Pay-For-Performance and Selective Referral in Long-Term Care*

Toshiaki Iizuka† Haruko Noguchi‡ Shinya Sugawara§

August 21, 2017

Abstract

We examine how pay-for-performance (P4P) affects long-term care (LTC), exploiting a natural experiment in Japan. Unique matched user/care manager/provider data are used to observe care managers' referral decisions. Care managers/providers can vertically integrate, and P4P creates new incentives for selective referrals. Overall, we found no robust evidence that P4P improves LTC outcomes. However, after P4P, LTC outcomes improved more when care managers referred users to affiliated providers than to non-affiliated providers. Moreover, care managers referred users whose care levels were more likely to improve to affiliated providers. Vertical integration and a lack of risk adjustment appear to explain selective referrals.

JEL classification: I11, I18, J14, J33
Key words: pay-for-performance, agency problem, referral, vertical integration

* This work is supported by Health Labour Sciences Research Grant #201503002A. The use of the data in this study was approved by the Ministry of Health, Labour, and Welfare under this grant. Sugawara also acknowledges support from a Grant-in-Aid for Young Scientists (B) No. 15K17070 from the Japanese Ministry of Education, Science, Sports, Culture, and Technology and JST CREST. The authors thank Kensaku Kishida and the seminar and conference participants at the International Health Economic Association, the Japan Health Economic Association, Keio University, Waseda University, the National Institute of Population and Social Security Research, Technische Universität Dresden, the Tokyo Labor Economics Workshop, and the University of Tokyo for their helpful discussions and comments. All remaining errors are our own.
† Corresponding author. The University of Tokyo: toshi.iizuka@gmail.com.
‡ Waseda University: h.noguchi@waseda.jp.
§ Tokyo University of Science: shinya_sugawara@rs.tus.ac.jp.

1. Introduction

The rapid aging of the population means that long-term care (LTC) expenditure is expected to increase dramatically in the near future. Since 2000, LTC expenditure has grown by 9% annually, outpacing the growth of public medical expenditure (4% annually). Moreover, the proportion of people over the age of 80 is projected to increase from 3.9% of the population (2010) to 10% by 2050 (OECD 2013). LTC expenditure typically increases as users' care needs increase. Thus, if the government can help people stay healthy as they become older, this will not only help people to achieve a better quality of life but also reduce LTC expenditure. Thus, one of the important goals of public policy is to achieve "healthy aging."

Although no one objects to healthy aging, current LTC payments do not necessarily provide an incentive to achieve this goal. In many countries, LTC providers are commonly paid a fee-for-service (FFS), reflecting the expected cost of providing care. One important drawback of FFS is that if a provider successfully improves a user's care level, the provider's revenue is reduced because the user requires less care. Flexibly adjusting the skill level of caregivers or the hours they work is often difficult. Thus, by improving the level of care needed, the provider is likely to reduce its profit. One potential solution to this incentive problem is to pay fees based on the outcome of care rather than on an FFS basis.

The aim of this study is to investigate the effects of pay-for-performance (P4P) on LTC, focusing on adult day-care services. We exploit a natural experiment in Japan in which Shiga prefecture, one of 47 prefectures in Japan, introduced a bonus payment for LTC providers based on the outcome of care. Under this payment scheme, a provider receives a bonus payment in addition to nationally uniform FFS payments if it performs well in terms of improving or maintaining a user's care-need level. Following Shiga, several other local governments began experimenting with a similar P4P payment scheme,[1] and the Japanese government is now considering implementing P4P for the entire population.[2]

Although a P4P payment for LTC could potentially address the incentive problem inherent in FFS payments, very few governments and insurers have implemented such a payment (OECD 2013).[3] Accordingly, empirical studies of the impacts of P4P on LTC are quite scarce as we discuss later. Thus, we contribute to the literature by providing

[1] The timing of P4P experiments with day-care service in areas other than Shiga did not overlap with our research periods.
[2] Source: Newspaper article in Nihon Keizai Shimbun, November 11, 2016
[3] This is in stark contrast to increasing cases of P4P for medical care, as we discuss later.

empirical evidence of the effects of P4P on LTC.

We also contribute to the literature by studying how P4P affects user selection. Under a P4P payment, providers have an incentive to choose profitable users whose outcomes are more likely to improve. Thus, a common concern about P4P is the selection of profitable users.[4] An interesting feature of the Japanese LTC system is that care managers and LTC service providers can be under the same ownership, and care managers can legally refer users to an affiliated provider. This means that the new P4P payment created additional financial incentives for integrated care managers to selectively refer profitable users – those whose performance measures are more likely to be improved or maintained – to affiliated providers but not to non-affiliated providers. Importantly, some care managers have affiliated care providers but others do not, and, moreover, integrated care managers refer users to both affiliated and non-affiliated care providers. We exploit these variations to identify the impact of financial incentives on referral decisions.

For this study, we construct unique matched user-care manager-provider data in which care managers' referral decisions are directly observed. The data come from the universe of LTC claims data obtained from the Ministry of Health, Labour, and Welfare (MHLW). The data cover the vast majority of public LTC service uses in Japan, except for a small number of people who live in municipalities that do not permit the secondary use of data.[5] The data are monthly and cover the period between April 2012 and April 2014. We can track each user's service utilization along with the user's attributes, such as age, gender, and level of care needed. The claims data also contain provider attributes, including the size of the establishment. In addition to the claims data, we utilize another government survey on LTC providers that allows us to identify when an establishment provides both adult day-care services and care management services.

We start our empirical analysis by estimating the standard difference-in-difference (DID) model, comparing the outcomes in the treatment prefecture with those of the other 46 control prefectures before and after the introduction of P4P. This allows us to identify the overall effects of P4P at the population level.

We then attempt to identify selective referrals using the triple difference approach. Here, we exploit the fact that even within the treatment prefecture, care managers' referral incentives differ depending on whether the care manager has an affiliated LTC provider

[4] In fact, the Japanese government expressed concern that P4P for LTC may create an incentive to select profitable users and to avoid those whose disability is less likely to improve, arguing that implementing a performance-based payment scheme is difficult in the case of LTC. Please see the policy discussion in Okayama prefecture, http://www.kantei.go.jp/jp/singi/tiiki/sogotoc/sinsei/dai3/iken/ikenokayama.pdf (Accessed September 10, 2016).

[5] As of 2014, 1,606 of the 1,718 municipalities were included in the data.

to whom the manager refers the user. Specifically, independent care managers that do not have an affiliated provider will not be affected by P4P. Moreover, care managers who have an affiliated provider would refer more profitable users to the affiliated provider and not to a non-affiliated provider. The advantage of the triple difference approach is that it allows us to control for the shocks that coincide with the implementation of P4P in the treatment prefecture, which is not possible using the standard DID approach.

As an alternative method of identifying selective referral, we also examine how P4P affects the type of users that integrated care managers refer to their own providers. The day-care providers we interviewed agreed that improvement in the care-need level is more likely to occur for new users. One reason for this is that many users who are discharged from a hospital and who use day-care services have a "disuse syndrome" due to prolonged bed rest. For these users, there is a good chance that day-care services can improve their care-need level. Thus, using a discrete choice model, we examine whether care managers refer more new users, whose care levels are relatively more easily improved, to affiliated providers after P4P relative to non-affiliated providers.

At the population level, we find no robust evidence that P4P improves LTC outcomes. Moreover, even when we find a significant result in one specification, the impact of P4P is very small. Additionally, we do not find evidence that P4P affects day-care expenditure. Thus, the overall effect of the current P4P payment seems negligible, if it exists at all.

In contrast, for the analysis of selective referrals, we find that financial incentives affect the referral decisions of integrated care managers. After P4P, LTC outcomes improve more when care managers with affiliated providers refer users to affiliated providers than when they refer to non-affiliated providers. We also find that LTC outcomes improve more when care managers with affiliated providers refer users to their own providers relative to the average LTC outcomes of users referred by independent care managers. These results are indicative that care managers refer those whose conditions are more likely to be improved or maintained to affiliated providers. Additionally, a separate analysis that examines the types of users referred by integrated care managers reveals that integrated care managers are more likely to refer new users of LTC, whose care level is likely to be more easily improved, to affiliated providers. This result also suggests that financial incentives induce care managers to selectively refer profitable users to affiliated providers.

Overall, our results indicate that although P4P for LTC may have potential for improving LTC outcomes, one common concern about P4P—that P4P leads to user selection—cannot be dismissed. A lack of risk adjustment and vertical integration of care managers and providers appear to explain the selection. An effective P4P scheme would

need to address the drawbacks that are highlighted in this paper.

The literature on P4P for LTC is very scarce as we discussed previously and not conclusive.[6] A notable early contribution is Norton (1992), who finds that a bonus payment to improve care actually improved the disability levels of Medicaid nursing home residents during the 1980s. However, more recent studies by Werner et al. (2013) and Grabowski et al. (2016) both of which examined P4P for nursing homes in the United States fail to find any clear effects of P4P. None of these studies ask whether financial incentives result in user selection.

Far more studies have examined the effects of P4P in medical care. As reviewed by Eijkenaar (2013), many of these studies find that P4P has little or no effect on health outcomes. Studies have also found that P4P has unintended consequences, such as user selection (e.g., Chen et al. 2011), gaming (e.g., Gravelle et al. 2010), and worsened non-incentivized care (e.g., Mullen et al. 2010). Again, none of these studies examines the relationship between financial incentives and referral decisions.

This paper also contributes to the growing body of literature examining the agency problem created by the vertical integration of complementary services in medical care (e.g., Afendulis and Kessler, 2007, Iizuka 2007, 2012, Ho and Pakes 2014, Baker et al. 2016, and Chen et al. 2016). In particular, Ho and Pakes (2014) and Baker et al. (2016) examined the impact of hospital-physician integration on patient hospital choices, motivated by recent attempts to coordinate and integrate providers through accountable care organizations (ACOs) and health maintenance organizations (HMOs) in the United States. Their results indicate that integrated physicians tend to refer patients to hospitals at which the integrated party's profits are higher. Like these studies, we find that financial incentives created by vertical integration influence agents' referral decisions. Unlike these studies, we directly observe agent referral behavior, i.e., which care manager refers a user to which provider, enabling more direct inferences.[7] Moreover, whereas all previous studies examine physician agency in medical care, to the best of our knowledge, our study is the first to examine the agency problem in LTC service.

The remainder of the paper is organized as follows. In Section 2, we briefly discuss LTC insurance in Japan and the Shiga P4P payment scheme. Section 3 introduces our empirical model, and Section 4 describes our data. In Section 5, we report our estimation results. Section 6 discusses alternative explanations. We conclude the paper in Section 7.

---

[6] Please also see Norton (2017) for related literature.

[7] The above two studies observe patient hospital choices but not physician referrals and thus cannot directly examine agent choices.

2. Background

2.1. Japan's Public Long-term Care Insurance

In response to increasing demand for LTC, Japan implemented a mandatory public long-term care insurance (LTCI) system in 2000, which makes all people aged 65 and older eligible to receive services.[8] The insured are categorized into one of seven care-need levels based on a questionnaire on activities of daily living (ADLs). The seven care-need levels consist of two major categories of "assistance required" (*Yoshien,* hereafter AR) and "care required" (*Yokaigo,* hereafter CR).[9] Eligibility is reassessed every two years or every six months (for those who need a lower level of care) or as requested in the event of deteriorating or improving health status.

(Table 1 here)

The reimbursement depends on the FFS schedule standardized by the government, and beneficiaries need to pay a 10% co-insurance before they reach a ceiling amount. The benefits covered by the insurance depend on a recipient's designated care-need level. The monthly ceiling amounts are summarized in Column 1 of Table 1, where one unit denotes 10 JPY, with a slight regional variation. Any additional costs beyond the coverage need to be paid entirely by out-of-pocket expenses of the care recipients.[10]

Under the current LTCI scheme, service providers have little financial incentive to improve or even sustain the designated care-need levels of old persons. This is because the reimbursement the provider receives would be reduced if a client's care-need level improves. Flexibly adjusting the skill level of caregivers or the hours they work is often difficult. Thus, by improving the recipient's care-need level, the provider's profit is likely to decrease.

2.2. Service Provision and Care Management in LTCI

[8] People between the ages of 40 and 64 years can also be beneficiaries of LTC insurance if they become disabled because of geriatric causes, such as a stroke or dementia.

[9] Focusing on preventive LTC, AR consists of two groups: those who have no cognitive problems and need ADL support for 25–32 minutes (AR1) and for 32–50 minutes (AR2) per day. Likewise, CR includes five groups: those who require daily support for 32–50 minutes per day, which is the same criterion as that of AR, but who have some cognitive problem and/or dementia (CR1), followed by those who need care for 50–70 minutes, 70–90 minutes, 90–110 minutes, and 110 minutes and longer per day (CR2 to CR5, respectively).

[10] Additional costs are reduced for low-income households below the poverty line (i.e., annual income less than 3,000,000 JPY)

The introduction of LTCI created new markets for various formal elderly care services. Among them, this study focuses on adult day care (*Tsuusho Kaigo*), which is a major segment in the home-based care sector. In 2012, the day-care service was utilized by more than 30% of elderly people who used public LTC and accounted for 15% of the total LTCI expenditure in that year. The adult day-care service provides care in a facility during the daytime with the aim of reducing the burden of informal family care or maintaining and improving users' physical functions. The service includes transport to a facility, help with daily living such as meals and bathing, and functional training for daily life. Provider payments are determined based on the size of the provider and the hours of service provided. We provide a typical payment schedule in the second column of Table 1.

Under the Japanese LTCI system, a care manager plays a key role of coordinating LTC users and service providers.[11] An LTC service user first chooses a care manager, and then the care manager creates a menu of LTC services to be provided, called a *care plan*.[12] A care plan describes the services offered and the specific providers for all services. After creating a care plan, the care manager must visit the user at least once a month to monitor the plan and to update it if necessary. Whereas day-care providers are rewarded by an FFS payment, care managers are paid by capitation, with a fixed reward for each care plan regardless of the content of the plan.

Importantly, the government allows an establishment to provide both care-management and day-care services. In fact, many care managers work for establishments that offer both care-management and day-care services.[13] These care managers can legally refer profitable users to affiliated providers. We exploit this institutional detail in our analysis later.[14] We should also note that the government reduces the capitation rate mentioned above if a care manager refers more than 80% of users to a specific service provider.[15] Because of this rule, care managers who have an affiliated provider refer users

[11] Germany introduced a similar care-management mechanism, partly based on Japan's experience (Campbell et al. 2010). In addition, care management is commonly seen in the long-term care sector in developed countries when the elderly return home after hospitalization (Johri et al. 2003).

[12] Technically speaking, consumers can also create their own care plan. However, the vast majority of plans are made by care managers. Thus, we focus on users whose care plans are created by care managers in this study.

[13] As shown later, in our data, approximately half of all day-care services are referred by care managers who have an affiliated day-care provider.

[14] Using municipality-level data, Sugawara and Nakamura (2016) find a positive correlation between care manager density and care costs, which suggests the existence of care-manager-induced demand.

[15] In this case, capitation rate is reduced by approximately 20%. Please see Sugawara and Nakamura (2016) for more details.

not only to an affiliated provider but also to a non-affiliated provider. This variation also helps us to identify how financial incentives affect care managers' referral decisions.

2.3. P4P in Shiga Prefecture

Shiga prefecture implemented P4P for LTC for three years in the budgetary years 2012–2014.[16] This program provides a bonus payment to LTC providers in addition to nationally uniform FFS payments based on the outcome of care. Day-care providers that provide specific training programs (hereafter, "training") to improve physical and cognitive functions, nutrition status, and oral functions in a day-care center, day-care center with rehabilitation, or a day-care center for people with dementia are eligible for the bonus (Shiga Prefecture 2016).[17]

To be candidates for the P4P program, LTC providers must meet the following three requirements in the year in which they apply: (1) the average number of users per day is 10 or more for the year from January 1 through December 31; (2) 60% or more of users utilize specific training programs to improve their physical and cognitive functions, nutrition status, and oral functions; and (3) a score of 0.7 or more must be achieved based on the formula [the number of users who maintain their care-need level + the number of users who improve their care-need level × 2] divided by [the total number of users who use the above training programs for three months or longer and who have their care-need level re-evaluated by municipalities]. Importantly, there is no risk adjustment when calculating this score.

Additionally, the guideline states that only those people whose care-need level is re-evaluated by the end of October can be included in this calculation (Shiga Prefecture 2016). This implies that the P4P program is unlikely to have had an effect in the first year (i.e., 2012) because providers were notified about the initiation of the program in late September 2012 and had little time to respond to the program in the first year. For this reason, in our empirical analysis, we consider that the program, in effect, started in 2013. Each year, approximately 100 out of 500 day-care providers in Shiga applied to the P4P bonus. Based on the performance measures discussed above, Shiga prefecture identified top 20 service providers each year. These providers, along with the activities and

[16] The Japanese budgetary year starts on April 1 and ends on March 31.

[17] Prior to this policy in Shiga prefecture, the MHLW introduced a similar bonus payment for prevention in LTC in 2006, targeting users with AR categories (as opposed to CR categories in Shiga). The performance measures used in Shiga are the same as its predecessor. See Tsutsui and Muramatsu (2007) for more detail. Because the government introduced the 2006 policy nationwide, the impact of the policy is difficult to identify.

functional programs that they provide, were publicly announced.

The amount of the bonus was 10,000 JPY (approximately 100 USD) per capacity per month. This means that a provider with a facility capacity of 10 people received 1.2 million JPY (approximately 12,000 USD) per year. This is a considerable amount to compensate for the reduced reimbursement due to an improvement in users' care levels. For example, as seen in Column 2 of Table 1, if a user improves his/her care need by one level in a normal-sized firm, the FFS payment will be reduced by 106 units (1,060 JPY, approximately 10 USD) per person per day. Because day-care services are used ten times per month on average, the improvement will reduce FFS payments by approximately 100 USD per month. The P4P bonus of 100 USD is likely to compensate for the reduced FFS payments because the bonus will be paid for the capacity, whereas not all users will improve their CR level.

Shiga prefecture financed this program using funding set aside to stabilize the financial status of LTCI, called "*Kaigo Zaisei Anteika Kikin*." In 2011, the MHLW encouraged prefectures to utilize the funds effectively, and in response, Shiga decided to utilize the funds to implement the P4P payment to motivate service providers to sustain or improve the physical and cognitive functions of those insured with LTCI because they expected a rapid increase in LTC costs in the near future due to aging. According to the administrative officials, the decision was completely independent of and not influenced by, for example, the improvement or deterioration of its LTC outcomes relative to those of other prefectures.[18]

3. Empirical Model

3.1. Overall Effects of P4P

We first examine whether the implementation of P4P affects LTC outcomes at the population level. If P4P creates incentives to provide better care, we may find that P4P leads to better LTC outcomes. We also look at the effect of P4P on LTC expenditures. We identify the effects of P4P using a standard DID model, as follows:

$$y_{ijkt} = A_k + B_t + P4P_{kt}\delta_1 + Z_{jt}\delta_2 + X_{it}\delta_3 + \varepsilon_{ijkt}, \quad (1)$$

where i denotes a user, j denotes a provider, k denotes the prefecture in which j locates,

[18] Based on an interview by the authors with administrative officials of the Shiga Prefecture government.

and *t* denotes a month. Our sample consists of the elderly in the month when they utilize the day-care service. Because the P4P bonus applies to providers located in Shiga prefecture, we adopt the location of the provider rather than the user to define *k*.

The dependent variable y_{ijkt} takes either LTC outcomes or total day-care expenditures. The first outcome variable, LTC *Outcome Score*, is zero when the care-need level of the user worsens between month *t* and *t* + 1,[19] one when the level remains the same, and two when it improves. Note that this is the same performance measure as Shiga prefecture used in evaluating the performance of care providers. The second outcome measure, LTC *Outcome Dummy*, is a dummy variable that takes the value of one when the care-need level remains the same or improves and zero when it worsens. The third dependent variable is the total number day-care service units that *i* received at *t*.

Because the adult day-care service is provided only for care-need levels between CR1 and CR5, we do not include those elderly in categories AR1 and AR2 in our study. Transitions to death and to hospital are treated as a worsening transition (i.e., *Outcome Score* or *Outcome Dummy* = 0) regardless of the care-need level at time *t,* whereas a transition from CR to AR is treated as a transition to a better status (i.e., *Outcome Score* = 2, or *Outcome Dummy* = 1). Any other kind of attrition is treated as missing.[20] Attrition may include transitions to a state without a care-need level because we cannot distinguish these transitions from the other reasons for attrition, such as moving to a different municipality.

Among the independent variables, our main variable of interest is $P4P_{kt}$, which is equal to one if user *i* received a day-care service from a provider in Shiga after January 2013, when the incentive became effective, and zero otherwise. Z_{jt} and X_{it} are vectors of provider and individual characteristics, respectively. For Z_{jt}, we include a dummy variable for for-profit providers[21] and three dummy variables—normal, big (I), and big (II)—that correspond to the size of the provider.[22] The small size is the reference category. The individual characteristics, X_{jt}, include age, squared age, a male dummy, and dummy variables for care-need levels CR2–CR5, where CR1 is assigned as a reference category. A_k and B_t are prefecture and month fixed effects, and ε_{jjkt} is the error term. We allow

[19] In the claims data, any transition other than attrition happens at the end of a month.
[20] We also exclude information on users when they obtain exceptional-deemed (*minashi*) or transitional (*keikateki*) certification as their care-need levels.
[21] We categorize social welfare corporations (*shakai fukushi hojin*), a semi-public type of ownership, as part of not-for profit providers.
[22] The size of day-care providers is determined based on the average monthly number of users in the previous fiscal year (April to March). The size is small if the average number of users × days in the month is 300 or less, normal if between 301 and 749, big (I) if 751 to 900, and big (II) if 901 or more.

the error term to be correlated within a prefecture over time and employ an estimation for the clustered standard error.

To identify the policy effect via the DID approach, it is required that the average outcomes for the treated and control groups follow parallel paths over time without the treatment. In this context, our identification approach will not work if, for example, Shiga had a different time trend in LTC outcomes relative to those of other prefectures prior to the implementation of the payment scheme. However, Shiga introduced the P4P payment because the Long-Term Care Insurance Act permitted each prefecture government to utilize LTC funds in 2011, which had been put aside to stabilize the LTC finances of each prefecture. Thus, Shiga prefecture did not implement the P4P payment scheme because of its own worsening (or improving) LTC outcomes relative to those of other prefectures. Although we cannot rule out all potential concerns, we believe the endogeneity concern may not be severe in our case.

3.2 Financial Incentives and Selective Referral

3.2.1 Effects of P4P by Referral Types

A frequently raised concern about P4P is that it may create an incentive to select users to improve measured performance. We attempt to identify the presence of user selection by exploiting Japan's unique institutional setup. As discussed in Section 2, in Japan, a care manager refers a user to a service provider. Importantly, both care-management and day-care services can be offered under the same ownership. In this environment, P4P will create a new incentive for care managers with affiliated providers to engage in user selection by referring users who are more likely to improve or maintain their care level to affiliated providers as opposed to non-affiliated providers. If this is true, we may find that after the P4P payment, the LTC outcomes of these users improve more than those of individuals who are referred to non-affiliated providers.

To investigate this possibility, we examine the effects of P4P on LTC outcomes by referral types. Specifically, we distinguish the following three referral types:

1. *Integrated referral*: a care manager refers a person to a day-care provider under the same ownership.
2. *Non-integrated referral*: a care manager has a day-care provider under the same ownership but refers to a non-affiliated provider.
3. *Independent referral*: a care manager does not have a day-care provider under the same ownership and refers to any provider.

We attempt to identify selective referral using the triple difference approach. We first obtain standard DID estimates by referral types and then compare the DID estimates across referral types. We are particularly interested in two comparisons. First, we test whether the DID estimate for integrated referrals is larger than that for non-integrated referrals. Because care managers with affiliated providers decide whether to refer a user to an affiliated or a non-affiliated provider, the evidence that the DID estimate for integrated referrals is larger than that for non-integrated referrals is consistent with selective referral. Second, we examine whether the DID estimate for integrated referrals is larger than that for independent referrals. In the case of independent referrals, care managers' behavior is not affected by P4P because independent care managers do not have an affiliated day-care service. Thus, the evidence that the DID estimate for integrated referrals improves more than that for independent referrals supports the selective referral argument.

The triple difference approach has an advantage because it controls for various shocks in Shiga that coincide with the introduction of P4P and thus allow us to concentrate on policy effects that differ by referral types. For example, P4P may create an incentive to game the performance measure by manipulating the number of "eligible users," which is the denominator of the measure. Because all of the providers in Shiga are affected by the same incentive, a comparison by referral type is less likely to be affected by such a manipulation, if there is any. The identifying assumption of the triple difference approach is that there is no contemporaneous shock in Shiga that differentially affects providers by referral types in the same year as the implementation of P4P. This is not a strong assumption, and we are not aware of such shocks.

3.2.2 Types of Users Referred to Affiliated Providers

In Section 3.2.1, we attempted to infer selective referral from the differential effects of P4P by referral types. Alternatively, we can examine how P4P affects the types of users that integrated care managers refer to their own service providers. Again, we expect that after P4P, care managers would refer those whose care level is more likely to improve or maintain to affiliated providers. Unfortunately, our data do not contain detailed information on the user's medical or LTC conditions, and we cannot identify users with significant comorbidity or disability, as has been done previously (e.g., Chen et al., 2011). However, according to our interviews, there is a common notion among day-care providers that an improvement in the care-need level is more likely to occur for new

users.[23] One reason for this is that many users who are discharged from a hospital and who use day-care services have a "disuse syndrome" due to prolonged bed rest. For these users, there is a good chance that day-care services can improve their care-need level.

Empirically, we consider a discrete choice model in which a care manager with an affiliated provider refers a user either to an affiliated or a non-affiliated day-care provider. The dependent variable takes the value of one if the manager refers to an affiliated provider and zero otherwise. We exclude users who are referred by independent care managers from this analysis.

We employ two explanatory variables that capture the extent to which the user's care need level may be more difficult or easier to improve. The first variable, *NewLTCI*, takes one for user *i* in month *t* if it is the first month that the user used an LTCI service and zero otherwise.[24] The second explanatory variable, *Before65*, takes one if the user started using LTC services before age 65 and zero otherwise. As noted in Section 2, people can use public LTCI before age 65 only if they have one of 16 specific diseases and the doctor determines that an LTC service is required. These users typically have age-related diseases for a longer period; thus, their care-need level may be more difficult to improve.

The right-hand side of the linear probability model contains *NewLTCI*, *Before65*, their interaction terms with *P4P,* and all other explanatory variables in Equation (1). We are interested in the coefficients for the interaction terms. The selective referral argument will be supported by a positive (negative) coefficient for *NewLTCI*P4P* (*Before65*P4P*). As before, we allow the error term to be correlated within a prefecture over time.

3.3 Analysis of "Target" Providers

Because day-care providers must satisfy certain conditions to apply for the P4P bonus (please see Section 2), not all day-care providers may react to the incentive. Briefly, only those providers that provided training for more than 60% of their users are eligible for the P4P bonus. Because of this requirement, providers might have responded to the P4P incentive only if they had already provided training for close to 60% of their users before the policy implementation. Thus, we also provide a subsample analysis in which we focus on the "target" providers – those that provided training for more than 60% of their users in 2012 – and examine the effects of P4P on the users who received care from these providers. An important advantage of focusing on the "target" providers is that we can compare more homogeneous groups of day-care providers, which allows us to more

[23] This is based on our interviews with several day-care providers in Shiga prefecture.
[24] We constructed this variable by using all claims data after May 2009.

clearly identify the effects of the P4P incentive. In Section 5.4, we also report the results when we alter the threshold value for "target" providers from 60% to 40% of users in 2012.

4. Data

4.1. Description of Data

We conduct our empirical analysis using matched user-care manager-provider data. Our data set combines two sources, namely, claims data and an establishment survey. The claims data are the Survey of Long-term Care Benefit Expenditures, which is conducted by the MHLW. We obtain all claims in more than 90% of municipalities; the remaining municipalities do not permit the secondary use of data. As a result, more than 80% of all LTCI users are included in our data. All municipalities in Shiga permit the secondary use of data.

The claims data contain detailed information on the usage of LTCI services by individuals. The data are accompanied by each user's age, gender, and the care-need level for each month as well as the reason for stopping service utilization, such as death, hospitalization, or moving to institutional care. The data also contain an ID for a care management office and a service provider, but their affiliation status, which is required for the analysis by referral types, is not available from the claims data.

To obtain the affiliation status, we use an establishment survey, namely, the Survey on Institutions and Establishments for Long-Term Care, which is conducted annually by the MHLW. Each year, the MHLW sends a survey to all establishments that perform any home-based LTCI service, and the establishment lists all types of LTC services it provides. If the establishment provides both care management and day-care services, we determine that they are under the same ownership. Because this survey also contains the IDs of the care-management office and the day-care service provider, we can incorporate the affiliation status in our regression analysis.

Our data cover the periods between April 2012 and April 2014, where P4P became effective in January 2013. We chose to study this period because there was a major fee-schedule change of the LTCI in April 2012.[25] We focus on the elderly aged 65 or older,

[25] Although our regression analyses use data from April 2012, we utilize information from claims data as old as April 2009 to construct *Before65* and *NewLTCI*. Moreover, because claims data often arrive at the MHLW several months later, we use data up to April 2015 to incorporate any late arrival data in our data set.

all of whom are covered by LTCI.[26] Although we include all day-care service users in Shiga in the data set, for the users in all other 46 prefectures, we draw a 1% random sample to handle the large size of the group. We restrict our attention to the months when users received day-care services from only one provider. This is necessary to identify the association between the provider and LTC outcomes. Less than 15% of observations are excluded by this procedure.

Among the three types of providers that could obtain the P4P bonus (i.e., day-care providers, day-care providers with rehabilitation, or day-care providers for people with dementia), we concentrate on users who received care from day-care providers because they are by far the largest of the three. Moreover, the P4P bonus was only awarded to adult day-care services, which suggests that P4P may be more effective for day-care services than for the other two.

4.2. Descriptive Statistics

(Tables 2 and 3 here)

Table 2 reports the descriptive statistics for the variables used in the study. We separately show the statistics for Shiga and for all other prefectures. We have approximately 1.2 million observations, among which Shiga users constitute 24% of the sample. LTC outcomes are close to one for both *Outcome Score* and *Outcome Dummy*, and their values are similar in both the control and treatment groups. The distribution of care-need levels is also similar across the two groups. It also shows that day-care services are used across the board, although we have more users with lower care-need levels. In terms of referral types, approximately half of all referrals are categorized as independent referrals, and integrated referrals have somewhat larger share of the remaining half. New users of LTCI services and those who started using day care service early (i.e., before age 65) consist of approximately 1% and 3%, respectively.

Table 3 summarizes the monthly transition of care-need levels in Shiga and in all other prefectures before and after P4P. It indicates that approximately 97% of users maintain their care-need level in each month, whereas slightly more than 2% (less than 1%) of users worsen (improve) their care level. The upper block (1) shows the statistics for all providers, indicating that the transition probabilities in Shiga are very much the same before and after P4P. In contrast, in control prefectures, the transition to “worse”

[26] As discussed in Section 2, LTCI is also available for those between 40 and 64 years old when they have specific age-related diseases that require long-term care.

somewhat increased and the transition to "better" somewhat decreased after P4P. The differential transition patterns between Shiga and the control prefectures are even clearer in the lower block that focuses on users who received care from "target" providers. For these users, we find that the transition to "worse" ("better") decreased (increased) after P4P in Shiga. However, we find a completely opposite pattern in non-Shiga prefectures. These statistics appear to indicate that P4P had some positive effects on the measured performance in Shiga especially among those who received care from "target" providers.

(Table 4 around here)

In Table 4, we report LTC outcome measures in Shiga and in all other prefectures before and after the implementation of P4P. The upper block (1) presents outcome measures for all providers. It shows that both *Outcome Score* and *Outcome Dummy* are almost identical in both Shiga and in all other prefectures before and after the introduction of the P4P payment. At the same time, however, there is a small reduction in *Outcome Score* in all other prefectures between 2012 and 2013 as indicated by the significant mean difference between the two periods. These results indicate that at the population level, the impact of P4P on LTC outcome may exist, but it is likely to be very small if it exists.

In contrast, the lower block statistics for "target" providers show more significant changes after the introduction of P4P. In particular, *Outcome Score* for integrated referrals significantly improved in Shiga after 2013 (by approximately 0.008 point), whereas the same measure slightly declined in all other prefectures (by approximately -0.002 point). Thus, the DID measure, which takes the difference between these two (i.e., 0.008-(-0.002)=0.01), indicates that LTC outcomes of integrated referrals improved in Shiga after P4P relative to control prefectures. We can also calculate the DID measure for non-integrated and independent referrals as 0.003 and 0.004, respectively.

The triple difference estimator further takes the difference between the DID for integrated referrals and the DID for non-integrated (or independent) referrals. These mean differences, which are 0.007 (=0.01-0.003) and 0.006 (=0.01-0.004), respectively, are statistically significant (not reported), indicating that P4P had a greater positive effect for integrated referrals relative to non-integrated or independent referrals. These results are consistent with the view that P4P payment creates an incentive for care managers to refer users to affiliated providers whose care levels are more like to be improved or maintained. We also observe similar but weaker changes in the mean values before and after P4P for *Outcome Dummy*.

5. Estimation Results

5.1. Overall Effects of P4P on LTC Outcomes

(Table 5 here)

Table 5 reports the results of the regression analysis for Equation (1) using the full sample of users for all providers. In Columns (1) and (2), we find that *P4P* has a significant coefficient for *Outcome Score* but not for *Outcome dummy*. Additionally, the estimated impact for *Outcome Score* is very small relative to the mean of *Outcome Score*, which is 0.985 in Shiga. Thus, the P4P incentive may have some effect on LTC outcomes, but its impact for the entire population seems very small, if any. This result is consistent with the descriptive statistics reported in Table 4, where we observed a very small change in outcomes between Shiga and all other prefectures before and after P4P. Column (3) reports the results for total day-care units. The coefficient for *P4P* is not significantly different from zero, indicating that the current P4P does not affect total day-care spending. Thus, at the population level, the impact of P4P seems negligible.

The estimation results for other explanatory variables are as expected. We briefly report these results because they are primarily used as controls. Columns (1) and (2) report the results for LTC outcomes. In both columns, the coefficient for *Male* is negative and significant, indicating that males' care-need levels are more likely to worsen than those of females after controlling for age, care-need level, and other attributes. The coefficients for the dummy variables *CR2–CR5* are all positive and significant, and their impacts become larger as the care-need levels increase. This implies that a lower level of disability (e.g., *CR1*) is more likely to transition to a worse state than is a higher level of disability (e.g., *CR4*). This pattern of transition is also found in previous studies (e.g., Tajika and Kikuchi 2005). In contrast to the individual characteristics, the effects of provider characteristics on outcome scores are far less clear. The for-profit dummy does not have a significant coefficient in either column. Similarly, any of the provider-size dummies are statistically significant at the conventional level.

Column (3) reports the results for total day-care expenditure. The coefficient for *Male* is negative and significant, indicating that males use fewer day-care services than females do. The positive and significant coefficients for *CR2–CR5* in Column (3) indicate that day-care expenditures increase with higher disability, as expected. With regard to provider characteristics, the coefficient of the for-profit provider dummy variable is positive, which implies that for-profit providers supply more day-care services than not-for-profit

providers do. There is no clear pattern in terms of the effects of provider size on units of all day-care services.

5.2 Financial Incentives and Selective Referral

5.2.1. Effects of P4P by Referral Types

In this section, we examine whether the effects of P4P vary by referral types. Under the current Japanese LTCI system, care managers can legally refer more profitable users to an affiliated provider. Thus, we expect that in the presence of user selection, the effect of P4P on LTC outcomes would be better for integrated referrals than for non-affiliated referrals. Additionally, we expect that LTC outcomes would be better for integrated referrals than for independent referrals. In the following, we first consider the impact of P4P for the entire population (i.e., all providers) and then focus on the users who received care from "target" providers that are expected to be more sensitive to P4P.

Table 6 reports the results. To examine differential responses by referral types, we interact the *P4P* dummy with the three referral types and include them as regressors. To save space, only the coefficients for the interaction terms are reported in the table.

In the upper block, we report the DID estimates for *Outcome Score*. The Column (1) results for all providers indicate that after the implementation of P4P, *Outcome Score* improved somewhat in Shiga for integrated and independent referrals relative to all other prefectures. However, as indicated by the F-statistics reported in the table, the impact of P4P does not appear to differ by referral types; the F-statistics do not reject the null hypothesis that the coefficients for integrated and non-integrated (or independent) referrals are equal. Thus, for all providers, we do not find evidence that indicates selective referral.

In Column (2), we report the results when we narrow the sample to "target" providers. The differential effects of P4P by referral types are expected to be more pronounced because "target" providers would be more responsive to the P4P incentive. We find a positive and significant coefficient for the interaction term for all referral types. These DID estimates suggest that if we focus only on "target" providers, P4P improved LTC outcomes in any type of referrals in Shiga relative to all other prefectures. These results may reflect the increased provider efforts under P4P.

More importantly, the F-statistics shown in the table indicate that the coefficient for integrated referrals is larger than that for non-integrated referrals. The result that the DID

estimate for integrated referrals is larger than that for non-integrated referrals is consistent with selective referral by care managers. To increase the chance of receiving a P4P bonus, care managers with an affiliated provider might have referred users who are more likely to improve or maintain their care level to an affiliated provider and to refer others to non-affiliated providers. The F-statistics also indicate that the coefficient for integrated referrals is larger than that for independent referrals. This result also supports the selective referral argument because under the null hypothesis of no user selection, we would expect that LTC outcomes would be no different between integrated and independent referrals.

In the lower block, we report the results for LTC *Outcome Dummy*. The results are even stronger for the lower block. In Columns (1) and (2), we find that the impact of P4P is significantly larger for integrated referrals than for non-integrated or independent referrals. Again, these results provide supportive evidence that P4P encourages care managers to selectively refer profitable users to affiliated providers.

5.2.2. Types of Users Referred to Affiliated Providers

(Table 7 here)

In the previous section, we inferred selective referral from the differential effects of P4P on LTC outcomes by referral types. Another way to examine the presence of user selection is to directly investigate care managers' referral decisions. As we discussed in Section 3.4, we pay attention to the users whose care levels are more (or less) likely to be improved or maintained and examine whether care managers refer more (or fewer) of these users to affiliated providers after P4P.

Table 7 reports the results. To save space, we only report the coefficients for the interaction terms of interest (i.e., *P4P*NewLTCI* and *P4P* Before65*). Column (1) shows the results for all care managers with affiliated day-care providers. We find that the coefficient for *P4P *NewLTCI* is positive and statistically significant, indicating that P4P increases the probability of new users to be referred to affiliated providers by 4.5 percentage points. The coefficient for *Before65* interacted with *P4P* is negative and significant, indicating that those who started using LTC services long ago (i.e., before age 65) are less likely to be referred to affiliated providers after P4P by 4.6 percentage points. These results are consistent with the selective referral behavior of care managers.

In Column (2), we report the results when we estimate the model by focusing on care managers who are affiliated with a "target" provider. The results are qualitatively the same as the Column (1) results and indicate that the P4P payment encourages care managers to

refer more new users and fewer long-time users to affiliated providers. These results provide additional evidence that P4P creates an incentive for care managers to engage in selective referral.

5.4. Additional Analysis

5.4.1. Top 20 Announcement Effect

As we noted in Section 2.3, the P4P program not only rewards the top 20 providers but also publicly announces these winners. One may wonder whether the public announcement may also have an effect on the LTC outcomes independent of the effect of P4P. For example, the top 20 recognition may improve LTC outcomes by promoting the matching between providers and users or by allowing the provider to hire better care workers. To the extent that the top 20 recognition is correlated with referral types, our estimates by referral types may be biased. We attempt to control for this potential bias by including a dummy variable, *Top20*, that equals one if the provider was a top 20 provider in the previous year and zero otherwise. The *Top20* dummy is expected to capture the effect of both P4P and being recognized as a top 20 provider in the previous year.

(Table 8 here)

Table 8 reports the estimation results. We continue to find that even after controlling for the *Top20* dummy, the differential effects of P4P by referral types remain, especially for "target" providers. Specifically, the Column (2) results for "target" providers show that for both *Outcome Score* and *Outcome Dummy*, we continue to find that the effects of P4P are significantly larger for integrated referrals relative to non-integrated or independent referrals. This suggests that P4P affects a broad range of "target" providers, and they appear to engage in selective referral after the implementation of P4P. Note also that in all regressions, the top 20 dummy variable has a positive and significant coefficient, indicating that top 20 providers in the previous year tend to have a better outcome in the following year.

5.4.2. Alternative Threshold for "Target" Providers

This subsection provides a robustness check by experimenting with a different threshold for "target" providers: instead of the 60% threshold that we used in our main analysis, we

define "target" providers as day-care providers that provided training for more than 40% of users in 2012.

(Table 9 here)

First, we re-estimate the effects of P4P on LTC outcomes by referral types using the 40% threshold for "target providers." The first and second blocks of Table 9, Column (2) report the estimation results for the 40% threshold value. For reference, the baseline results using the 60% threshold are reported in Column (1). As shown in Columns (1) and (2), the results are qualitatively the same regardless of the threshold value. At both thresholds, we find that for both *Outcome Score* and *Outcome Dummy*, the coefficients for integrated referrals are positive and significant and their magnitudes are statistically larger than those for non-integrated or independent referrals. Thus, we consistently obtain an indication of selective referral for the alternative threshold value.

Second, we re-estimate care managers' referral decisions using the 40% threshold for "target" providers. As shown in the third block of Table 9, the results are also qualitatively the same regardless of the definition of the "target" providers.

6. Alternative explanations

In Section 5.2.1, we found evidence suggesting that care managers engage in selective referrals. That is, after P4P, LTC outcomes experiences greater improvement when care managers with affiliated providers refer users to affiliated providers than when they refer users to non-affiliated providers or compared to independent referrals. One alternative explanation for this result is that after P4P, care managers with affiliated providers made better use of information about the user's disability condition, helping improve the user's LTC outcomes. If true, the better LTC outcomes result from better coordination between affiliated care managers and care providers, not from selection.

We argue that our results do not support this alternative explanation for two reasons. First, as we discussed in Section 5.2.2, there is other evidence that P4P has led care managers with affiliated providers to refer those who are more likely to improve or maintain their care levels to affiliated providers and those who are not to non-affiliated providers. Second, if the better LTC outcomes result solely from better coordination, we would expect P4P to result in a net improvement in overall LTC outcomes because better coordination will not have external effects on other players. However, as we reported in Section 5.1, we did not find clear evidence that P4P improved LTC outcomes at the

aggregate level. This suggests that P4P has an external effect. Indeed, if we estimate equation (1) only for non-target providers—those which provided a training to fewer than 60% of users in 2012—the coefficient for P4P becomes negative for both *Outcome Score* and *Outcome Dummy* and weakly significant for *Outcome Score* (not reported). This further indicates that P4P has an external effect, contrary to the better coordination explanation.

One may also argue that one of our results from the comparison of integrated and independent referrals (please see Section 5.2.1) may be explained by endogenous integration. For example, P4P may encourage productive care managers and providers to vertically integrate. If true, this may bias our estimation results. We believe this endogeneity concern is not severe in our case for two reasons. First, Shiga prefecture announced upfront that for budgetary reasons, P4P payments will be implemented for only three years. Thus, it is unlikely that players responded to the incentive by changing organizational structure. Second, our data cover only the first two years since the introduction of P4P, which was announced at the end of the first year. Thus, very few providers would have had time to change ownership structure during our study period.

7. Conclusions

This study examined how P4P affects LTC, focusing on adult day-care services. The growth of LTC expenditure is projected to be substantial, and P4P for LTC could potentially improve welfare by providing better outcomes at lower costs. However, few governments and insurers worldwide have implemented such a policy so far, and we know little about the effects of P4P on LTC outcomes and expenditure. We exploited a natural experiment in Japan, where Shiga prefecture introduced an outcome-based bonus payment for adult day-care services in addition to nationally uniform FFS payments. We used unique user-care manager-provider matched data in which we directly observed who referred a user to whom.

At the population level, we found weak evidence that P4P affects LTC outcomes (i.e., care-required level), but the magnitude appeared to be very small, if any. Additionally, we did not find evidence that P4P affected total day-care expenditures. Thus, the overall effect of P4P on day-care services seems negligible.

We also studied whether providers engage in user selection. This has often been raised as a major concern by opponents of P4P. The Japanese example provides a unique testing ground because both a care-management service and a day-care service can be offered under the same ownership, and the P4P program is expected to provide a new

incentive for integrated managers to refer more profitable users to affiliated providers. We tested this hypothesis using our data and found that after P4P, LTC outcomes improved more when a care manager referred a user to an affiliated provider as opposed to a non-affiliated provider. We also directly examined care managers' referral choices and found that after P4P, care managers selectively refer those whose care levels are more likely to improve to their affiliated providers instead of non-affiliated providers. Both of these results are consistent with the selective referral hypothesis.

The fact that the primary effect of P4P was to change the distribution of outcomes among providers without clear overall improvements is concerning to policymakers. A major drawback of the current P4P payment is the lack of risk adjustment. Without risk adjustment, providers have a clear incentive to select users, and doing so allows them to achieve a better score without improving users' care levels. As in all P4P payments, properly risk-adjusting performance measure is critical. Such an attempt is extremely important for P4P for LTC as well.

This study also highlighted the potential problem of common ownership of complementary services in LTC services. Vertical integration of care management and day-care services creates an incentive to refer profitable users to own providers. This resembles the agency problem in medical care in which ownership of ancillary services such as laboratory and imaging services and pharmacies creates a conflict of interest. A remedy such as separating ownership by making care management independent of service provision would need to be considered to address this incentive issue.

References

Afendulis, Christopher C., and Daniel P. Kessler. 2007. "Tradeoffs from Integrating Diagnosis and Treatment in Markets for Health Care." *American Economic Review* 97 (3): 1013–20.

Baker, L. C., Bundorf, M. K., & Kessler, D. P. (2016). The effect of hospital/physician integration on hospital choice. *Journal of Health Economics*, 50, 1-8.

Briesacher, B. A., Field, T. S., Baril, J. L., and Gurwitz, J. H. (2009). Pay-for-performance in nursing homes, *Health Care Financing Review*, 30(3), 1-13

Campbell, J. C., Ikegami. N. and Gibson. M. J. (2010). Lessons from public long-term care insurance in Germany and Japan. *Health Affairs*, 29(1), 87-95.

Chen, B. K., Gertler, P. J., & Yang, C. Y. (2016). Physician ownership of complementary medical services. *Journal of Public Economics*, 144, 27-39.

Chen, T. T., Chung, K. P., Lin, I., and Lai, M. S. (2011). The Unintended Consequence of Diabetes Mellitus Pay‐for‐Performance (P4P) Program in Taiwan: Are Patients with More Comorbidities or More Severe Conditions Likely to Be Excluded from the P4P Program?. *Health Services Research*, 46(1), 47-60.

Eijkenaar, F., Emmert, M., Scheppach, M., and Schöffski, O. (2013). Effects of pay for performance in health care: a systematic review of systematic reviews. *Health Policy*, 110(2), 115-130.

Grabowski, D. C., Stevenson, D. G., Caudry, D. J., O'Malley, A. J., Green, L. H., Doherty, J. A., & Frank, R. G. (2016). The Impact of Nursing Home Pay for Performance on Quality and Medicare Spending: Results from the Nursing Home Value Based Purchasing Demonstration. *Health Services Research*.

Gravelle, H., Sutton, M., & Ma, A. (2010). Doctor behaviour under a pay for performance contract: treating, cheating and case finding?. *The Economic Journal*, 120(542), F129-F156.

Ho, K., & Pakes, A. (2014). Hospital choices, hospital prices, and financial incentives to physicians. The American Economic Review, 104(12), 3841-3884.

Iizuka, Toshiaki. 2007. “Experts’ Agency Problems: Evidence from the Prescription Drug Market in Japan.” *RAND Journal of Economics* 38 (3): 844–62.

Iizuka, Toshiaki. 2012. “Physician Agency and Adoption of Generic Pharmaceuticals” *American Economic Review* 102(6): 2826-2858.

Johri, M., Beland, F., and Bergman, H. (2003). International experiments in integrated care for the elderly: a synthesis of the evidence. *International Journal of Geriatric Psychiatry*, 18(3), 222-235.

Ministry of Health, Labor and Welfare (Various years) *Survey of Long-term Care Benefit*

Expenditures (Kaigo Kyuufu-hi Jittai Chousa, in Japanese), Tokyo, Japan

Ministry of Health, Labor and Welfare (Various years), *Survey on Institutions and Establishments for Long-Term Care (Kaigo Sabisu Shisetsu, Jigyousho Chousa, in Japanese)*, Tokyo, Japan

Ministry of Health Labour and Welfare, (Various years), Annual Report on LTCI (Kaigo Hoken Jigyou Joukyou Houkoku, in Japanese). Tokyo, Japan.

Ministry of Health, Labour and Welfare. (2003). *Perspective of long-term care for old persons in 2015: for establishing long-term care system to support the dignity of old persons* (2015 nen no Koreisha Kaigo: Koreisha no Songen wo Sasaeru Care no Kakuritu he Mukete). In Japanese. http://www.mhlw.go.jp/topics/kaigo/kentou/15kourei/. (accessed Aug 7, 2016).

Ministry of Health, Labour and Welfare. (2010). *Ceiling amounts of payment pertaining to an allowance for long-term care insurance by designated care level-documents for Social Security Council* (Kubun Shikyu Gendo Kijungaku ni tsuite). In Japanese. http://www.mhlw.go.jp/shingi/2010/05/dl/s0531-13d_17.pdf. (accessed Aug 7, 2016).

Ministry of Health, Labour and Welfare. (2014). *Ceiling amounts of payment pertaining to an allowance for long-term care insurance by designated care level-documents for Social Security Council* (Kubun Shikyu Gendo Kijungaku ni tsuite). In Japanese. http://www.mhlw.go.jp/file/05-Shingikai-12601000-Seisakutoukatsukan-Sanjikanshitsu_Shakaihoshoutantou/0000049257.pdf. (accessed Aug 7, 2016).

Mullen, K. J., Frank, R. G., & Rosenthal, M. B. (2010). Can you get what you pay for? Pay‐for‐performance and the quality of healthcare providers. *The Rand Journal of Economics*, 41(1), 64-91.

Nihon Keizai Shimbun, "Prime minister Abe mentioned long-term care will aim at recovery and independence by increasing rewards for improvement (in Japanese)" November 11, 2016

Norton, E. C. (1992). Incentive regulation of nursing homes. *Journal of Health Economics*, 11(2), 105-128.

Norton, E. C. (2017). Long-term care and pay-for-performance programs. ADBI Working Paper Series No. 680. Asian Development Bank Institute.

OECD/European Commission. 2013. *A good life in old age? Monitoring and improving quality in long-term care*. Paris: OECD Health Policy Studies, OECD Publishing. Available from: http://dx.doi.org/10.1787/9789264194564-en

Shiga prefecture. (2016). P4P for LTC providers in private sector. http://www.pref.shiga.lg.jp/e/lakadia/sidou/minnkansyudou_youkaigodokaizen.html. (accessed Aug 7, 2016).

Sugawara, S. and Nakamura, J. (2016) Gatekeeper incentives and demand inducement: An empirical analysis of care managers in the Japanese Long-Term Care Insurance program, *Journal of the Japanese and International Economies*, 40, 1-16

Tsutsui, T and Muramatsu, N. (2007). Japan's universal long-term care system reform of 2005: Containing costs and realizing a vision. *Journal of the American Geriatrics Society* 55(9): 1458–1463.

Werner, R. M., Konetzka, R. T., and Polsky, D. (2013). The Effect of Pay‐for‐Performance in Nursing Homes: Evidence from State Medicaid Programs. *Health Services Research*, 48(4), 1393-1414.

Care-need level	Upper bound of units/month (up to March 31, 2014)	Day-care units/day: Normal size, 5-7 hours
CR1	16580	602
CR2	19480	708
CR3	26750	814
CR4	30600	920
CR5	35830	1,026

Table 1: Upper bound of monthly units and typical payment schedule for day-care providers.
Source: All-Japan Federation of National Health Insurance Organizations (https://www.kokuho.or.jp/system/lib/20150331_korosho_jimurenraku1-2-2-1.pdf, accessed June 1, 2016)

		All providers			
		Shiga		Not Shiga	
		Mean	S.D.	Mean	S.D.
Dependent variables	Outcome score	0.985	0.172	0.987	0.178
	Outcome dummy	0.978	0.148	0.978	0.148
	Units of day care	8452.2	5148.3	8586.1	6005.7
Care-need levels	CR1	0.337	0.473	0.353	0.478
	CR2	0.304	0.460	0.303	0.460
	CR3	0.186	0.389	0.178	0.382
	CR4	0.112	0.315	0.106	0.307
	CR5	0.062	0.241	0.060	0.238
Age		84.1	8.0	83.5	8.2
Male		0.309	0.462	0.315	0.465
New user of LTCI		0.011	0.105	0.016	0.126
New user of day care (within 6 months)		0.133	0.340	0.145	0.352
User of LTCI before 65		0.032	0.177	0.037	0.189
Referral type	Integrated	0.269	0.443	0.275	0.447
	Non-integrated	0.252	0.434	0.199	0.399
	Independent	0.479	0.500	0.526	0.499
Ownership	For-profit	0.335	0.472	0.420	0.494
Provider size	Small	0.258	0.438	0.241	0.428
	Normal	0.651	0.477	0.605	0.489
	Big (I)	0.077	0.267	0.088	0.283
	Big (II)	0.008	0.089	0.064	0.245
#Observations		282,168		895,491	

Table 2: Descriptive statistics

(1) All providers		Shiga		Not Shiga	
		2012	2013–	2012	2013–
#Observations		105,904	176,264	331,434	564,057
Transition of care-need levels(%)	Worse	2.24	2.22	2.22	2.25
	Same	97.01	97.02	96.79	96.84
	Better	0.75	0.74	0.98	0.91
(2)Target providers		Shiga		Not Shiga	
		2012	2013–	2012	2013–
#Observations		21,621	39,670	110,853	190,547
Transition of care-need levels(%)	Worse	2.27	2.12	2.01	2.08
	Same	97.03	97	96.87	96.93
	Better	0.7	0.88	1.12	0.99

Table 3. Monthly transition of care-need levels

All providers		Shiga					Not Shiga				
		2012	2013–	Difference (2013–2012)		#Observations	2012	2013–	Difference (2013–2012)		#Observations
Outcome score		0.985	0.985	0.000		105,904	0.988	0.987	−0.001	**	331,434
Outcome dummy		0.978	0.978	0.000		105,904	0.978	0.978	0.000		331,434
Target providers											
Outcome score	Integrated	0.978	0.986	0.008	**	5,332	0.988	0.986	−0.002	*	30,826
	Non–integrated	0.985	0.987	0.002		5,460	0.992	0.991	−0.001		22,532
	Independent	0.987	0.989	0.002		10,829	0.992	0.990	−0.002	**	57,495
Outcome dummy	Integrated	0.973	0.978	0.005	*	5,332	0.978	0.977	−0.001		30,826
	Non–integrated	0.979	0.979	0.000		5,460	0.980	0.980	0.000		22,532
	Independent	0.979	0.979	0.000		10,829	0.981	0.980	−0.001		57,495

Table 4. Descriptive statistics for LTC outcomes

Note: The upper block shows statistics for all providers. The lower block shows statistics for target providers. *** p<0.01, ** p<0.05, * p<0.1.

		(1)		(2)		(3)	
		Outcome score		Outcome dummy		Units of day care	
		Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
P4P		0.0007**	(0.000)	0.0002	(0.000)	10.1322	(23.672)
Age		0.0020***	(0.000)	0.0011***	(0.000)	356.0953***	(47.198)
Squared age		0.0000***	(0.000)	0.0000***	(0.000)	−1.9818***	(0.289)
Male		0.0046***	(0.000)	0.0043***	(0.000)	−679.6505***	(43.386)
Care-need level	CR2	0.0107***	(0.000)	0.0063***	(0.000)	1,772.1569***	(55.652)
(Reference: CR1)	CR3	0.0176***	(0.001)	0.0105***	(0.001)	3,938.0705***	(161.945)
	CR4	0.0263***	(0.001)	0.0151***	(0.001)	4,942.7725***	(268.411)
	CR5	0.0308***	(0.001)	0.0238***	(0.001)	5,607.0516***	(403.214)
For profit		−0.0005	(0.000)	−0.0004	(0.000)	1,075.4149***	(206.024)
Provider size	Normal	−0.0010*	(0.000)	0.0000	(0.000)	−553.4761***	(98.543)
(Reference: small)	Big (I)	−0.0007	(0.001)	0.0010	(0.001)	−174.8431*	(100.201)
	Big (II)	0.0005	(0.001)	0.0016*	(0.001)	−197.5948	(184.356)
Constant		0.9277***	(0.009)	0.9524***	(0.007)	11,097.6326***	(1,870.846)
Observations		1,177,659		1,177,659		1,177,659	

Table 5: Regression results for users in all providers. Note: Prefecture dummies and month dummies are included but abbreviated. Cluster standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Outcome score		All providers		Target providers	
		(1)		(2)	
P4P	P4P * Integrated	0.0008*	(0.000)	0.0073***	(0.001)
	P4P * Non-integrated	0.0000	(0.000)	0.0043***	(0.001)
	P4P * Independent	0.0011***	(0.000)	0.0048***	(0.001)
#Observations		1,177,659		362,691	
F-statistics					
(P4P * Integrated) - (P4P * Non-integrated) = 0		2.02		9.79***	
(P4P * Integrated) - (P4P * Independent) = 0		0.38		7.96***	
Outcome dummy		All providers		Target providers	
		(1)		(2)	
P4P	P4P * Integrated	0.0010**	(0.000)	0.0041***	(0.001)
	P4P * Non-integrated	-0.0001	(0.000)	0.0024***	(0.001)
	P4P * Independent	-0.0000	(0.000)	0.0013*	(0.001)
#Observations		1,177,659		362,691	
F-statistics					
(P4P * Integrated) - (P4P * Non-integrated) = 0		8.62***		5.02**	
(P4P * Integrated) - (P4P * Independent) = 0		9.61***		17.40***	

Table 6: Regression results by referral types

Note: Only the coefficients for P4P-related variables are reported. Cluster standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Dependent variable	Integrated		Integrated, target	
	(1)		(2)	
P4P * Before65	−0.0457**	(0.017)	−0.0896***	(0.025)
P4P * NewLTCI	0.0448***	(0.006)	0.0218**	(0.010)
#Observations	571,755		162,733	

Table 7: Regression results for care manager's referral decision

Note: Only the coefficients for the interaction terms are reported. The dependent variable equals one if a care manager with an affiliated provider refers the user to the affiliated provider and zero otherwise. Column (1) examines the decisions of all care managers with affiliated providers. Column (2) shows the same for "target" providers. Cluster standard errors in parentheses, *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Outcome score		All providers		Target providers	
		(1)		(2)	
P4P	P4P * Integrated	0.0006	(0.000)	0.0070***	(0.001)
	P4P * Non-integrated	-0.0005	(0.000)	0.0027**	(0.001)
	P4P * Independent	0.0006	(0.000)	0.0036***	(0.001)
Top20		0.0092***	(0.000)	0.0067***	(0.000)
#Observations		1,177,659		362,691	
F-statistics					
(P4P * Integrated) - (P4P * Non-integrated) = 0		4.68**		21.55***	
(P4P * Integrated) - (P4P * Independent) = 0		0.00		14.67***	
Outcome dummy		All providers		Target providers	
		(1)		(2)	
P4P	P4P * Integrated	0.0009**	(0.000)	0.0039***	(0.001)
	P4P * Non-integrated	-0.0005	(0.000)	0.0014	(0.001)
	P4P * Independent	-0.0003	(0.000)	0.0006	(0.001)
Top20		0.0051***	(0.000)	0.0041***	(0.000)
#Observations		1,177,659		362,691	
F-statistics					
(P4P * Integrated) - (P4P * Non-integrated) = 0		12.29***		11.62***	
(P4P * Integrated) - (P4P * Independent) = 0		12.69***		25.21***	

Table 8: Regression results with Top 20 providers

Note: Top 20 providers dummy takes one if the provider was recognized as a top 20 provider in the previous year and zero otherwise. Cluster standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

			Baseline target (60% in 2012)	Alternative target (40% in 2012)
			(1)	(2)
Outcome score	Overall		0.0053***(0.001)	0.0043***(0.001)
	Referral	P4P*Integrated	0.0073***(0.001)	0.0064***(0.001)
		P4P*Non-integrated	0.0043***(0.001)	0.0031***(0.001)
		P4P*Independent	0.0048***(0.001)	0.0040***(0.001)
F-statistics				
(P4P * Integrated) - (P4P * Non-integrated) = 0			9.79***	14.41***
(P4P * Integrated) - (P4P * Independent) = 0			7.96***	9.86***
Outcome dummy	Overall		0.0023***(0.001)	0.0016**(0.001)
	Referral	P4P*Integrated	0.0041***(0.001)	0.0038***(0.001)
		P4P*Non-integrated	0.0024***(0.001)	0.0016**(0.001)
		P4P*Independent	0.0013*(0.001)	0.0007(0.001)
F-statistics				
(P4P * Integrated) - (P4P * Non-integrated) = 0			5.02**	10.39***
(P4P * Integrated) - (P4P * Independent) = 0			17.40***	25.88***
	#Observations		362,691	407,420
CM's decision		P4P * Before65	-0.0896***(0.025)	-0.1011***(0.023)
		P4P * NewLTCI	0.0218**(0.010)	0.0361***(0.009)
	#Observations		162,733	184,087

Table 9: Regression results for alternative definitions of target providers

Note: Only the coefficients for *P4P* using samples of target providers are reported. Column (1) shows the baseline results under the original definition. Cluster standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.