashi, Asaminami-ku, Hiroshima, Hiroshima 731-3166, Japan
e-mail: kota@hcu.ac.jp
doi: 10.1298/ptr.E10159

Because a prolonged LOS makes it difficult to discharge patients to the community, various studies have investigated the factors contributing to a prolonged LOS among inpatients with psychiatric disease. Longer LOS has been associated with older age, male sex, ethnicity, accommodations and family environment, multiple psychiatric diagnoses, medical-psychiatric comorbidity, and restraint use during current admission[4–7)]. In addition, a study that analyzed the LOS from the perspective of stratification into patient, hospital, and community levels found that, as in other previous studies, patient- and hospital-level factors affect LOS; furthermore, community level factors such as living situation before hospitalization and available financial support are also reported to have an effect[8)]. Thus, it is clear that not only psychiatric disease but also physical and psychosocial factors can prolong hospitalization, suggesting the importance of treatment and the difficulty of discharging patients to the community.

It has been reported that patients with schizophrenia have poor physical health and physical activity levels[9]. Several systematic reviews and meta-analyses[10–12] have reported that exercise therapy, such as aerobic exercise, strength training and yoga, are effective in improving positive and negative symptoms, quality of life, cognition, and hippocampal plasticity and in increasing hippocampal volume in patients with schizophrenia. However, until recently, a medical fee structure for physiotherapy for patients in psychiatric long-term care wards in Japan had not been established, so insufficient physiotherapy was provided to patients. Thus, to promote the transition of inpatients to the community, fee-for-service reimbursement for physiotherapy that was provided to a patient hospitalized in psychiatric long-term care wards who have developed a specific disease was offered beginning in fiscal year 2020. Few studies have investigated the effects of physiotherapy on patients admitted to psychiatric long-term care wards in Japan.

The purposes of this study were to investigate the outcomes of physiotherapy in psychiatric long-term care wards in Japan and to identify the characteristics of patients who have been discharged to the community.

Methods

Experimental design

This was a retrospective cohort study.

Subjects

The subjects comprised 59 patients admitted to psychiatric long-term care wards at four different hospitals in Japan who were prescribed physiotherapy. Inclusion criteria were those who were admitted to a psychiatric long-term care ward and those who received physical therapy through medical service fees for rehabilitation examined according to the disease group. Exclusion criteria were those for whom physical therapy was discontinued due to hospital transfer or sudden change in condition.

Fifty patients were included in the analysis after excluding five patients who were transferred to another hospital and four patients who discontinued physiotherapy due to sudden health changes.

The study was conducted from April 2020 to March 2021 in accordance with the tenets of the Declaration of Helsinki and was approved by the Ethics Committee of Hirakawa Hospital (approval number: H30-1).

Intervention

General physiotherapy for the patient's disease was provided, including muscle strengthening exercises, joint range of motion exercises, balance exercises, and activities of daily living (ADL) exercises, including ambulation exercises. Each patient was treated by one physiotherapist who provided physiotherapy for 40 minutes a day, approximately five times a week.

Outcome measures

The main outcome was whether the patient was discharged to the community, such as to a home or a group home (discharged group), or remained hospitalized (hospitalized group) at the end of physiotherapy. Basic subject characteristics, including age, sex, classification of mental disorders, classification of diagnosis that led to physiotherapy, LOS in the hospital, and length of physiotherapy, were collected from the medical records. LOS was classified into three categories: less than 6 months, 6 months to 3 years, and more than 3 years.

The Functional Independence Measure (FIM) was completed by the physiotherapist in charge of subjects at the initial evaluation and final evaluation based on information from the ward nurse.

Statistical analysis

Basic subject characteristics, the FIM score at the start of physiotherapy (starting FIM), the FIM score at the end of physiotherapy (final FIM), and the increase in FIM score (FIM gain) were compared between the discharged group and the hospitalized group. FIM gain was calculated by subtracting the starting FIM from the final FIM. To compare the two groups, Welch's two-sample t-test (r), Fisher's exact test (φ or Cramer's V), and the Mann-Whitney U test (r) were employed.

Comparisons between two groups in changes of FIM score at the starting and end of physiotherapy were analyzed by two-way mixed-design analysis of variance (ANOVA).

We conducted all analyses using R version 4.0.3. The threshold for significance was $p < 0.05$.

Results

At the end of physiotherapy, there were 14 subjects in the discharged group and 36 subjects in the hospitalized group (Fig. 1).

Basic subject characteristics are listed in Table 1. There were significant differences in the classification of diagnosis, LOS, and LOS classification between the discharged and hospitalized groups (Table 1). In terms of FIM scores at the start of physiotherapy, there were significant differences in eating, grooming, bathing, comprehension, expression, memory, motor subscore, cognitive subscore, and the total score between the two groups (Table 2). In terms of FIM scores at the end of physiotherapy, there were significant differences for each of the items except toileting, walking/wheelchair, social interaction and problem solving, and the total score between the two groups (Table 2). In terms of FIM gain, there were significant differences in grooming, bathing, dressing upper body, dressing lower

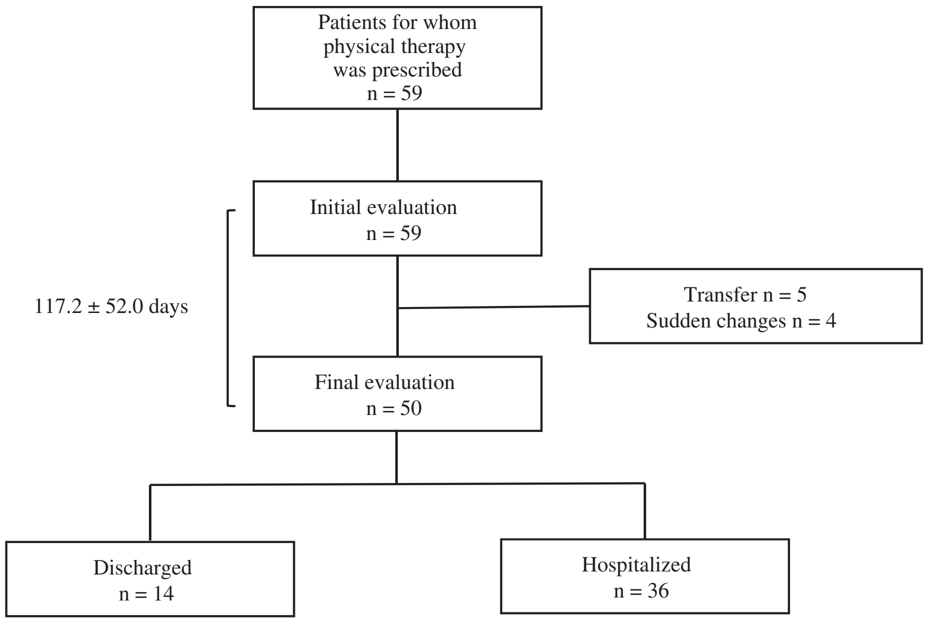


Fig. 1. Flow chart of participant selection

body, toileting, transfer (tub, shower), stairs, motor subtotal score, and the total score between the two groups (Table 2).

In comparisons between two groups in changes of FIM score at the starting and end of physiotherapy, there were significant interactions between phases and groups in self-care, transfer, locomotion, motor subscore, and the total score (Table 2).

Discussion

In this study, patients who received physiotherapy in psychiatric long-term care wards were divided into two groups according to whether they were discharged, and basic subject characteristics and FIM scores were compared between groups.

There was a significant difference between the discharged group and the hospitalized group in the classification of diagnosis that led to physiotherapy. Fractures were the most common proportion in the discharged group, while disuse syndrome and musculoskeletal ambulation disability symptom complex were more common in the hospitalized group (Table 1). People with schizophrenia are known to be at significantly increased risk of fractures[13)], and complex psychopharmacological treatment[14)] and excessive bone mineral density loss[15)] have been identified as contributing factors. Physiotherapy for femoral neck fractures yielded poorer 1-year functional outcomes in patients with schizophrenia than in patients without schizophrenia but caused no increase in postoperative complications and no difference in 1-year mortality[16)], indicating a certain effect of physical therapy. In addition, the longer the LOS, the less likely the patients were to be discharged at the end of physiotherapy (Table 1). Previous studies[17,18)] of patients who did not receive physiotherapy showed that prolonged hospitalization was an impediment to the decision to discharge. Among psychiatric wards, patients with schizophrenia have been found to have particularly prolonged LOS[17)], and in addition, a large percentage of patients admitted to psychiatric long-term care wards also have locomotive syndrome, and it has been pointed out that locomotive syndrome is also associated with prolonged LOS[18)]. In this study, the same was true for those who received physiotherapy. As another perspective, the fewer number of outreach service such as visits for psychiatric nursing care at home has also been pointed out as a cause of prolonged LOS[19)]. In this study, we were not able to examine the relationship between the involvement of physiotherapists other than ADL and dischargeability, such as whether the involvement of physiotherapists was effective in promoting the use of social resources. This may be an issue for future research.

The comparison of FIM scores at the start of physiotherapy showed significant differences between the two groups, mainly in the subcategories of self-care and communication (Table 2). Eating, grooming, and bathing are basic activities necessary for living in any environment, and even if the patient is discharged from the hospital to a group home, it is expected that these activities can be performed independently or with supervision. Comprehension, expression, and memory require thinking that involves meaningful words, and it has also been suggested that thought disorder leads to social limitations[20)]. These results suggest that the degree of disability in self-care, communication, and memory at the start of physiotherapy may be a predictor of whether a patient will be discharged. Sakiyama et al.[21)] reported that factors that cause patients with mental disorders to be rehospitalized after discharge from the hospital

Table 1. Basic subject characteristics and a comparison between the discharged group and the hospitalized group

Variables	Overall (n = 50)	Discharged (n = 14)	Hospitalized (n = 36)	p	ES	
Age, years	67.0 ± 11.0	63.8 ± 10	68.2 ± 11.2	0.182	0.184	a
Female, n (%)	30 (60.0)	7 (50.0)	23 (63.9)	0.522	0.127	b
Classification of mental disorders				0.439	0.290	c
Alzheimer's disease, n (%)	4 (8.0)	0 (0)	4 (11.1)			
Schizophrenia, n (%)	38 (76.0)	11 (78.6)	27 (75.0)			
Mood disorders, n (%)	3 (6.0)	1 (7.1)	2 (5.6)			
Eating disorders, n (%)	1 (2.0)	1 (7.1)	0 (0)			
Disorder onset in childhood, n (%)	4 (8.0)	1 (7.1)	3 (8.3)			
Classification of diagnosis				0.005	0.546	c
DS, n (%)	19 (38.0)	3 (21.4)	16 (44.4)			
MADS, n (%)	14 (28.0)	2 (14.3)	12 (33.3)			
Fractures, n (%)	9 (18.0)	6 (42.9)	3 (8.3)			
OOD (%)	6 (12.0)	1 (7.1)	5 (13.9)			
ND, n (%)	2 (4.0)	2 (14.3)	0 (0)			
LOS, days	2804.2 ± 4646	611.6 ± 1272.8	3656.8 ± 5191	0.002	0.297	a
<6 M, n (%)	14 (28.0)	10 (71.4)	4 (11.1)	<0.001	0.608	c
≥6 M, <3 Y, n (%)	12 (24.0)	2 (14.3)	10 (27.8)			
≥3 Y, n (%)	24 (48.0)	2 (14.3)	22 (61.1)			
Length of physiotherapy, days	117.2 ± 52.0	114.8 ± 74.3	118.2 ± 41.5	0.873	0.030	a

Mean ± SD

a: Welch two-sample t test, r
b: Fisher's exact test, φ
c: Fisher's exact test, Cramer's V

ES, effect size; disorder onset in childhood, behavioral and emotional disorders with onset usually occurring in childhood and adolescence; DS, disuse syndrome; MADS, musculoskeletal ambulation disability symptom complex; OOD, other orthopedic diseases; ND, neurological diseases; classification of diagnosis, classification of the diagnosis that led to physiotherapy

are associated with disorders in self-care and thought disorders, and the present study supports the results of their study.

The comparison of FIM scores at the end of physical therapy between the two groups showed differences in many subitems in addition to the total score and subtotal scores; the subitems that did not show differences were transfer to the toilet, walk/wheelchair, social interaction, and problem-solving (Table 2). Although there was no significant difference for transfer to the toilet, it showed a moderate effect size ($r = 0.267$), indicating a tendency for a difference between the two groups. On the other hand, for walk/wheelchair, there was no significant difference between the two groups, and the effect size was small ($r = 0.190$). Regarding FIM gait, it can be inferred that in the discharged group, many of the patients (42.9%) underwent physiotherapy due to fracture, and although the median value of FIM gait improved from 5 to 6 with physiotherapy, many of them still required walking aids such as canes at the end of the physiotherapy. On the other hand, in the hospitalized group, the median value of FIM gait at the start of physiotherapy was 5, which was similar to that of the discharged group. However, most of the diseases causing physiotherapy were disuse syndrome and musculoskeletal ambulation disability symptom complex, and physiotherapy was not effective in improving walking independence.

Social cognition in the form of social interaction and problem-solving is known to influence functioning in patients with schizophrenia[22]; however, in this study, these values for these items at the end of physiotherapy were not significantly different between the discharged group and the hospitalized group. The reason for these results is that the FIM items for social interaction and problem-solving do not sufficiently account for the frequency and degree of positive and negative symptoms caused by schizophrenia and other psychiatric disorders, suggesting that there may have been confounding factors that reduce reliability, such as intertester variability.

The analysis of FIM gain between the two groups showed significant differences in several subitems. In the category of self-care, significant differences were found in all subitems except for eating. There was a ceiling effect on the eating assessment because of the high level of independence at the start of physiotherapy. Although there were no significant differences in bath transfer and stairs between the two groups at the start of physiotherapy, there were significant differences in FIM gain; significant differences were also observed at the end of physiotherapy. These items suggest that although the patients did not actively practice these items during hospitalization for safety reasons, they may have practiced these items after

Table 2. Comparison of FIM scores between the discharged group and the hospitalized group

	At the start of physiotherapy				At the end of physiotherapy				FIM gain				P × G
	DG (n = 14)	HG (n = 36)	p	ES (r)	DG (n = 14)	HG (n = 36)	p	ES (r)	DG (n = 14)	HG (n = 36)	p	ES (r)	p
Self-care													[a]
Eating	7 (0.5)	6 (1)	0.001	0.471	7 (0.5)	6 (1)	<0.001	0.500	0 (0)	0 (0)	0.944	0.052	[a]
Grooming	5 (0.5)	3 (2)	0.004	0.402	6 (1)	3 (2)	<0.001	0.496	0.5 (0.5)	0 (0)	0.007	0.369	[a]
Bathing	4 (1.375)	2 (1.5)	0.013	0.348	5 (1.75)	2 (1.5)	0.005	0.391	0 (0.375)	0 (0)	0.031	0.312	[a]
Dressing – upper body	5 (1.25)	4 (2)	0.069	0.257	6.5 (1.5)	4 (2.5)	0.005	0.387	1 (0.875)	0 (0)	0.005	0.413	[a]
Dressing – lower body	5 (1.25)	2.5 (2)	0.071	0.256	6 (1.5)	3.5 (2.5)	0.008	0.373	1 (0.875)	0 (0)	0.001	0.449	[a]
Toileting	6 (1.625)	3 (2.5)	0.066	0.260	6.5 (1)	4 (3)	0.024	0.317	0.5 (0.875)	0 (0)	0.039	0.295	[a]
Sphincter control													[a]
Bladder management	6 (1)	2.5 (2.625)	0.063	0.265	6.5 (0.875)	2.5 (2.625)	0.007	0.380	0 (0.375)	0 (0)	0.057	0.274	[a]
Bowel management	5 (1.25)	1.5 (2.125)	0.057	0.269	5.5 (0.875)	2 (2.5)	0.011	0.354	0 (0.375)	0 (0)	0.060	0.265	[a]
Transfers													[a]
Bed, chair, wheelchair	6 (1.375)	5 (2.125)	0.141	0.209	7 (0.875)	5.5 (2)	0.049	0.278	0 (0.5)	0 (0.5)	0.642	0.069	[a]
Toilet	6 (1.375)	4.5 (2.625)	0.127	0.217	6.5 (1)	5 (2)	0.060	0.267	0 (0.5)	0 (0.5)	0.620	0.073	[a]
Tub, shower	4.5 (1.5)	3 (2)	0.244	0.168	5 (1.5)	3 (2)	0.038	0.293	0.5 (0.5)	0 (0)	0.004	0.418	[a]
Locomotion													[a]
Walker/wheelchair	5 (1.5)	5 (2.125)	0.676	0.061	6 (0.875)	5 (1.625)	0.182	0.190	1 (1)	0 (0.5)	0.092	0.240	[a]
Stairs	1 (0)	1 (0)	0.606	0.081	3 (2.5)	1 (0)	0.019	0.335	0 (1.875)	0 (0)	0.004	0.393	[a]
Communication													[a]
Comprehension	5 (0.875)	4 (1.25)	0.012	0.354	5 (0.875)	4 (1.5)	0.005	0.389	0 (0)	0 (0)	0.372	0.121	[a]
Expression	5 (0.5)	4.5 (1.5)	0.027	0.311	5 (0.5)	4.5 (1.5)	0.007	0.378	0 (0)	0 (0)	0.155	0.238	[a]

(*Continued*)

Table 2. (*Continued*)

	At the start of physiotherapy				At the end of physiotherapy				FIM gain					P × G
	DG (n = 14)	HG (n = 36)	p	ES (r)	DG (n = 14)	HG (n = 36)	p	ES (r)	DG (n = 14)	HG (n = 36)	p	ES (r)		p
Social cognition													a	
Social interaction	5 (0.75)	3 (2)	0.196	0.185	5 (0.5)	3.5 (2)	0.140	0.211	0 (0)	0 (0)	0.806	0.020	a	
Problem-solving	4 (1.375)	3 (1.5)	0.325	0.141	4.5 (1.375)	3 (1.5)	0.182	0.190	0 (0.375)	0 (0)	0.288	0.156	a	
Memory	5 (1.875)	3 (1.125)	0.030	0.305	5 (1.75)	3 (1.25)	0.014	0.345	0 (0)	0 (0)	0.906	0.027	a	
Summaries														
Self-care	23 ± 8.7	15.9 ± 10.0	0.019	0.319	26.6 ± 8.2	16.7 ± 10.6	0.001	0.411					b	0.005
Sphincter control	9.9 ± 4.3	6.8 ± 5.1	0.037	0.280	10.7 ± 3.8	6.9 ± 5.2	0.007	0.340					b	0.267
Transfers	14.4 ± 5.7	11.6 ± 6.6	0.140	0.202	16.8 ± 3.6	12.6 ± 6.3	0.005	0.319					b	0.020
Locomotion	6.4 ± 3.6	5.7 ± 3.5	0.580	0.081	9.0 ± 3.8	6.4 ± 3.5	0.040	0.314					b	0.006
Communication	10.4 ± 2.8	7.8 ± 3.2	0.009	0.358	10.7 ± 2.3	7.8 ± 3.1	0.001	0.410					b	0.357
Social cognition	4.7 ± 1.4	3.8 ± 2.1	0.081	0.213	4.9 ± 1.5	3.9 ± 2	0.067	0.236					b	0.749
Motor subscore	53.7 ± 20.8	40.0 ± 23.7	0.054	0.265	63.1 ± 17.6	42.6 ± 24.3	0.002	0.382					b	0.012
Cognitive subscore	23.4 ± 6.9	18.0 ± 8.4	0.027	0.295	24.5 ± 6.2	18.4 ± 8.3	0.008	0.340					b	0.410
Total score	77.1 ± 26.4	58.0 ± 30.6	0.037	0.285	87.6 ± 20.9	61.0 ± 31.0	0.001	0.391					b	0.023

Median (quartile deviation); mean ± SD

a: Mann–Whitney U test, ES: r
b: Pairwise t test, ES: r
P × G: Two-way mixed-design ANOVA, Interaction of Phase × Groups

DG, discharged group; HG, hospitalized group; ES, effect size; ANOVA, analysis of variance physiotherapy started in anticipation of discharge from the hospital, with a view to the range of activities they would be performing and safe living after discharge. It has been pointed out that patients with schizophrenia have impaired ADL abilities due to a multitude of factors, including impaired physical functioning[23], and it has also been reported that impaired ADL abilities are also present in patients with bipolar disorder[24]. In addition, rehabilitation interventions for patients with schizophrenia have been reported to improve ADL abilities, although to a limited extent, which is consistent with the results of this study[25]. Although the number of FIM subitems that showed improvement was limited in this study, it can be inferred that in some cases, discharge planning was conducted from the beginning of physiotherapy, and practice of ADL was carried out in anticipation of life after discharge.

The FIM subscores that showed that improvement in ADL by physiotherapy influenced dischargeability by two-way ANOVA were self-care, transfer, and locomotion. All these subscores observed significant difference including subitems that showed a moderate to large effect size of 0.4 in the FIM gain analysis of FIM subitems. These subitems were dressing in the self-care category, transferring to the tub in the transfer category, and stairs in the locomotion category. These subitems are ADLs that are not necessarily of high priority in terms of the need for independence during hospitalization, when care can be expected, but are necessary during life in the community, and improvement in these subitems is thought to have influenced discharge to the community.

Limitations of this study include the mixture of multiple psychiatric disorders and the insufficient assessment of psychiatric symptoms. Psychiatric symptoms and motor functions that cannot be assessed by FIM may affect the degree of improvement in FIM and whether the patient can be discharged to the community. Therefore, it is necessary to limit the classification of mental disorders and diagnoses that led to physiotherapy in the future. In addition, the reference to the effects of physiotherapy is limited because we did not conduct a randomized controlled trial with or without physical therapy intervention. Another limitation of the study was that the sample size was too small to conduct multivariate analysis on the factors influencing whether patients are discharged from the hospital. In particular, classification of diagnosis that led to physiotherapy and LOS can be important confounders, so continued research is needed. Despite these limitations, it is significant that physiotherapy had a certain outcome in improving ADL ability and facilitating discharge to the community in patients in a psychiatric long-term care ward.

Conclusion

The outcomes of physiotherapy in a psychiatric long-term care ward were analyzed by examining the results of discharged and hospitalized groups. The results showed that there were differences between the two groups in basic characteristics such as classification of diagnosis that led to physiotherapy and length of hospital stay. It was also found that the discharged group had higher FIM scores at the start of physiotherapy and a greater FIM gain.

Acknowledgments: We thank the staff members of all the hospitals who collaborated in this study. We are grateful to the following institutions for facilitating this multicenter study: Hirakawa Hospital, Akitsu Konoike Hospital, and Maple Hill Hospital.

Conflict of Interest: The authors declare no conflicts of interest associated with this manuscript.

References

1) Oshima I, Mino Y, *et al.*: Institutionalisation and schizophrenia in Japan: social environments and negative symptoms: nationwide survey of in-patients. Br J Psychiatry. 2003; 183: 50–56.
2) Ministry of Health and Welfare: Patient survey: Statistics and Information Department, Minister's Secretariat, Ministry of Health and Welfare, Tokyo, 2017, p. 12. (in Japanese)
3) Ministry of Health, Labour and Welfare [Internet]. Tokyo: Central Social Insurance Medical Council; 40th Result verification committee for medical fee revision. Ken 1-1 [Updated 2014 Jun 25; Cited 2021 Sep 22]. Available from: https://www.mhlw.go.jp/file/05-Shingikai-12404000-Hokenkyoku-Iryouka/0000049161.pdf
4) Cheng J, Shumway M, *et al.*: Patient factors associated with extended length of stay in the psychiatric inpatient units of a large urban county hospital. Community Ment Health J. 2016; 52: 658–661.
5) Jansen L, van Schijndel M, *et al.*: Health-economic outcomes in hospital patients with medical-psychiatric comorbidity: a systematic review and meta-analysis. PLoS ONE. 2018; 13: e0194029.
6) Newman L, Harris V, *et al.*: Factors associated with length of stay in psychiatric inpatient services in London. UK Psychiatr Q. 2018; 89: 33–43.
7) Zeshan M, Waqas A, *et al.*: Factors predicting length of stay in an adolescent psychiatric unit, South Bronx, NY: a short report. J Can Acad Child Adolesc Psychiatry. 2018; 27: 142–147.
8) Miller DA, Ronis ST, *et al.*: The impact of demographic, clinical, and institutional factors on psychiatric inpatient length-of-stay. Adm Policy Ment Health. 2021; 48: 683–694.
9) McNamee L, Mead G, *et al.*: Schizophrenia, poor physical health and physical activity: evidence-based interventions are required to reduce major health inequalities. Br J Psychiatry. 2013; 203: 239–241.
10) Vancampfort D, Probst M, *et al.*: Systematic review of the benefits of physical therapy within a multidisciplinary care approach for people with schizophrenia. Phys Ther. 2012; 92: 11–23.
11) Girdler SJ, Confino JE, *et al.*: Exercise as a treatment for schizophrenia: a review. Psychopharmacol Bull. 2019; 49: 56–69.

12) Firth J, Stubbs B, *et al.*: Aerobic exercise improves cognitive functioning in people with schizophrenia: a systematic review and meta-analysis. Schizophr Bull. 2017; 43: 546–556.

13) Stubbs B, Gaughran F, *et al.*: Schizophrenia and the risk of fractures: a systematic review and comparative meta-analysis. Gen Hosp Psychiatry. 2015; 37: 126–133.

14) Sørensen HJ, Jensen SOW, *et al.*: Schizophrenia, antipsychotics and risk of hip fracture: a population-based analysis. Eur Neuropsychopharmacol. 2013; 23: 872–878.

15) Kishimoto T, Hert MD, *et al.*: Osteoporosis and fracture risk in people with schizophrenia. Curr Opin Psychiatry. 2012; 25: 415–429.

16) Ng JPH, Ho SWL, *et al.*: Functional outcomes of patients with schizophrenia after hip fracture surgery: a 1-year follow-up from an institutional hip fracture registry. J Bone Joint Surg Am. 2021; 103: 786–794.

17) Kono T, Shiraishi H, *et al.*: Discharge dynamics and related factors of long-stay patients in psychiatric hospitals. Seishin Shinkeigaku Zasshi. 2015; 117: 713–729. (in Japanese)

18) Kota M, Uezono S, *et al.*: Relationship between whether the planned discharge destination is decided and locomotive syndrome for admitted patients in psychiatric long-term care wards. Phys Ther Res. 2020; 23: 180–187.

19) Niimura J, Nakanishi M, *et al.*: Regional supply of outreach service and length of stay in psychiatric hospital among patients with schizophrenia: national case mix data analysis in Japan. Psychiatry Res. 2017; 258: 295–298.

20) Marggraf MP, Lysaker PH, *et al.*: The link between formal thought disorder and social functioning in schizophrenia: a meta-analysis. Eur Psychiatry. 2020; 63: e34.

21) Sakiyama S, Iida J, *et al.*: Factors of good outcome after discharge from support house (engoryou) for schizophrenia. Psychiatry Clin Neurosci. 2002; 56: 609–615.

22) Couture SM, Penn DL, *et al.*: The functional significance of social cognition in schizophrenia: a review. Schizophr Bull. 2006; 32: S44–S63.

23) Strassnig M, Signorile J, *et al.*: Physical performance and disability in schizophrenia. Schizophr Res Cogn. 2014; 1: 112–121.

24) Moore K, Merritt B, *et al.*: ADL skill profiles across three psychiatric diagnoses. Scand J Occup Ther. 2010; 17: 77–85.

25) Ayres H, Ngo H, *et al.*: Limited changes in activities of daily life performance ability among people with schizophrenia at clinical settings and the factors moderating the changes. Schizophr Res Cogn. 2019; 16: 29–35.

Physical Therapy Research

http://www.jstage.jst.go.jp/browse/ptr

Editor-in-chief

Hiroyuki SHIMADA (National Center for Geriatrics and Gerontology)

Editorial Board

Takehiko DOI (National Center for Geriatrics and Gerontology)
Kazuhiro HARADA (Kibi International University)
Hiroyuki HASHIDATE (Kyorin University)
Daisuke HIGUCHI (Takasaki University of Health and Welfare)
Tome IKEZOE (Kansai Medical University)
Tomoya ISHIGAKI (Nagoya Gakuin University)
Fuminari KANEKO (Tokyo Metropolitan University)
Tatsuya KASUYAMA (Health Science University)
Kenichi KONO (International University of Health and Welfare)
Ryo KOZU (Nagasaki University)
Masafumi KUBOTA (Kanazawa University)
Yasuaki KUSUMOTO (Fukushima Medical University)
Noriaki MAEDA (Hiroshima University)
Hyuma MAKIZAKO (Kagoshima University)
Shinichiro MORISHITA (Fukushima Medical University)
Hideki MORIYAMA (Kobe University)
Ken MURAMATSU (Kyorin University)
Koutatsu NAGAI (Hyogo Medical University)
Yasuhide NAKAYAMA (The Jikei University School of Medicine)
Ippei NOJIMA (Shinshu University)
Junya OZAWA (Hiroshima International University)
Harutoshi SAKAKIMA (Kagoshima University)
Kenichi SUGAWARA (Kanagawa University of Human Services)
Makoto TAKAHASHI (Hiroshima University)
Hiroshige TATEUCHI (Kyoto University)
Eiki TSUSHIMA (Hirosaki University)
Masayuki UCHI (Toho University Omori Medical Center)
Minoru YAMADA (Tsukuba University)
Tomofumi YAMAGUCHI (Juntendo University)

Editorial Advisory Board

Yoshiteru AKEZAKI
Mutsuaki EDAMA
Yoshihiro FUKUMOTO
Tatsuya HIRASE
Junichiro INOUE
Mizuri ISHIDA
Kentaro KAMIYA
Keisuke KAWAKAMI
Nobuhiro KITO
Masayoshi KUBO
Hideaki MATSUO
Shu MORIOKA
Kazuaki NAGASAKA
Kazuki NAKANISHI
Manabu NANKAKU
Masafumi NOZOE
Rei ONO
Masakazu SAITOH
Megumi SUMIZONO
Minoru TABATA
Akiyoshi TAKAMI
Takako TANAKA
Atsuhiro TSUBAKI
Shintaro UEHARA
Shigeru USUDA
Tatsuki YOSHIMATSU
Masaya ANAN
Hidemi FUJINO
Masatoshi HANADA
Hiroto HONDA
Yu INOUE
Kazuhiro IZAWA
Osamu KATAYAMA
Yuriko KIHARA
Ryoji KIYAMA
Hiroshi MAEJIMA
Saori MIYAGISHIMA
Tomoyuki MORISAWA
Toru NAKA
Jiro NAKANO
Tomohiko NISHIGAMI
Shinya OGAYA
Michihiro OSUMI
Noboru SEKIYA
Kazuaki SUYAMA
Satoshi TAKAGI
Katsuhiko TAKATORI
Masashi TANIGUCHI
Toshiya TSURUSAKI
Kazuki UEMURA
Satoru WATANABE
Tadayoshi ASAKA
Hiroyuki FUJISAWA
Yumi HIGUCHI
Hikaru IHIRA
Kazuto ISHIDA
Naoto KAMIDE
Michitaka KATO
Ippei KITADE
Makoto KOMIYA
Tadamitsu MATSUDA
Toshiaki MIYAMOTO
Motoyoshi MORISHITA
Masatoshi NAKAMURA
Tatsuya NAKANOWATARI
Satoshi NOBUSAKO
Yohei OKADA
Shotaro OTSUKA
Hisato SUGATA
Risa SUZUKI
Yasuyuki TAKAKURA
Shigeo TANABE
Koki TOMIDA
Kota TSUTSUMIMOTO
Susumu URAKAWA
Masami YOKOGAWA

PTR is published by Japanese Society of Physical Therapy. c/o Japanese Physical Therapy Association, 7-11-10 Roppongi, Minato-ku, Tokyo 106-0032, Japan, FAX +81-3-5843-1748.

No part of this publication may be reproduced or transmitted in any form, or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission in writing from the Society. All rights reserved.