

Table 2 Age-specific number and incidence of fractures after the first hip fracture

Age group (years)	n	Person (years)	Subsequent fractures						Incidence (per 1,000 person-years)			Rate ratio of hip fractures ^a
			Once	Twice	Total patients	Total fractures	Hip fractures	Clinical vertebral fractures	All fractures	Hip fractures	Clinical vertebral fractures	
65–74	276	275	15	2	17	19	10	3	69.0	36.3	10.9	18.6
75–84	919	906	59	2	61	63	26	15	69.6	28.7	16.6	3.3
85–94	1,025	1,003	62	3	65	68	36	7	67.8	35.9	7.0	1.5
≥95	108	106	10	0	10	10	5		94.8	47.4		1.9
Total	2,328	2,289	146	7	153	160	77	25	69.9	33.6	10.9	4.0

^a Age- and gender-specific incidences, reported previously for the general population in Japan, were adopted to compare the risk of hip fracture among patients with a first hip fracture

walk without difficulty,” “to walk outside with a walking aid,” and “to walk only inside with an aid” and “dependent” was “unable to walk without support” and “completely unable to walk.” The proportion of patients with subsequent fractures was significantly higher in the “dependent” than in the “no aid” ambulatory group. The proportion of patients with subsequent fractures was significantly higher among patients treated with antiosteoporosis drugs than those without antiosteoporosis drugs during the posthospitalization duration. There was no significant difference between the two groups regarding other factors. A multivariate analysis was performed using variants such as body weight and ambulatory abilities before the first hip fracture, and only body weight was a significant risk factor for subsequent fractures (OR = 0.96, 95% CI 0.928–0.993).

Discussion

This study demonstrated a high risk of subsequent fractures in patients with an initial hip fracture. It also revealed that, after fractures, antiosteoporosis pharmacotherapy was prescribed in only 19.6% of patients during their hospitalization and in only 18.7% during the 1-year follow-up period after discharge from the first hospital stay. The strength of this study is the substantially large number of subjects with a wide age distribution. As a result, the age-specific incidence of a second hip fracture in patients with an initial hip fracture could be calculated.

In retrospective studies, the rate of sustaining a second hip fracture is reported to be 7.5–11.8% [9–12]. In a population-based cohort study, Melton et al. [13] estimated that the recurrence rate for an additional hip fracture was 1% at 1 year after the first fracture in Minnesota in 1943–1977. Another US population-based study showed almost identical data, with a cumulative incidence of a second hip fracture of 2.3–2.5% after 1 year [2, 14], while the cumulative incidence after 5 years varied between 8.2

and 20%. A recent report from Finland showed a higher risk of 5.08% for persons aged 60 years and over within the first year following an initial fracture [15]. These differences may mirror the large variations in the risk of hip fractures and life expectancies in different populations.

Very little data are available on the risk of sustaining a second hip fracture after an initial hip fracture in the Asian population, although there is a large difference in the incidence of fragility fractures between Asians and Caucasians [1]. In a prospective study, Yamanashi et al. [16] reported that the annual incidence of a second hip fracture was 0.038 per person-year during the first year after an initial hip fracture and 0.028 per person-year during the second year. The 1-year incidence rate as determined in our current study (3.40%) is very consistent with their data. The age- and gender-specific incidence rates of sustaining a second hip fracture established in this study indicate that the rate ratio of a second hip fracture compared to the general population in Japan is higher in patients with an initial hip fracture under 75 years of age than that in those 75 years or older. In a Danish study, age- and gender-specific incidence rates (per 1,000 person-years) in women were 40, 51, 62, and 73 in the age groups 60–69, 70–79, 80–89, and ≥90 years, respectively [14]. Although the incidence rates of a second hip fracture in the Japanese and Caucasian populations are thought to be equal [16], there are differences in the population over 70 years of age, possibly due to a lower hip-fracture incidence in the Japanese population compared to that in northern Europe.

In this study, second hip fractures most frequently occurred within 32 weeks after the first hip fracture. The rate of sustaining a second hip fracture in a previous Japanese study tended to rise during the first 8 months after the first hip fracture and then to plateau [16]. A recent nationwide study in Denmark with a large number of hip-fracture patients demonstrated that the risk of a second hip fracture was increased almost 12 times at 1 month and more than doubled at 1 year; however, the risk remained significantly elevated until 15 years of follow-up [14].

These findings indicate that the early period of 6 months after the first fracture is a window of opportunity to execute a strategy for fracture prevention.

It has been reported that a previous fracture at any site is an important risk factor for future fractures [17]. Although the risk factors for hip fractures have been well defined in Caucasian subjects, only two epidemiological studies have examined risk factors for hip fractures in Japanese subjects [18, 19]. In the current study, body weight was a risk factor for a second hip fracture; however, age and other comorbidities were not associated with this risk. The study by Yamanashi et al. [16] demonstrated that senile dementia and Parkinson disease are important risk factors for second hip fractures and that prefracture ambulatory ability was similar between the unilateral and bilateral hip-fracture groups. A large cohort Danish study demonstrated that prior fractures, alcoholism, living alone, higher income, and advanced age are risk factors for second hip fractures [14]. Lonroos et al. [15] found no significant risk of a second hip fracture with multiple comorbidities including dementia, whereas Berry et al. [2] reported a significantly increased risk in patients with a high level of functioning compared to those with moderate functioning. On the other hand, Chapurlat et al. [20] found that walking for exercise, which is an indirect marker for functional status, was a protective predictor. The explanation for these discrepancies might be caused by differences in the definition of comorbidity used in the various studies [14].

It is reported that the vast majority of patients who experience a hip fracture do not take antiosteoporotic therapy after the fracture [7, 8]. Among patients who begin antiresorptive osteoporosis treatment after fracture, the adherence to treatment decreases over time and remains suboptimal [8, 21]. Inadequate treatment after the first hip fracture became evident in the current study. In the current study we also found that patients undergoing osteoporotic therapy were more likely to sustain a new fracture than those not undergoing pharmacotherapy. This is probably due to the fact that more fractures occurred in those with severe osteoporosis; therefore, there was an increased likelihood of prescription of pharmacotherapy. It was not determined if prescribing antiosteoporosis drugs during the 1-year posthospitalization period was effective for fracture risk reduction. It is now known that oral bisphosphonate treatments for 3 years [22] or an annual infusion of zoledronic acid for 1.9 years [23] after repair of a hip fracture is associated with a reduction in the rate of new clinical fractures including hip fracture [22] as well as an improvement in survival [23].

There are some limitations to the current study. First, we estimated the risk of sustaining a second hip fracture based

on hospital records from the treating hospital and questionnaires from the patients and calculated rate ratios based on reported age-specific incidence rates derived from a different research method. Because mortality is increased after a hip fracture, this approach underestimates the true incidence and might lead to biased estimates. Although the mortality in this study was much lower than that reported previously, 304 patients were lost to follow-up and their true life span was unknown. Since further follow-up was difficult, this is a limitation of this study; therefore, the true mortality of this study must be interpreted with caution. Second, we included only female patients ≥ 65 years. It is well known that mortality and prognoses are poorer in male versus female patients, and this might affect our results. However, this effect would be limited since the number of female patients with a hip fracture is 3.8 times that of male patients in Japan [24]. Third, we collected data based on admission records and therefore could not fully eliminate patients with additional hip fractures treated at other hospitals. Each rural hospital from where our data originate is the main hospital in its region and has orthopedic specialists. Thus, most patients sustaining fractures visit these hospitals, and the number of patients with an additional hip fracture who are missed would be limited. However, the possible bias introduced by this would underestimate the fracture incidence. A lack of sufficient validation is a potential weakness. Unfortunately, we were not able to further validate our data. Finally, we could not compare the BMD between patients with and without a subsequent hip fracture since BMD was measured in only 314 patients. Although some reports have suggested that BMD between patients with and without an additional hip fracture was similar [10], another study showed that patients with a lower BMD are more likely to sustain an additional hip fracture [20].

In conclusion, the current study discovered a high risk of a subsequent fracture after an initial hip fracture and that the treatment for patients after the first hip fracture is not entirely adequate. Since hip-fracture patients are the most plausible candidates in the prevention of subsequent fractures, prescribing appropriate osteoporosis treatments is essential along with more aggressive interventions for preventing falls.

Acknowledgments The authors sincerely acknowledge the staff members of the hospitals that cooperated with this Prevention of Second Hip Fracture (POSHIP) study. The authors also acknowledge the Japanese Society for Fracture Repair. The authors express their sincere appreciation to Ms. Kaoru Yamabe and Ms. Machiko Abe for their help in the analyses. The authors also acknowledge Ms. Ryoko Ikehara for her secretarial assistance. This study was mainly supported by a grant from the Japanese Society for Fracture Repair and partially supported by a grant-in-aid from the Ministry of Health, Labor and Welfare of Japan (21-shi, 23-22 Choujyu Iryou).

References

1. Hagino H, Furukawa K, Fujiwara S, Okano T, Katagiri H, Yamamoto K, Teshima R (2009) Recent trends in the incidence and lifetime risk of hip fracture in Tottori, Japan. *Osteoporos Int* 20:543–548
2. Berry SD, Samelson EJ, Hannan MT, McLean RR, Lu M, Cupples LA, Shaffer ML, Beiser AL, Kelly-Hayes M, Kiel DP (2007) Second hip fracture in older men and women: the Framingham Study. *Arch Intern Med* 167:1971–1976
3. Leslie WD, O'Donnell S, Jean S, Lagace C, Walsh P, Bancej C, Morin S, Hanley DA, Papaioannou A (2009) Trends in hip fracture rates in Canada. *JAMA* 302:883–889
4. Melton LJ 3rd, Kearns AE, Atkinson EJ, Bolander ME, Achenbach SJ, Huddlestone JM, Therneau TM, Leibes CL (2009) Secular trends in hip fracture incidence and recurrence. *Osteoporos Int* 20:687–694
5. Abrahamsen B, Vestergaard P (2009) Declining incidence of hip fractures and the extent of use of anti-osteoporotic therapy in Denmark 1997–2006. *Osteoporos Int* 21:373–380
6. Orimo H, Yaegashi Y, Onoda T, Fukushima Y, Hosoi T, Sakata K (2009) Hip fracture incidence in Japan: estimates of new patients in 2007 and 20-year trends. *Arch Osteoporos* 4:71–77
7. Cadarette SM, Katz JN, Brookhart MA, Levin R, Stedman MR, Choudhry NK, Solomon DH (2008) Trends in drug prescribing for osteoporosis after hip fracture, 1995–2004. *J Rheumatol* 35:319–326
8. Roerholt C, Eiken P, Abrahamsen B (2009) Initiation of anti-osteoporotic therapy in patients with recent fractures: a nationwide analysis of prescription rates and persistence. *Osteoporos Int* 20:299–307
9. Boston DA (1982) Bilateral fractures of the femoral neck. *Injury* 14:207–210
10. Dretakis KE, Dretakis EK, Papakitsou EF, Psarakis S, Steriopoulos K (1998) Possible predisposing factors for the second hip fracture. *Calcif Tissue Int* 62:366–369
11. Dinah AF (2002) Sequential hip fractures in elderly patients. *Injury* 33:393–394
12. Shabat S, Gepstein R, Mann G, Kish B, Fredman B, Nyska M (2003) The second hip fracture—an analysis of 84 elderly patients. *J Orthop Trauma* 17:613–617
13. Melton LJ III, Ilstrup DM, Beckenbaugh RD, Riggs BL (1982) Hip fracture recurrence. A population-based study. *Clin Orthop Relat Res* 167:131–138
14. Ryg J, Rejnmark L, Overgaard S, Brixen K, Vestergaard P (2009) Hip fracture patients at risk of second hip fracture: a nationwide population-based cohort study of 169, 145 cases during 1977–2001. *J Bone Miner Res* 24:1299–1307
15. Lonnroos E, Kautiainen H, Karppi P, Hartikainen S, Kiviranta I, Sulkava R (2007) Incidence of second hip fractures. A population-based study. *Osteoporos Int* 18:1279–1285
16. Yamanashi A, Yamazaki K, Kanamori M, Mochizuki K, Okamoto S, Koide Y, Kin K, Nagano A (2005) Assessment of risk factors for second hip fractures in Japanese elderly. *Osteoporos Int* 16:1239–1246
17. Klotzbuecher CM, Ross PD, Landsman PB, Abbott TA III, Berger M (2000) Patients with prior fractures have an increased risk of future fractures: a summary of the literature and statistical synthesis. *J Bone Miner Res* 15:721–739
18. Suzuki T, Yoshida H, Hashimoto T, Yoshimura N, Fujiwara S, Fukunaga M, Nakamura T, Yoh K, Inoue T, Hosoi T, Orimo H (1997) Case-control study of risk factors for hip fractures in the Japanese elderly by a Mediterranean Osteoporosis Study (MEDOS) questionnaire. *Bone* 21:461–467
19. Fujiwara S, Kasagi F, Yamada M, Kodama K (1997) Risk factors for hip fracture in a Japanese cohort. *J Bone Miner Res* 12:998–1004
20. Chapurlat RD, Bauer DC, Nevitt M, Stone K, Cummings SR (2003) Incidence and risk factors for a second hip fracture in elderly women. The study of osteoporotic fractures. *Osteoporos Int* 14:130–136
21. Rabenda V, Vanoverloop J, Fabri V, Mertens R, Sumkay F, Vannecke C, Deswaef A, Verpooten GA, Reginster JY (2008) Low incidence of anti-osteoporosis treatment after hip fracture. *J Bone Joint Surg Am* 90:2142–2148
22. Osaki M, Tatsuki K, Hashikawa T, Norimatsu T, Chiba K, Motokawa S, Furuichi I, Doiguchi Y, Aoyagi K, Shindo H (2011) Beneficial effect of risedronate for preventing recurrent hip fracture in the elderly Japanese women. *Osteoporos Int*. doi: 10.1007/s00198-011-1556-7
23. Lyles KW, Colon-Emeric CS, Magaziner JS, Adachi JD, Pieper CF, Mautalen C, Hyldstrup L, Recknor C, Nordsletten L, Moore KA, Lavecchia C, Zhang J, Mesenbrink P, Hodgson PK, Abrams K, Orloff JJ, Horowitz Z, Eriksen EF, Boonen S (2007) Zoledronic acid in reducing clinical fracture and mortality after hip fracture. *N Engl J Med* 357(18):1799–1809
24. Hagino H, Sakamoto K, Harada A, Nakamura T, Mutoh Y, Mori S, Endo N, Nakano T, Itoi E, Kita K, Yamamoto N, Aoyagi K, Yamazaki K (2010) Nationwide one-decade survey of hip fractures in Japan. *J Orthop Sci* 15:737–745

介護予防の効果に関する実証分析

一「介護予防事業等の効果に関する総合的評価・分析に関する研究」における傾向スコア調整法を導入した運動器の機能向上プログラムの効果に関する分析一

伊藤 和彦¹⁾ 大淵 修一²⁾ 辻 一郎³⁾

本論では介護予防を目的とした運動器の機能向上プログラムへの継続的な参加の効果について、厚生労働省が実施した介護予防継続的評価・分析（辻他，2009）の研究結果およびデータを用いて、先行研究では明示的に扱われなかった選択バイアスを調整したうえで、生存時間分析および「差の差」の手法によりその効果を分析した。

はじめにKaplan-Meier法により継続的に参加した高齢者の効果を確認し、次に加速ハザードモデル（AFTモデル）により推計された偏回帰係数を用いて、運動器の機能向上プログラムの参加の有無によって、加速要素（Acceleration Factor）がどのくらい異なるのかを推計した。他の条件を一定として、それぞれリスクが発現する時間が遅くなる効果は、非参加者に比べ特定高齢者で1.6倍、要支援者群で3.1倍と推計された。また、偏回帰係数の検定結果から、特定高齢者群、要支援者群といったカテゴリーによって違いがあるものの、ものわずれテストの点数、既往歴などが介護予防の効果に影響を与えていることを確認した。

さらに、基本チェックリストを用いた「差の差」の手法から、運動器の機能向上プログラムに1年を超えて継続的に参加した高齢者では、そうでない高齢者に比べ、基本チェックリストの点数で測った改善は、特定高齢者で2.0～2.5点程度、要支援者群で0.45～0.5点程度、トータルで0.8点程度、特に要支援者群と全数を用いた推計では有意な改善効果を計量的に確認することができた。

キーワード 介護予防、介護予防継続的評価・分析、傾向スコア、生存時間分析、「差の差」の手法

1. はじめに

日本の平均余命は世界的にトップレベルにある。長寿を保つ高齢者が増える一方、介護が必要な状態で人生の末期を過ごす高齢者も少なくない。高齢化の進展に伴い、介護保険制度が導入され

た2000年度には3.6兆円だった介護保険財政の規模も、2009年度予算では7.2兆円に膨れ上がり、結果、利用者負担の目安となる第一号被保険者月額保険料の全国加重平均は、2,911円から4,160円にまで上昇。要介護状態を防ぐと同時に介護費用の抑制が喫緊の課題となっている。

高齢者の健康の維持、介護費用の抑制という二つの課題に応えるべく、2005年の介護保険法改正にあわせて、予防重視型システムへの変換がはかられ、介護予防サービスが予防給付サービスとし

¹⁾ 東京都教育庁

²⁾ 地方独立行政法人東京都健康長寿医療センター研究所

³⁾ 東北大学大学院医学系研究科

て取り入れられることになった。

介護予防とは、「要介護状態の発生をできる限り防ぐ(遅らせる)こと、そして要介護状態にあってもその悪化をできる限り防ぐこと」と定義される(厚生労働省「介護予防マニュアル(改訂版)」)。介護予防は、高齢者の運動機能の向上や栄養状態管理を通じて、個々の高齢者の活動レベルや役割の向上をもたらし、それによって一人ひとりの生きがいや自己実現のための取り組みを支援して、生活の質(QOL)の向上をめざすものである。

本論の目的は、厚生労働省が実施した介護予防継続的評価・分析(辻他, 2009)のデータを用いて、運動器の機能向上を目的とした各種プログラムへの継続的な参加による介護予防の効果を実証することである。より具体的には、その嚆矢である辻他(2009)で展開された分析を踏まえ、後述する選択バイアス(セレクションバイアス)にも配慮したうえで、介護予防の効果のうち、運動器の機能向上を目的とした各種プログラムへの継続的な参加による現在の状態を維持することへの効果や改善の程度を計量的に確認する。

運動器の機能向上プログラムとは、生活機能を向上させることを目的に、筋力トレーニング、バランストレーニング、柔軟性トレーニングなどを通じて基本的な動作能力を高めるプログラムである。虚弱高齢者に適用することに配慮して、コンディショニング期、筋力増強期、機能的トレーニング期の3期に分けて徐々に負荷量を増加し実施されている。高齢期では体力の個人差が大きいことから、評価に基づき、個別のプログラムを作成し実施することとされている。

具体的には、運動器の機能向上プログラムに継続的に参加した特定高齢者^{注1)}および要支援者が、参加しなかった特定高齢者および要支援者に比

注1) 特定高齢者という名称については2010年8月より「二次予防事業対象者」と名称変更となっているが、本論では調査時点での名称である「特定高齢者」をそのまま使用している。

べ、現在の状態をより長く維持できるのか、維持できるとすれば年齢や性別、既往歴、生活環境などの影響を受けるのか、また、基本チェックリストの点数などの高齢者の健康状態を把握する具体的なアウトカム(成果)を示す指数はどのくらい改善されたのか、生存時間分析(サバイバル分析)や「差の差」の手法(Difference in Differences = DID)といった近年のマイクロ計量経済分析の研究成果を利用してこのような疑問を明らかにする。

本論の構成は以下の通りである。第2節では、辻他(2009)等の先行研究の成果を簡単にサーベイし、今回使用したデータについて概観したうえで、選択バイアスの存在を指摘する。第3節では、本論での分析手法について言及する。そこでは運動器の機能向上プログラムへの参加効果の確認のための生存時間分析や傾向スコア(Propensity Score Matching = PSM)を利用した「差の差」の手法について言及する。第4節では推計結果を用いての考察が行われる。そこでは、運動器の機能向上プログラムへの参加効果を生存時間分析によって確認するとともに、健康状態を示す具体的なアウトカム(基本チェックリストの点数)がどのくらい改善されたのかを明らかにする。最終節では考察結果を総括したうえで、政策的含意が示される。

2. 先行研究とデータ

1) 先行研究

大規模な長期追跡データに基づき介護予防の効果を実証した日本で最初の本格的な研究として辻他(2009)があげられる。同研究は運動器の機能向上、口腔機能の向上や栄養改善の3つのプログラムの効果をカバーしており、介護予防導入による費用対効果を数量的に明らかにしたうえで^{注2)}、要介護認定状況や基本チェックリストの点数が改善または維持されたか、あるいは悪化したかとい

注2) 大久保(2009)

う二値を従属変数とし、年齢や性別、基本チェックリストの点数、既往歴や日常生活における役割、独居区分などのダミー変数を共変量とするロジスティック回帰分析を運動器の機能向上プログラム、口腔機能の向上プログラム、栄養改善プログラムのそれぞれで実施している。ロジスティック回帰分析から得られたオッズ比から判断して、いずれのプログラムにおいても介護予防の効果を計量的に確認している。

大測 他 (2010) では、運動器関連指標の計測が行われている同研究のデータベースの3,852人を対象として、特定高齢者および要支援者それぞれを対象とした運動器関連指標の分析を実施した。男女とも、特定高齢者と要支援者とは、運動器関連指標の分布が異なり、特に移動能力において差が大きいことを明らかにしたうえで、特定高齢者と要支援者では、運動器の機能向上プログラムの効果判定には、異なる評価基準を用いる必要性を指摘した。そのほか、辻 他 (2009) の研究に参加した多くの研究者が各分野において同研究のデータベースを用いて詳細な研究を実施し、研究の蓄積が進みつつあるが、運動器の機能向上プログラムに焦点を絞って、選択バイアスを考慮して参加者の効果を実証的に分析した先行研究は、筆者らの知りうる範囲では存在しない。

2) データ

本分析において使用するデータは、「介護予防事業等の効果に関する総合的評価・分析に関する研究」に使用されたデータ (サンプル数19,827件) で、2007年1月より2008年12月までの期間において、3ヶ月に1回の調査を基本として集計された個票データと調査に参加した83の地方公共団体に関するデータからなる。個票データについては、地域包括支援センターで収集され、同センターで匿名化されたうえ、厚生労働省を經由して研究班に提供された。本研究課題は、2008年10月24日、東北大学医学部倫理審査委員会で承認された。

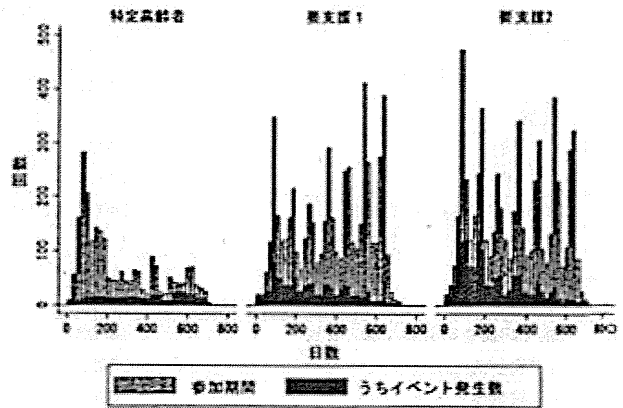


図1 参加期間 (日数) とイベント発生までの日数の分布 (カテゴリー別)

個票データについては、単純に集計すると不完全な12期のパネルデータになるが、開始日がまちまちであることに加えて、脱落例や設問による部分的な欠損値も多く、そのままパネルデータとして使うと問題がある。辻 他 (2009) では、1回しか調査が行われていない症例、および介護予防サービスを開始してから第1回目の調査まで3ヶ月以上を経過している症例を除いた総数9,105の症例を用いて、最初に調査が行われた期を第1期、12ヶ月後またはそれ以前に終了したもののについては終了時点を第2期とする2期間のデータとして分析した。しかし、本分析では選択バイアスにも十分留意すること、利用できる症例は最大限利用すべきとの考えに基づき、19,827件のうち1回しか調査されていないデータ1,010件を除いた18,817件から、個別の設問で回答が脱落している症例を除いた説明変数として利用できる最大数である14,002件を基本として分析を進める。そのうち男性は3,078人、女性が10,924人、特定高齢者は2,330人 (うち男性487人)、要支援1は5,548人 (同1,178人)、要支援2は6,124人 (同1,413人) である。その他の主な変数の基本統計量は表1に示される。

分析の対象となった高齢者の運動器の機能向上プログラムへの参加期間 (日数) およびイベント発生までの日数の分布状況は図1に示される。本調査の終了である2008年12月時点で調査開始時点

表1 基本統計量

変数	全数 (N=14002)			特定高齢者 (N=2330)			要支援者群 (N=11672)			
	サンプル数	平均	標準誤差	サンプル数	平均	標準誤差	サンプル数	平均	標準誤差	
gender	性別	14002	0.22	0.41	2330	0.21	0.41	11672	0.22	0.42
	1: 男性	3078			487			2591		
	0: 女性	10924			1843			9081		
age	第1回調査時点の満年齢	14002	82.00	6.57	2330	79.92	6.35	11672	82.42	6.53
single	第1回調査時点の独居区分	14002	0.36	0.48	2330	0.25	0.43	11672	0.39	0.49
	1: 独居	5076			572			4504		
	0: 同居	8926			1758			7168		
role	第1回調査時点の役割の有無	14002	0.37	0.48	2330	0.58	0.49	11672	0.33	0.47
	1: 役割あり	5235			970			3875		
	0: 役割なし	8767			1360			7797		
cognitive	第1回調査時点の認知的活動テストの点数	14002	16.32	4.37	2330	17.36	4.22	11672	16.11	4.37
gds	第1回調査時点のGDSテストの点数	14002	5.21	3.46	2330	4.40	3.20	11672	5.37	3.49
hds	第1回調査時点のもの忘れテストの点数	14002	24.13	4.97	2330	25.20	4.47	11672	23.92	5.03
checklist	第1回調査時点の基本チェックリストの点数	14002	10.55	4.15	2330	8.71	4.03	11672	10.92	4.07
brain	脳血管疾患の既往歴	14002	0.16	0.37	2330	0.09	0.29	11672	0.18	0.38
	1: 既往あり	2268			210			2058		
	0: 既往なし	11734			2120			9614		
rheumat	関節疾患(リウマチ等)の疾患既往歴	14002	0.25	0.43	2330	0.22	0.41	11672	0.26	0.44
	1: 既往あり	3540			513			3027		
	0: 既往なし	10462			1817			8645		
diments	認知症の既往歴	14002	0.04	0.20	2330	0.04	0.18	11672	0.04	0.20
	1: 既往あり	586			82			504		
	0: 既往なし	13416			2248			11168		
fracture	骨折・転倒の既往歴	14002	0.17	0.38	2330	0.10	0.30	11672	0.19	0.39
	1: 既往あり	2268			232			2191		
	0: 既往なし	11734			2098			9481		
weaken	高齢による衰弱による既往歴	14002	0.09	0.28	2330	0.09	0.28	11672	0.08	0.28
	1: 既往あり	1195			206			989		
	0: 既往なし	12807			2124			10683		
indept	日常生活の自立度	14002	0.73	0.44	2330	0.96	0.19	11672	0.68	0.46
	1: ランクJ	10226			2244			7982		
	0: その他	3776			86			3690		
TRT	運動器の機能向上プログラムを継続して参加している(※)	14002	0.37	0.48	2330	0.60	0.49	11672	0.32	0.47
	1: 参加	5167			1399			3768		
	0: 非参加	8835			931			7904		
over75yrs	第1回調査時点の後期高齢者	14002	0.85	0.36	2330	0.77	0.42	11672	0.87	0.34
	1: 75歳以上	11921			2591			10130		
	0: 75歳未満	2081			9081			1542		
t	時間t (第1回目と観測された最終回の日数)	14002	348.55	194.91	2330	265.84	195.66	11672	365.06	190.51

(※) 継続かどうかの基準として、観測された最終回の観測日まで継続している場合は継続とした。ただし、継続期間が90日に満たないものは継続には含めていない。

の介護認定が継続している症例、および家族の反対、本人の希望、転居等によりそれ以前に調査を終了した症例^{注3)}については、後述する生存時間分析において「右打ち切り例」(Right Censored Case)として扱った。また、第1回目の記入日

注3) 本調査は終了事由についても調査している。

以前から運動器の機能向上プログラムに参加していた場合は「左打ち切り例」(Left Censored Case)として統計的なバイアスの原因となる可能性がある。しかし、サービス開始日についての質問項目は8,000近くの症例が欠損値となっており、サービス開始日を正確に判定できないことから、開始日を起算点として使用した(サービス開

表2 運動器の機能向上プログラムの選択に関わるバイアス

被説明変数 TRT	推定回帰係数	標準誤差
gender	-0.04	0.05
age	0.16	0.05
age ²	0.00	0.00
single	-0.55	0.04
role	0.10	0.04
cognitive	0.01	0.00
gds	-0.02	0.01
hds	0.01	0.00
checklist	-0.01	0.01
braint	0.30	0.05
rheumat	0.17	0.04
diments	-0.50	0.10
fracture	0.03	0.05
weaken	-0.20	0.07
サンプル	14002	
対数尤度	-9014.9	
決定係数 (psued R squared)	0.022	

* p<.05 ; ** p<.01 ; *** p<.001

始日と第1回目の記入日の双方がわかるデータの多くが同じ月であり、影響は少ないと思われる。)

3) 選択バイアスの存在

本調査では、公募により選択された地方公共団体の症例が抽出されている。さらに、登録者の選定にあたり、参加した団体は登録者をランダムに選定したわけではなく、調査に協力的な高齢者が選ばれた可能性もある。このため、たとえば「健康状態の良い」、「運動が得意」といった高齢者が運動器の機能向上プログラムを選択し、そうでない高齢者が選択しない場合の両者のアウトカムを比較したとすれば、得られた推定量に上方バイアスが生じている可能性がある。その場合の推計結果は参加した高齢者の体調や運動能力を反映した結果であり、継続的な参加の効果を反映したものではない。この点を確認するため、第1回調査時点で運動器の機能向上プログラムを選択した(=1)、しない(=0)を従属変数とし、年齢、生活環境、疾患履歴などを代理する説明変数からなる多変量ロジスティック回帰分析を行った^{注4)}。結果は表2に示されているとおり、ほとんどの説

明変数の偏回帰係数の検定結果が有意であることから、選択バイアスの可能性が示唆された。

3. 実証分析の方法と仮説

1) 分析方法

新しい治療方法の有用性を検証するといった場合と同様に、介護予防の効果を厳密に検証するにあたっては、選択バイアスを回避するために無作為割付けを伴う実験研究が行われる必要がある。しかし、実際に生活している高齢者を対象とする研究では、身体状況や生活環境を考慮して無作為割付けを行うことは、事実上不可能に近く、さらに疾患履歴や心身の状況が複雑に交絡しているため、研究者の側でこの点を考慮した分散分析などの手法が広く用いられている。にもかかわらず、共分散分析などの手法は従属変数と共変量の関係を事前に線形関数として指定する必要があるなど様々な制約があるため、Rosenbaum and Rubin (1983, 1985) が提案した傾向スコアによる共変量調整法が近年利用されるようになりつつある(星野・岡田, 2006)。選択バイアスを考慮してもなお、運動器の機能向上プログラムへの継続的な参加の効果を確認するため、ここでは、ロジスティック回帰分析により算出した傾向スコアを利用し、以下の2通りの方法^{注5)}により、運動器の機能向上プログラムへの継続的な参加による効果を確認する(図2)。第1に、傾向スコアによりリサンプリングされたデータを用い生存時間分析を実施し、運動器の機能向上プログラムに継続的に参加した高齢者群(処理群)とそうでない高齢者群(対照群)の生存時間曲線に有意差があるかを確認したうえで、パラメトリックモデルを推計し、年齢、性別、既往歴などの生存時間関数への

注4) ここで用いられている説明変数は、辻他(2009)で用いられているものと同様のものを使った。なお、変数の定義については、表1の基本統計量および3-3)生存時間分析で説明される。

注5) 分析の手順はGuo, Barth and Gibbons(2006)を参考にしている。

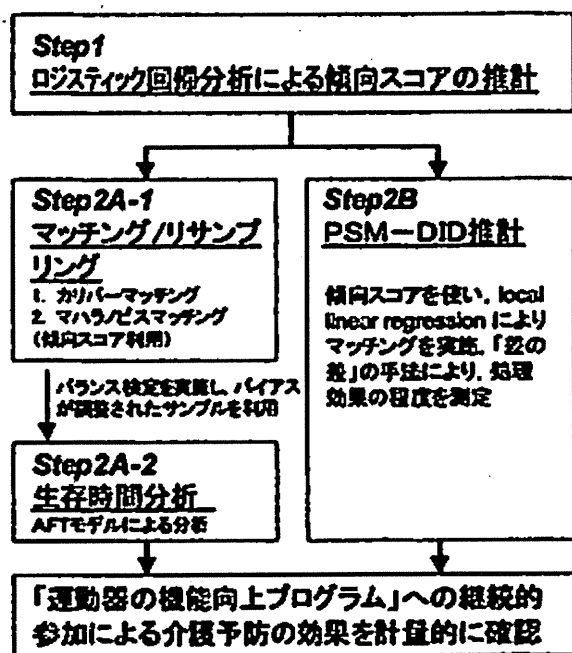


図2 分析の手順

影響を考察する。第2に、傾向スコアを利用した「差の差」(PSM-DID)の手法を用いて、高齢者の健康状態を示すアウトカム(基本チェックリスト)がどのくらい改善したのかを検討する²⁶⁾。

2) プログラム評価の手法と傾向スコア、処理群と対照群のマッチング²⁷⁾

運動器の機能向上プログラムの効果を評価するにあたり、同じ高齢者が参加と不参加を同時に経験できないという困難に直面する。しかし、ここでの関心は個人の効果ではなく、政策対象者である高齢者全体への効果であることから、処理群の平均的効果(Average Treatment Effect on the Treated=ATT)、つまり、処理群の参加した場合とそうでない場合の差、具体的には $ATT = E(Y_1|D=1) - E(Y_0|D=1)$ を求める場合を考える。ここで Y_1 は参加した場合の効果、 Y_0

は参加しなかった場合の効果、 D は処理効果を表す指標で、 $D=0$ であれば対照群、 $D=1$ は処理群であることを示す。容易に判るように、第1項については実際のデータから計算することができるが、第2項は得ることができない。実際に計算できるのは、「参加しなかった」高齢者の平均的な効果 $E(Y_0|D=0)$ である。いま、実際に計算できる Y_1 の期待値と参加の有無を示す D は独立であると仮定すると、処理群の平均的効果(ATT)を次のように近似することができる。

$$ATT \approx E(Y_1|D=1) - E(Y_0|D=0) \quad (1)$$

しかし、社会的実験の場合にこのような無作為割付けの状態を設定することは不可能に近いことから、この問題に対処するために、Rosenbaum and Rubin(1983, 1985)は傾向スコアを利用したマッチング手法を提案した。「要介護状態になる」という結果に影響を与えると考えられる複数の変数を使って、ロジスティック回帰分析などによりプログラムに参加する確率 $Pr[d_i=1|x]$ (傾向スコア) を求めることで、複数の説明変数を一つに集約し、傾向スコアが近い処理群と対照群のサンプルをマッチング(あるいはリサンプリング)させる手法である²⁸⁾。マッチングの方法は多数あるが、本分析では、他の分析例に従い²⁹⁾ 複数

26) 推計にあたっては、STATA SE 11.2およびPSMについては、ユーザー作成のソフトウェアLeuven and Sianesi (2003) psmatch2を利用した。

27) ここでの記載は北村(2009)、星野(2009)、Heinrich, Maffioli and Vazquez (2010)を参考にしている。

28) 傾向スコア計算のためのロジスティック回帰分析で利用する説明変数は以下の通り。性別、年齢、基本チェックリストの点数、認知的活動テストの点数、GDSテストの点数、ものわすれテストの点数、日常生活における役割の有無を示すダミー変数、独居を示すダミー変数、既往の有無を示すダミー変数(脳血管疾患、関節疾患、認知症、転倒・骨折、高齢による衰弱の5項目のそれぞれ)、日常生活の自立度(自立か寝たきりかのダミー変数)、要支援者を示すダミー変数、直前の3ヶ月の入院の有無を示すダミー変数、低所得者ダミー変数(年間所得80万円以下)、83の参加公共団体を示すダミー変数を用いた(変数の定義の詳細は表1で説明される)。

29) たとえばGuo, Barth and Gibbons (2006)。なお、マッチングの詳細については、星野(2009)、Guo and Franser (2009)に詳しい。

のマッチング方法を採用した。一つは最近傍カリパーマッチング (nearest-neighbor matching with caliper) で、もう一つはマハラノビスマッチング (mahalanobis matching including propensity score) である。

この2通りの方法によるリサンプリングが有効であるかどうかを確認するため、リサンプリング後、処理群と対照群の両グループの平均値の差が0と有意に異なるか否かをt検定によって判断するバランス検定 (Balancing test) を実施する。各変数について両グループの間に統計的に有意な差が認められなければ、参加した高齢者と同じような属性を持つ非参加高齢者がリサンプリングされていると判断することができる。次に、リサンプリング後のプログラム参加者と非参加者をプールしたサンプルでロジスティック回帰分析を実施し、リサンプリングが適切に行われていることの確認として、説明変数が何ら説明力を有しないことを確認する。

3) 生存時間分析

次に、傾向スコア調整法によりリサンプリングされたデータを用いて、生存時間分析¹¹¹⁾により運動器の機能向上プログラムへの参加効果や状態の持続期間に関する評価を行う。

生存時間分析では、あるイベントが起こるまでの時間を確率関数である生存時間の関数として表現し分析する。たとえば、時間の経過に伴ってどのように特定高齢者の身体状態が変化するかに関心があるとき、特定高齢者が要支援または要介護になる時間を確率変数 T で表す。 T が時間 t より大きくなる確率を生存時間関数とよび、 $S(t) = Pr(T > t)$ と表す。

実証研究においては、パラメトリックモデル¹¹¹⁾によってパラメータを推定し、説明変数の

生存時間に対する影響を考察することができるが、その前提として実データから生存時間の分布を仮定する必要がある¹¹²⁾。今回の調査における生存時間の分布は、はじめの200日くらいまでは増加し、その後は減少する傾向が見受けられる。したがって、今回の分析では、対数ロジスティック分布または対数正規分布を仮定して分析を進めるが、具体的なパラメトリックモデル¹¹³⁾の特定にあたっては、(1) ワイブル分布の可能性についても考慮し比例ハザード性の仮定が成立するかをSchoenfeld残差によりテストし、(2) 対数ロジスティックモデルが成立するかどうかを判断するために、イベント発生のおッズ比が時間を通して一定、 $\frac{1-S(t)}{S(t)} = \lambda t^\gamma$ 、その対数オッズ比と時間

の対数 t が $\ln\left(\frac{1-S(t)}{S(t)}\right) = \ln(\lambda) + \gamma \ln(t)$ という

¹¹¹⁾ 生存時間分析において広く用いられているCOX比例ハザードモデルは、生存時間に特定の分布を仮定することなく推計できるが、ハザード比が時間の経過に関わらず一定であることを前提とする「比例ハザード性の仮定」が成立している必要がある。具体的には、Schoenfeld残差に対する時間の関数の傾きが0ではないかを検定する。検定の結果、Schoenfeld残差は時間に依存していて、COXモデルを利用できるかどうかは疑わしい結果となった。ただし、後述するように、特定高齢者のみのデータを用いた場合は、比例ハザード性が成立する可能性があるテスト結果となっている。

¹¹²⁾ 時間の経過とともにイベントが単調に増加すると考えれば、ワイブルモデルや指数モデルの利用が示唆されるが、最初は増加し後に予防効果が奏功し減少に転じる分布パターンでは、対数正規分布や対数ロジスティック分布が適している (Cameron and Trivedi, 2005)。

¹¹³⁾ ワイブルモデルはハザード関数が各主体に共通のベースラインハザードと各主体に独自の相対ハザードの積で表される比例ハザードモデル (PHモデル) としても、ベースラインハザードが時間とともに加速していくことを想定している加速ハザードモデル (AFTモデル) としても推計することができるが、対数ロジスティック分布および対数正規分布モデルでは加速ハザードモデルでしか推計することはできないため、ここではAFTモデルを選択した。

¹¹⁰⁾ ここでの記述はKleinbaum and Klein (2005) を参考にしている。

関係を利用して、処理群、対照群ごとにプロットした対数オッズ比と時間の対数の関係を示した曲線が平行であるかどうかを確認した。ここで λ はハザード率、 γ はシェイブパラメータである。さらに、(3)対数尤度が最大であるか、またはAIC基準値が最小であるかにより最も適したモデルを判断する²¹⁰。対数ロジスティック分布を仮定した場合の生存時間関数 $S(t)$ 、ハザード関数 $\lambda(t)$ は次のとおりである²¹⁵。

$$S(t) = \frac{1}{1 + (\lambda t)^\sigma} \quad (2)$$

$$\lambda(t) = \frac{\lambda \gamma (\lambda t)^{\gamma-1}}{1 + (\lambda t)^\sigma} \quad (3)$$

ここで σ を $\lambda(t)$ の誤差とすると、 $\gamma = 1/\sigma$ となる。また、説明変数である i 番目の高齢者の属性や身体状況、生活環境などの生存関数およびハザード関数に対する影響はそれぞれ、次のような関数によって推定される。

$$\lambda_i = \exp(-\beta X_i) \quad (4)$$

β は推定されるパラメータ、 X_i は説明変数ベクトルである。

被説明変数としては、現在の介護状態が維持できなくなるまでの時間、すなわち、特定高齢者で

あれば、要支援または要介護と認定されるまでの、要支援であれば、要介護になるまでの時間を用いる。また、説明変数ベクトル X は、辻他(2009)で用いられた説明変数。具体的には、性別、年齢、基本チェックリスト²¹⁶の点数、認知的活動テスト²¹⁷の点数、GDSテストの点数²¹⁸、ものわすれテストの点数²¹⁹、日常生活における役割の有無を示すダミー変数、独居を示すダミー変数、脳血管疾患、関節疾患、認知症、転倒・骨折、高齢による衰弱といった既往の有無を示す5つのダミー変数。さらに、運動器の機能向上プログラムを継続して参加したかどうかを示すダミー変数からなる²²⁰。

年齢は若年であるほど改善確率を高めるが、年齢の上昇によって改善確率は低下することから二乗項は負の関係(age: $\beta_2 > 0$, sq_age: $\beta_3 < 0$)を、基本チェックリストの点数は低いほど改善確率が上昇する可能性が高いことから負の関係

210 実際の推計にあたっては、Cleves *et al.* (2010)を参照した。なお、本誌の匿名のレフリーより、「本データでは、おそらく死亡発生により要介護になる前に打ち切りになるなど、イベント発生に近づくほど打ち切りになりやすい傾向があるのではないかと。より複雑なモデルによる解析が必要であると考え。」とのコメントを頂戴した。要介護になるといったリスク以外にも他疾患での入院といったリスクもあり、競合するリスクの取扱については今後の課題としたい。

215 ワイブル分布モデル、対数正規分布モデル、一般化ガンマ分布モデルの場合の生存時間関数、ハザード関数については、Kleinbaum and Klein (2005) p.263やCameron and Trivedi (2005) p.585などを参照。なお、各モデルにおける推計すべき補助的なパラメータ(ancillary parameter)であるシェイブパラメータは、表4においては、ワイブルモデルが ρ 、対数ロジスティックモデルが γ 、対数正規分布モデルが σ で示されている。

216 基本チェックリストとは、「暮らしぶり」「運動器の関係」「栄養・口腔機能の関係」「こころの状況」に関連する計25項目の質問で構成されており、その合計点によって介護予防の必要性を判断する厚労省のガイドラインに沿った基準。その点数が高いほど介護予防の必要性が高いと判断される。

217 認知的活動とは、「テレビを見る」「ラジオを聴く」「新聞を読む」「雑誌を読む」「本を読む」「(トランプなど)ゲームをする」の項目を、(1)ほとんど毎日、(2)週に数回、(3)月に数回、(4)年に数回、(5)年に1回以下かどうかの5段階で評価、点数化し、その合計点が高いほど認知的状況が高いと判断する。

218 GDSテストとは、15項目からなる世界中で広く利用されている、うつ状態を測定する検査。ここでは、最近1ヶ月の状況について評価し、5点以上でうつ傾向があり、10点以上がうつ状態とされる。

219 「長谷川式簡易知能評価スケール(HDS-R)」はものわすれの程度を判断するテストで、現在の年齢や日付・曜日感覚、単純な単語の記憶、簡易な引き算などを問うもので、点数が高いほど正常な記憶が保たれていると判断される。

220 調査最終回の直前の回まで継続して参加の場合に「継続して参加した」に分類した。ただし、継続期間が90日に満たないものは除いた。

(checklist: $\beta_9 < 0$) を想定した。また、日常生活における役割の有無は、役割があるほどよりアクティビティが高くプログラムの効果を高めることから正の関係 (role: $\beta_3 > 0$) を、認知的活動状況に関するテスト、ものわすれテストの点数については高得点者ほど正常な記憶機能が保たれており効果が得られると考えられることから、それぞれ正の関係 (cognitive: $\beta_6 > 0$, hds: $\beta_8 > 0$) を、落ち込みやすい傾向が強いほど改善意欲に乏しく効果が望みにくいことからGDSテストの点数との間では負の関係 (gds: $\beta_7 < 0$) を想定した。また、既往がある場合、その身体的ダメージから改善確率を下げる可能性がありいずれも負の関係を、運動器の機能向上プログラムに継続して参加した場合、一定の効果があると考えられることから、正の関係 (TRT: $\beta_{15} > 0$) を想定した。具体的な回帰式としては以下の通りである。

$$\ln(t) = \beta_0 + \beta_1 \text{gender}_i + \beta_2 \text{age}_i + \beta_3 \text{sq_age}_i + \beta_4 \text{single}_i + \beta_5 \text{role}_i + \beta_6 \text{cognitive}_i + \beta_7 \text{gds}_i + \beta_8 \text{hds}_i + \beta_9 \text{chklist}_i + \beta_{10} \text{brain}_i + \beta_{11} \text{rheumat}_i + \beta_{12} \text{diments}_i + \beta_{13} \text{fracture}_i + \beta_{14} \text{weaken}_i + \beta_{15} \text{TRT} + \theta_{\text{municipal}} + u_i \quad (5)$$

ここで、 i は高齢者、 u は誤差項、 β_j ($j=1, \dots, 15$) は推計されるパラメータである。全数に加えて、特定高齢者、要支援者群のカテゴリ別にも推計を実施する。また、調査を実施した地方公共団体に共通の観測不能な異質性 (shared frailty) $\theta_{\text{municipal}}$ がハザード関数に存することを考慮した Frailtyモデル (ガンマ分布) も推計する。

4) 「差の差」の手法

最後に、調査開始から365日を越えて調査が継続されている症例について、第1回目と最終回では、身体状況がどのくらい改善されているのか、具体的には、基本チェックリストの点数がどのくらい改善されたかを推計する¹²¹⁾。推計にあたっては、傾向スコアを利用した「差の差」の手法 (PSM-DID) を利用する。

「差の差」の手法とは、処理群、対照群ともに処理前 (第1期)、処理後 (第2期) のデータが利用できる場合に、処理群における第2期と第1期の差から対照群における同様の差を引くことで求めることができる。参加した高齢者 (処理群) の時間を通じての処理効果は、 $ATT = E(Y_1 | D=1) - E(Y_0 | D=1)$ であり、参加しなかった高齢者 (対照群) の時間を通じての効果、ATUは $ATU = E(Y_1 | D=0) - E(Y_0 | D=0)$ ¹²²⁾ であるが、後者の時間を通じての変化は、実際の処理効果ではなく、処理効果以外の理由により時間の経過とともに変化した差分であり、この差分を差し引いて評価する必要がある。

$$ATT - ATU = E(Y_1 - Y_0 | D=1) - E(Y_1 - Y_0 | D=0) \quad (6)$$

もし、処理群、対照群の選択が完全に無作為に行われているとすれば、 $ATT - ATU$ によって DID推定量を求めることができるが、処理群への参加確率に対していずれかの変数が影響を与えているとすれば、バイアスによって正しい推定量が得られない。

これを克服するために、PSMとDIDを組み合わせた実証分析の手法が政策プログラム評価などの分野で幅広く採用されている。PSM-DIDの基本的な考え方は、運動器の機能向上プログラムに参加した高齢者とプログラムへの参加確率が同程度でかつ同様の特徴を持つ非参加者のデータをマッチングさせることにより、実際に参加した高齢者のアウトカムと、もし参加したとすれば得られたであろう仮想的なデータの差により効果の差を測

¹²¹⁾ 今回の調査では、運動器の機能向上プログラムの非参加者から握力 (kg) や歩行速度 (m/秒) といった具体的な体力データは取得しておらず、このため、今回は高齢者の健康状態を示す総合的な指標の一つである基本チェックリストを利用することとしたが、実際はそれらのデータを用いた方が、身体状況の改善という意味ではわかりやすい説明が可能であったかもしれない。

¹²²⁾ ATU = Average Treatment Effect on Untreated.

表3 傾向スコアのためのロジスティック回帰分析およびバランステスト<全数ベース>

		バランス検定										傾向スコアのための ロジスティック回帰			
		カリバーマッチ					マハラノビスマッチ					調整後			
		平均	%reduct	t-test		平均	%reduct	t-test		調整前	カリバー マッチ後	マハラノビス マッチ後			
処理群	対照群	%bias	bias	t	p値	処理群	対照群	%bias	bias	t	p値				
gender	before	0.22	0.22	0.4	0.22	0.83	0.22	0.22	0.4	0.22	0.83	0.068	-0.014	0.350	
	after	0.23	0.23	-0.4	-0.19	0.85	0.22	0.17	12.8	-3250.3	6.83	0.00	---	---	
age	before	81.76	82.14	-5.8	-3.31	0.00	81.76	82.14	-5.8	-3.31	0.00	0.085	-0.050	-0.347	
	after	81.84	81.86	-0.3	95.1	-0.13	0.90	81.76	82.20	-6.8	-16.3	-3.71	0.00	---	
single	before	0.29	0.41	-24.9	-14.09	0.00	0.29	0.41	-24.9	-14.09	0.00	-0.403	0.013	-0.117	
	after	0.32	0.32	1.3	94.9	0.56	0.57	0.29	0.31	-4.6	81.6	-2.41	0.02	---	
rule	before	0.40	0.36	7.5	4.32	0.00	0.40	0.36	7.5	4.32	0.00	-0.120	0.009	0.012	
	after	0.36	0.36	-0.9	88.7	-0.38	0.71	0.40	0.40	-0.1	98.4	-0.06	0.95	---	
cognitive	before	16.61	16.15	10.7	6.11	0.00	16.61	16.15	10.7	6.11	0.00	0.031	0.006	-0.199	
	after	16.43	16.48	-1.2	88.5	-0.54	0.59	16.61	16.45	3.7	65.6	1.98	0.05	---	
gds	before	4.96	5.35	-11.3	-6.41	0.00	4.96	5.35	-11.3	-6.41	0.00	0.003	-0.019	-0.106	
	after	5.11	5.05	1.8	84.0	0.80	0.42	4.96	4.70	7.5	33.2	4.13	0.00	---	
hds	before	24.46	23.94	10.8	6.10	0.00	24.46	23.94	10.8	6.10	0.00	0.047	-0.023	-0.106	
	after	24.32	24.36	-0.8	92.4	-0.37	0.71	24.46	24.56	-2.0	81.3	-1.13	0.26	---	
checklist	before	10.30	10.70	-9.8	-5.56	0.00	10.30	10.70	-9.8	-5.56	0.00	-0.008	0.008	-0.086	
	after	10.51	10.46	1.3	86.9	0.57	0.57	10.30	10.32	-0.6	93.4	-0.35	0.72	---	
brain	before	0.19	0.15	10.5	6.10	0.00	0.19	0.15	10.5	6.10	0.00	0.331	-0.038	0.454	
	after	0.18	0.18	0.3	96.7	0.15	0.88	0.19	0.14	11.4	-7.9	5.80	0.00	---	
rheumat	before	0.27	0.24	7.2	4.14	0.00	0.27	0.24	7.2	4.14	0.00	0.106	-0.027	0.290	
	after	0.27	0.27	0.1	99.2	0.03	0.98	0.27	0.26	2.4	66.8	1.20	0.23	---	
diments	before	0.03	0.05	-11.1	-6.15	0.00	0.03	0.05	-11.1	-6.15	0.00	-0.340	-0.096	0.306	
	after	0.03	0.04	-1.2	89.2	-0.55	0.58	0.03	0.02	1.8	83.8	1.10	0.27	---	
fracture	before	0.17	0.17	-0.6	-0.33	0.74	0.17	0.17	-0.6	-0.33	0.74	0.040	-0.016	0.415	
	after	0.18	0.18	0.5	17.1	0.21	0.84	0.17	0.14	8.7	-1404.0	4.62	0.00	---	
weaken	before	0.07	0.09	-8.0	-4.51	0.00	0.07	0.09	-8.0	-4.51	0.00	-0.139	-0.006	0.422	
	after	0.07	0.07	-0.1	98.8	-0.04	0.97	0.07	0.05	7.3	9.7	4.22	0.00	---	
サンプル数		7724					5450					14002	7722	5450	
ここでは、主要な変数のみを掲載したが、これ以外に、年齢、基本チェックリスト、HDS-R、GDS		対数尤度		-7680.4		-5327.2		-3603.7							
の各2乗項、要支援者数、過去3ヶ月の入院歴、自立度がJランクかどうかを示すダミー、参加83		カイ二乗値		3077.9		50.5		347.9							
公共団体ごとの地域ダミー変数なども推計式に含まれている。これらの変数についても、カリバー		AIC基準値		15580.8		10870.5		7427.4							
マッチングのケースではマッチング後の平均値の差に有意差はない(1%水準)が、マハラノビスマ		類似決定係数		0.1669		0.0047		0.0461							
ッチングでは、ここに示されていない多くの変数において有意差が認められており、信頼性が低い		c統計量		0.76		0.55		0.64							

* p<.05 ; ** p<.01 ; *** p<.001

定する。前節の分析のマッチングは、処理群と対照群が1対1でマッチするようリサンプリングを行い、そのサンプルをもとに生存時間分析を行ったが、ここでは、Heckaman, Ichimura and Todd (1997, 1998)により開発された、ノンパラメトリック回帰分析を用いてマッチングウェイトを算出し、ウェイト付けされた対照群のサンプルと処理群のサンプルのマッチングによりATTを求める。したがって、ここでは、前節の分析と異なり、マッチングと分析をひとつの過程で行うことができる。

ここでは先行研究を参考に^{註23)}、PSM-DIDに

註23) Guo, Barth and Gibbons (2006)

よって処理群と対照群の改善の差を推計する。調査開始後365日以上を経た平均観察期間545日の症例を利用し、マッチングウェイトによりone-to-many matchingを実施する^{註24)}。

註24) 推計の感度を確認する目的から、カーネル関数に利用されるバンド幅については、既定値である0.06のほか、0.01、0.3、0.8の場合の4通りを推計し、トリミングレベルとして2%、5%、10%の各ケースを推計した。DID推定量の標準誤差および95%信頼区間を得るためにブートストラッピングを利用した。

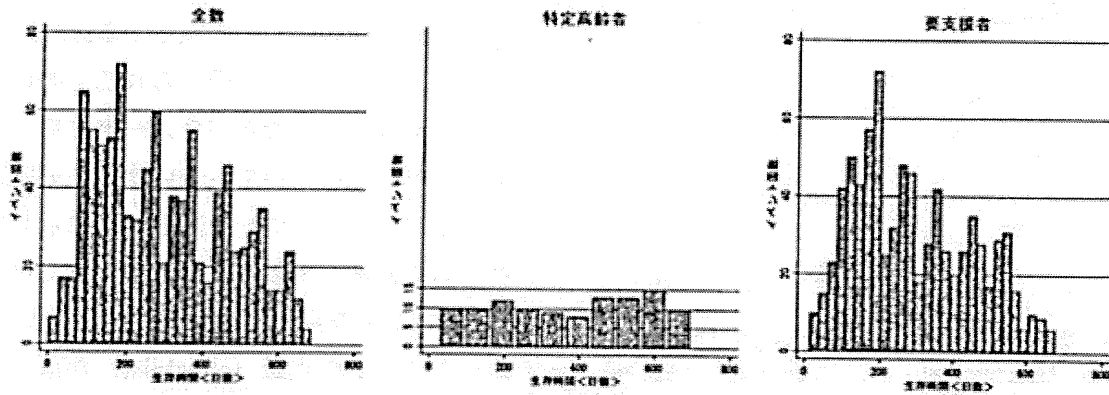


図3 イベント発生件数の生存時間別内訳 (カテゴリー別)

4. 推計結果に対する考察

1) 傾向スコアによるデータリサンプリングについて

表3にはバランス検定の結果を示した。「強く無視できる割当て」(Ignorable treatment assignment) 条件が成立している、つまりこの条件が満たされるような変数が選択されているかどうかを判断する基準である調整前のロジスティック回帰分析のc統計量は、0.76であり、目安とされる0.8^{註25)}に近い数値が得られた^{註26)}。また、バランス検定の結果、カリバーマッチングについては、すべての変数で処理群と対照群の平均値の差に有意な差が認められず、良好な結果が得られているものの、マハラノビスマッチングのいくつかの変数については、平均値の差において有意差が認められており良好とはいえない。

2) 生存時間分析結果の解釈および議論

生存時間分析の推計結果を表4に示した。図3に示すとおり、全数および要支援者群については、プログラムへの参加後、200日までイベント発生が増加し、200日以降徐々に減少する生存時間の分布状況を踏まえ、対数ロジスティック分布または対数正規分布を仮定したうえで、両モデルで推計を実施し、AIC基準および対数尤度から全数では対数ロジスティックモデル、要支援者群では対数正規分布モデルが適当と判断し、推計結果として示した。特定高齢者については、生存時間の分布が時間を通じて大きな変化はないこと、比例ハザード性が成立するかどうかのSchenfeld残差が時間に依存するかどうかの検定から、Schenfeld残差は時間に依存しておらず、ワイブルモデルも含めて推計を実施し、AIC基準および対数尤度からワイブルモデルが適当であると判断した^{註27)}。

次に、選択バイアスによる影響を把握するため、傾向スコアによって調整する前のすべての症例(N=14,002)を用いた場合と、傾向スコアによって調整された症例(N=7,724)を用いた場合の、Kaplan-Meier法による生存時間曲線の比較を図

註25) c統計量は、ロジスティック回帰分析の適合率を評価する指標として広く用いられているものの一つで、ROC (=Receiver Operating Characteristic Curve) 曲線下の面積(AUC)を求め、これが1に近いほど、一般的にはモデルの適合率が高いと判断する。なお、「強く無視できる割当て」条件が成立しているかどうかについては、星野・岡田(2006)、星野(2009)により0.8が目安とされている。

註26) なお、特定高齢者のサンプルで0.86、要支援者群のサンプルで0.75とカテゴリー別でも良好な結果であった。

註27) 実際、すべてのカテゴリーにおいて、ワイブル分布、対数ロジスティック分布、対数正規分布を仮定したAFTモデルによる推計を実施したが、結果、モデルによって、パラメータの推定量や有意変数に大きな違いは見受けられなかった。

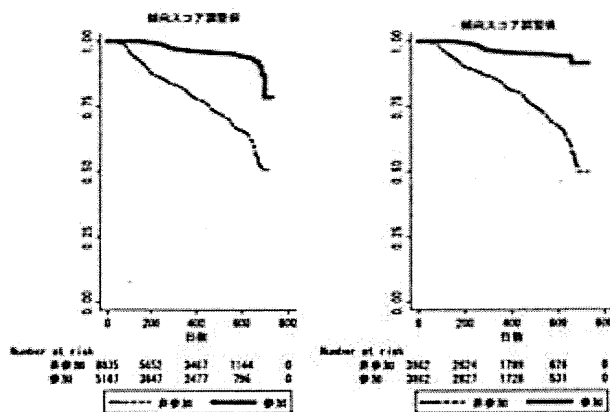


図4 Kaplan-Meier法による生存時間曲線比較 (参加対非参加)

4に示した。結果、調整後のデータを用いても容易に参加効果を確認することができる。また、運動器の機能向上プログラムへの継続的参加を示すタミー変数 (TRT) の調整前後のパラメータを比較した結果、調整前の症例を用いた場合が1.39であるのに対し、カリバーマッチによる調整後の症例を用いた場合は1.12と小さいことから、調整前の症例を用いた推計では選択バイアスによりパラメータの推計値が過大になっている可能性が示唆された。

次に、AFTモデルでは、推計されたパラメータ β を用いて、加速要素 (Acceleration factor = AF) を求めることができるが ($AF = \exp(\beta)$)、その結果から運動器の機能向上プログラムの参加の有無によって、現在の状態を伸ばす効果にどのくらいの差があるかを判断できる。他の条件を一定として、運動器の機能向上プログラムに継続的に参加した人 (TRT=1) とそうでない人 (TRT=0) でどのくらい現在の状態を維持できる日数を長くすることができるかを加速要素によって求めたところ、特定高齢者では1.6倍 ($\exp(0.472) = 1.603$)、要支援者群では3.1倍 ($\exp(1.141) = 3.127$) と推計された。このことから、特定高齢者では、参加した人とそうでない人では2倍弱、要支援者群では3倍強、それぞれリスクが発現する時間を延ばす効果があると考えられる。

最後に、パラメータ β を用いて生存時間関数に影響を与える変数について検討する。全体として、各カテゴリーで違いが見受けられるものの、プラス、マイナスといった各パラメータの符号条件はいずれのモデルにおいても違いはなく安定しており、GDSテスト (うつ) を除き、先に示した仮説と整合的である。全数ベースでの推計では、ものわすれテストの点数 (hds) が高い方が、基本チェックリストの点数 (checklist) が低い方が、運動器の機能向上プログラムを選択した (TRT=1) 方が、高い有意水準 (1%) で介護予防の効果が高いことが確認でき、脳血管障害 (brain) および認知症 (diments) の既往がある方が、効果が低いことを確認できる。特定高齢者の場合、運動器の機能向上プログラムに参加している方が、ものわすれテストの点数が高い方が同様に介護予防の効果を確認できる。要支援者群では、ものわすれテストの点数が高い方が、基本チェックリストの点数が低い方が、運動器の機能向上プログラムを選択した方が、独居で日常生活に役割を持つ高齢者の方が、高い有意水準 (1%) で介護予防の効果を確認でき、脳血管および認知症の既往がある方が、介護予防の効果が低いことが統計的に確認できる。

以上を総括すると、特定高齢者、要支援者群、全数のいずれのカテゴリーでも運動器の機能向上プログラムへの継続的参加による介護予防の一定の効果を確認するに至っており、特定高齢者では、参加した人とそうでない人では2倍弱、要支援者群では3倍強、それぞれリスクが発現する時間を延ばす効果があると考えられる。このようにカテゴリーによって介護予防の効果に違いがあることは、辻他 (2009) や大淵他 (2010) とも整合的な結果となっているが、その程度がどのくらいかは本論によって明らかにされた成果のひとつである。また、カテゴリーによって違いがあるものの、初期の基本チェックリストの点数やものわすれテストの点数、脳血管障害および認知症の既往歴が

表4 AFTモデルの推計結果

	全数 対数ロジスティック		特定高齢者 ワイブル		要支援者群 対数標準		(参考) 特定高齢者 対数ロジスティック	
	係数 (標準誤差)	係数 (標準誤差)	係数 (標準誤差)	係数 (標準誤差)	係数 (標準誤差)	係数 (標準誤差)	係数 (標準誤差)	係数 (標準誤差)
	gender	-0.080* (-2.05)	-0.134* (-2.58)	-0.225** (-3.30)	-0.202 (-2.22)	-0.081 (-1.83)	-0.073 (-1.23)	-0.262** (-3.29)
age	0.094* (2.29)	0.090 (1.48)	0.052 (0.46)	-0.019 (-0.12)	0.113* (2.53)	0.092 (1.50)	0.104 (0.84)	0.008 (0.04)
sq_age	-0.001* (-2.47)	-0.001 (-1.61)	-0.000 (-0.52)	0.000 (0.05)	-0.001** (-2.69)	-0.001 (-1.56)	-0.001 (-0.90)	-0.000 (-0.09)
single	0.190** (6.36)	0.225*** (4.40)	0.027 (0.41)	0.089 (1.04)	0.200*** (5.01)	0.199*** (3.49)	0.053 (0.71)	0.135 (1.34)
role	0.214*** (5.56)	0.160** (3.12)	0.022 (0.34)	0.042 (0.48)	0.236*** (5.47)	0.202*** (3.44)	0.050 (0.69)	0.088 (0.89)
cognitive	0.006 (1.49)	0.006 (1.07)	0.027** (3.69)	0.021* (2.16)	0.004 (0.81)	-0.001 (-0.09)	0.028*** (3.37)	0.019 (1.72)
gds	0.008 (1.40)	0.011 (1.44)	0.005 (0.43)	-0.001 (0.08)	0.007 (1.19)	0.012 (1.40)	0.004 (0.29)	-0.003 (-0.18)
hds	0.058*** (17.91)	0.056*** (12.26)	0.036*** (6.84)	0.043*** (5.56)	0.060*** (15.01)	0.057*** (10.94)	0.041*** (6.45)	0.046*** (5.14)
checklist	-0.051*** (-10.63)	-0.049*** (-7.50)	-0.014 (-1.53)	-0.011 (-0.97)	-0.059*** (-10.64)	-0.047*** (-6.21)	-0.012 (-1.20)	-0.011 (-0.80)
brain	-0.238*** (-5.56)	-0.321*** (-5.91)	-0.230** (-2.63)	-0.225 (-1.92)	-0.228*** (-4.73)	-0.301*** (-4.90)	-0.261** (-2.63)	-0.276 (-1.96)
rheumat	0.109** (2.75)	0.112* (2.15)	-0.094 (-1.21)	-0.145 (-1.39)	0.146** (3.29)	0.147 (2.49)	-0.149 (-1.74)	-0.210 (-1.75)
diments	-0.515*** (-8.28)	-0.278** (-2.83)	-0.063 (-0.52)	0.004 (0.02)	-0.566*** (-7.76)	0.532*** (4.65)	0.016 (0.12)	0.079 (0.40)
fracture	-0.042 (-0.98)	-0.092 (-1.60)	-0.102 (-1.10)	-0.194 (-1.67)	-0.023 (-0.48)	-0.081 (-1.28)	-0.130 (-1.24)	-0.223 (-1.59)
weaken	-0.077 (-1.37)	-0.026 (-0.32)	-0.041 (-0.47)	-0.008 (-0.07)	-0.079 (-1.21)	0.079 (0.80)	-0.049 (-0.48)	-0.002 (-0.01)
youshien	0.044 (0.70)	0.247 (3.44)						
TRT	1.391*** (24.98)	1.122*** (18.05)	0.661*** (10.39)	0.472*** (6.43)	1.487*** (25.26)	1.141*** (18.15)	0.703*** (4.47)	0.496*** (4.92)
_cons	2.075 (-2.47)	2.235 (-1.61)	3.375 (-0.52)	6.358 (0.05)	1.646 (-2.69)	2.430 (-1.56)	1.103 (-0.90)	5.136 (-0.09)
θ	0.348*** (-5.64)	0.309*** (-4.94)	0.533 (0.23)	0.638 (0.37)	0.397*** (-4.81)	0.332*** (-4.37)	0.361* (0.18)	0.311*** (0.30)
ρ			2.941*** (0.16)	2.843*** (0.21)				
γ	0.588*** (-26.58)	0.554*** (-21.30)					0.318*** (0.02)	0.336*** (0.03)
σ					1.200*** (7.95)	1.108*** (3.30)		
Frailty	あり	あり	あり	あり	あり	あり	あり	あり
マッチング方法	未調整	カリバー	未調整	カリバー	未調整	カリバー	未調整	カリバー
サンプル総数	14002	7724	2330	982	11672	6320	2330	982
対数尤度	-5570.2	-2595.3	-355.4	-201.5	-5136.0	-2232.7	-364.9	-212.6
カイ二乗値	2346.1	1073.2	217.5	105.6	2010.1	876.4	213.0	97.8
AIC基準値	11178.4	5228.6	746.9	438.9	10307.9	4501.5	765.8	461.2

* p<.05; ** p<.01; *** p<.001

介護予防の効果に影響を与えていることが統計的にも整合的である。に確認できたことは辻 他 (2009) の研究成果と

表5 PSM-DID推計結果

		全数	特定高齢者	要支援者群
調整前	処理群 (参加)	-0.071	0.313	-0.133
	対照群 (非参加)	0.134	2.164	0.084
	両グループ間の平均の差	-0.206 *	-1.851 ***	-0.217 *
	共変量調整 (OLS)	-0.339 ***	-1.774 ***	-0.281 **
基本 チェ ック リス ト 調整 後	Case1	-0.839 ***	-2.048 *	-0.496 **
	DID 推計 (バンド幅: 0.06) 95%ci.	(-1.02 - -0.34)	(-3.94 - 0.28)	(-0.85 - -0.24)
	Case2	-0.850 ***		-0.456 **
	DID 推計 (バンド幅: 0.01) 95%ci.	(-1.18 - -0.32)		(-0.74 - -0.14)
	Case3	-0.192 **	-2.250 ***	-0.453 **
	DID 推計 (バンド幅: 0.3) 95%ci.	(-1.03 - -0.40)	(-3.60 - -1.22)	(-0.76 - -0.18)
	Case4	-0.693 ***	-2.473 ***	-0.471 **
	DID 推計 (バンド幅: 0.8) 95%ci.	(-0.95 - -0.27)	(-3.82 - -1.19)	(-0.81 - -0.19)
	Case5	-0.780 ***	-2.000 *	-0.388 **
	DID 推計 (トリミング: 2%) 95%ci.	(-1.01 - -0.30)	(-4.00 - -0.03)	(-0.57 - -0.25)
	Case6	-0.600 ***	-1.904 *	-0.483 **
	DID 推計 (トリミング: 5%) 95%ci.	(-0.81 - -0.25)	(-3.99 - 0.37)	(-0.89 - -0.19)
	Case7	-0.655 ***	-1.713	-0.450 **
	DID 推計 (トリミング: 10%) 95%ci.	(-0.88 - -0.27)	(-3.69 - 0.48)	(-0.87 - -0.18)

* p<.05; ** p<.01; *** p<.001

共変量調整 (OLS) の推定式: $Y2 = \alpha + \beta X + \gamma TRT + \epsilon$ X は AFT モデルで用いた説明変数を成分にもつベクトルである。

3) PSM-DID推計結果に関する議論

推計結果を表5に示した。運動器の機能向上プログラムに参加した高齢者は、0.071点の改善⁽¹²⁸⁾がみられる一方、非参加高齢者は0.134点悪化しており、PSM-DIDによるバランス調整前の段階では、参加者群の平均改善が非参加者群のそれにくらべ0.206点上回っているにとどまっている。しかし、PSM-DIDによるバランス調整後の両者の平均差は0.839点と1点近い改善が見られ、バンド幅を変えても安定していること、また、95%信頼区間は-1.02から-0.34であり、95%信頼区間がすべて負の領域であることから、運動器の機能向上プログラムの参加がチェックリストの改善にポジティブな影響を与えている証左と考えられる。特定高齢者の場合で調整前の平均差が1.851点、調整後で2.048点(95%信頼区間では-3.94から0.28)、要支援者群では、調整前で0.217点、調整後で0.496点(95%信頼区間では-0.85から-0.24)である。特定高齢者の場合、全数および

要支援者群にくらべ、改善幅は大きいものの信頼区間にばらつきがあり、若干信頼性にかける推計となっている。また、バンド幅やトリミングを変えることによって推計の感度を確認したが、特定高齢者を推計した場合の一部を除き、運動器の機能向上プログラムへの継続的な参加の効果を有意に確認できる結果となっており、改善幅についても、トリミングを10%とした場合を除き、ほぼ同レベルの推定量が得られている。

5. 政策的含意 (インプリケーション)

本論は介護予防の効果を計量的に明らかにすることを目的に、運動器の機能向上プログラムへの参加の効果について、傾向スコアによる選択バイアスの調整を行ったうえで、生存時間分析およびPSM-DID推計により分析を行った。

結果、運動器の機能向上プログラムへの継続的な参加による介護予防の効果を生存時間分析により確認し、推計されたパラメータにより算出した加速要素から、現在の状態を伸ばす効果は、参加の有無によって特定高齢者で1.6倍、要支援者群で3.1倍異なると推計された。また、特に要支援

⁽¹²⁸⁾ 基本チェックリストは点数が少ないほど健康状態の良さを示すことから、マイナスが改善、プラスが悪化を示す。

者群では、ものわすれテストの点数、チェックリストの点数、脳血管障害や認知症といった既往歴などが介護予防の効果に影響を与えていることも確認できた。さらに、PSM-DID推計から、運動器の機能向上プログラムに1年を超えて継続的に参加した高齢者では、そうでない高齢者に比べ、基本チェックリストの点数で、特定高齢者で2.0～2.5点程度、要支援者群で0.45～0.5点程度改善しており、改善効果を確認することができた。

この研究のインプリケーションは以下に集約できる。高齢者が現在の健康状態を一日でも長く維持できることは、高齢者自身の生活の質の確保にとどまらず、喫緊の課題である介護費用の抑制にも効果的である。そのため、ひとりでも多くの高齢者が早期より介護予防に参加できるような枠組みを整備することが望まれる。要支援者になってからでも、介護予防に参加しない高齢者に比べ高い効果が本論からも確認されており、今後は、高齢者が気軽に介護予防に参加するための条件整備が急務である。

また、今回は運動器の機能向上プログラムの効果について確認したが、介護予防における運動器の機能向上プログラムの中には、マシンを利用した筋力トレーニングに主眼をおいたプログラムからグループ体操やレクリエーション・ゲームといったソフトな運動まであり、高齢者の状態によって、どのようなプログラムが適当であるのか、また、どのような頻度で行うのが適当であるか、他のプログラムとの併用は効果的であるかについて、今後、本格的な研究業績の蓄積が進むことが望まれる。

謝辞

本論文に関しては、本誌匿名のレフリーから細部にわたり助言および貴重なコメントをいただいた。この場を借りて深く感謝の意を表したい。また、「介護予防事業等の効果に関する総合的評価・分析に関する研究」の研究組織の構成員である植

田耕一郎日本大学歯学部教授、大久保一郎筑波大学大学院人間総合科学研究科教授、大野裕独立行政法人国立・精神神経医療研究センター認知行動療法センター長、大原里子東京医科歯科大学歯学部附属病院講師、杉山みち子神奈川県立保健福祉大学保健福祉学部教授、鈴木隆雄独立行政法人国立長寿医療研究センター研究所長、丹後俊郎医学統計学研究センター長、本間昭認知症介護研究・研修東京センター長、安村誠司福島県立医科大学医学部教授の各氏（五十音順）および曾根稔雅東北福祉大学健康科学部助教、また、東京都健康長寿医療センター研究所の三木明子研究員をはじめとする同センターの多数の研究員から多くのご教示およびコメントを頂戴した。この場を借りて深く感謝の意を表したい。にもかかわらず、コメントを活かしきれていない点や文中に残るすべての誤謬は、筆者3名の共同責任である。なお、ここでの考察のすべては筆者の個人的見解であり、筆者の所属する組織の見解を示すものではない。

なお、本研究は、平成20年度厚生労働省老人保健事業推進費補助金等による「介護予防事業等の効果に関する総合的評価・分析に関する研究」の一部として行われたものであり、本分析において使用したデータは、地域包括支援センターで収集され、同センターで匿名化されたうえ、厚生労働省を経由して研究班に提供された。また、本研究課題は、2008年10月24日、東北大学医学部倫理審査委員会承認されたものである。

参考文献

- 大久保一郎 (2009) 「介護予防サービスの費用対効果分析について」 <<http://www.mhlw.go.jp/shingi/2009/03/dl/s0326-12e.pdf>>2009年3月26日アクセス
 大洞修一、小島基水、三木明子、伊藤和彦、新井武志、辻一郎、大久保一郎、大原里子、杉山みち子、鈴木隆雄、曾根稔雅、安村誠司 (2010) 「介護予防対象者の運動器関連指標評価基準—介護予防ケアマネジメントのために」『日本公衆衛生雑誌』57 (11) : 988-995
 北村行伸 (2009) 「ミクロ計量経済学入門」日本評論社
 辻一郎、大洞修一、杉山みち子、植田耕一郎、大原里子、

- 安村誠司, 本岡昭, 大野裕, 鈴木孝雄, 大久保一郎, 丹後俊郎 (2009) 「介護予防事業等の効果に関する総合的評価・分析に関する研究 報告書」—介護保険制度の適正な運営・周知に寄与する調査研究事業—財団法人日本公衆衛生協会
- 星野崇弘, 岡田謙介 (2006) 「傾向スコアを用いた共変量調整による因果効果の推定と臨床医学・疫学・薬学・公衆衛生分野での応用について」『保健医療科学』55 (3) : 230-243
- 星野崇弘 (2009) 「調査観察データと統計科学—因果推論・選択バイアス・データ融合—」岩波書店
- Cameron AC and Trivedi PK (2005) *Microeconometrics: Methods and Applications*. Cambridge University Press.
- Cleves M, Gould W, Gutierrez RG and Marchenko YV (2010) *An Introduction to Survival Analysis Using Stata*. 3rd ed. STATA Press.
- Guo S, Barth RP and Gibbons C (2006) "Propensity Score Matching Strategies for Evaluating Substance Abuse Services for Child Welfare Clients." *Children and Youth Services Review*. 28 : 357-383.
- Guo S and Fraser MW (2009) *Propensity Score Analysis: Statistical Methods and Applications*. SAGE Publication Inc.
- Heckman JJ, Ichimura H and Todd P (1997) "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme." *Review of Economic Studies*. 64 (4) : 605-654.
- Heckman JJ, Ichimura H and Todd P (1998) "Matching as an Econometric Evaluation Estimator." *Review of Economic Studies*. 65 (2) : 261-294.
- Heinrich C, Maffioli A and Vazquez G (2010) "A Primer for Applying Propensity-Score Matching." *The Inter-America Development Bank. Impact-Evaluation Guidelines Technical Notes (No.IDB-TN-161)*.
- Kleinbaum DG and Klein M (2005) *Survival Analysis A Self-Learning Text*. 2nd ed, Springer.
- Leuven E and Sianesi B (2003) "PSMATCH2: Stata Module to Perform Full Mahalanobis and Propensity Score Matching, Common Support Graphing, and Covariate Imbalance Testing." *Statistical Software Components*. Vol. S432001.
- Rosenbaum PR and Rubin DB (1983) "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika*. 70 (1) : 41-55.
- Rosenbaum PR and Rubin DB (1985) "Constructing a Control Using: Multivariate Matched Sampling Methods That Incorporate the Propensity Score." *The American Statistician*. 39 (1) : 33-38.

(2010年11月12日受付, 2011年8月18日採用)

連絡先: 伊藤和彦
Kazuhiko_I_Ito@member.metro.tokyo.jp

How Effectively Does EXERCISE for the Elderly Work in Preventing Frailty? : Using Propensity Score Matching Methods to Evaluate the Effectiveness of Exercise Programs Aimed at Improvement of Physical Capabilities of the Potential Frail Elderly Persons

Kazuhiko Ito, M.A.¹⁾²⁾, Shuichi Obuchi, P.T., Ph.D.²⁾, Ichiro Tsuji, M.D., Ph.D.³⁾

Abstract

Based on the empirical findings from the observational studies conducted by the Ministry of Health, Labour and Welfare's specific study group (Tsuji *et al.*, 2009), we examine the effectiveness of continuous participation in the "Exercise Program aimed at Improvement of Physical Capabilities of Potentially Frail Elderly Persons" (EXERCISE), which was introduced to prevent frail elderly people from requiring a long-term caretaker in his/her aging daily living, using Survival analyses and Heckman's Difference in Differences (DID) estimation after the implementation of the Propensity Score Matching (PSM) method to avoid any harmful effect due to the selection bias.

As a result, we reconfirm the significant effect of continuous participation in EXERCISE with the result of survival analyses using an accelerated failure time model (AFT model). An estimated acceleration factor (AF), which describes "Stretching out" of the survival time when comparing the two groups (Participation and Non-Participation) and which is obtained by exponentiation of the estimated coefficient of TRT variable, giving an AF value of 1.6 for the "Person at High Risk of Frailty = PHRF (Tokuteikoureisha)" and an AF value of 3.1 for the "Person Requiring Support = PSR (Yoshiensha)" implies that the median survival time for the Participation group in PHRF is slightly less than double the survival time of the Non-Participation group and that in PSR is little more than triple, respectively. Moreover, we confirm that the estimated coefficients on predictor variables of the HDS-R Test score and on the dummy variables indicating the history of a certain disease obtained by the AFT parametric model indicate a significant effect on the results statistically.

In addition, DID estimations using the "Kihon-checklist" score indicate that the Kihon-checklist score for the Participation group of HRF increased by 2.0-2.5 units and by 0.45-0.5 units for PSR compared to that for the Non-Participation group of PHRF and that for PSR, respectively. Also, we found a significant effect on the score improvement of the Kihon-checklist in the both the estimation using samples for HRF and using for all samples.

Keywords : Prevention of Frailty, Exercise Program aimed at Improvement of Physical Capabilities for the Potentially Frail Elderly Person, Propensity score matching (PSM), Survival Analysis, Difference in Differences estimation

¹⁾ Office of Education of Tokyo Metropolitan Government

²⁾ Tokyo Metropolitan Institute of Gerontology

³⁾ Department of Public Health & Forensic Medicine, Toboku University Graduate School of Medicine