博士論文

Essays on Health Economics

（医療経済学に関する研究）

髙橋　雅生

Acknowledgements

I am grateful to Toshiaki Iizuka for his generous support throughout this project. I thank Daisuke Adachi, Rong Fu, Konan Hara, Hidehiko Ichimura, Takuya Ishihara, Yoko Ibuka, Yuki Ito, Kyogo Kanazawa, Keisuke Kawata, Daiki Kiyama, Yasuki Kobayashi, Hiroki Matsui, Haruko Noguchi, Seojung Oh, Hiroki Saruya, Hitoshi Shigeoka, Reo Takaku, Yuta Toyama, Gyo Uchida, Hideo Yasunaga, and Kohei Yata for thoughtful discussions and comments. I also thank participants at Applied Econometrics Conference 2018, AWEHE 2019, Asian Meeting of the Econometric Society 2019, Health Economics Workshop at GRIPS, Hitotsubashi University, IHEP, JEA Spring Meeting 2019, Waseda University, and WEAI 2019 for helpful comments. Fundings from Japan Society of Promotion of Science and Institute of Health Economics and Policy are gratefully acknowledged. All errors are my own.

Contents

Chapter 1

Overview

This doctoral dissertation consists of three empirical studies on health economics related to healthcare and long-term care system in Japan. As in other developed countries, the total health care expenditure in Japan has skyrocketed mainly due to the rapid aging of the population. To make health care systems sustainable, it is increasingly important to enhance the efficiency of the health care system and make effective use of limited resources. The purpose of this dissertation is to clarify the incentive structure of people in the healthcare and long-term care system and to contribute to building a more efficient social security system in the future.

Chapter 2 investigates how financial incentives affect hospitals' inpatient care by exploiting the introduction of a new payment system in Japan. The new payment system called DPC/PDPS (DPC) has a feature of per-diem prospective payment and it is applied to large hospitals as a replacement for the traditional fee-for-service (FFS) payment system. Using patient-level discharge data of circulatory disease during 2008-2009, I find that DPC reduces LOS and medication, but the surgical procedure is not affected. I also show that the number of discharged patients significantly increase right before the drop of per-diem reimbursement. Furthermore, I find that the health effect of DPC is negligible.

Chapter 3 addresses the empirical question that how does the generosity of social insurance coverage affects the demand for healthcare and health outcomes for the elderly. I examine the effect of insurance coverage on long-term care utilization and its health consequences using novel administrative data on public long-term care insurance (LTCI) system in Japan. The LTCI coverage is determined by a standardized health index and there exist multiple thresholds that generate discontinuous changes in the insurance coverage limit. Using a regression discontinuity design, I find that the coverage expansion significantly increases recipients' long-term care utilization even without changing the prices they face. I also find that utilizing more long-term care has little effect on health outcomes. Together, these results suggest that generous LTCI coverage could induce excessive utilization mainly due to behavioral biases without having health benefits.

Chapter 4 examines the economic consequences of the manipulation of social insurance benefits. This chapter is co-authored with Takuya Ishihara. We show that there is a bunching in a distribution of the health score determining a benefit level of LTCI in Japan. The observed distribution suggests that recipients of LTCI receive more generous benefits due to the manipulation of the score. To quantify the impact of manipulation on long-term care expenditure, we develop a nonparametric partial identification framework exploiting general shape restrictions on counterfactual distribution. We find that the manipulation increases annual long-term care expenditure by at least 629 USD per recipient.

Chapter 2

How Does Financial Incentive to Hospitals Affect Inpatient Care? Evidence from Per-Diem Prospective Payment

2.1 Introduction

In many developed countries, national health care expenditure has skyrocketed because of the rapid aging of the population. Under these circumstances, it is important to construct an effective reimbursement system that incentivizes physicians and hospitals to reduce the excessive provision of health care. Fee-for-service (FFS) is a traditional payment system that is designed to reflect the amount of medical treatment performed. This payment system has been criticized as leading to excessive medical treatment because it does not generally have a mechanism to prevent excessive medical treatment and reinforces "physician's induced demand". The pioneering payment system for reducing medical costs is the prospective payment system (PPS) introduced in 1983 in the US. Under the PPS, reimbursement level is fixed for each diagnostic related group (DRG) regardless of the cost of treatment, so healthcare providers have the incentive to reduce health care costs to increase their profits. Nowadays, the PPS-type payment system is adopted in many countries.

Japan, one of the world's fastest aging countries, is not an exception. The Japanese government started replacing traditional FFS with a new payment system called Diagnosis Procedure Combination / Per-Diem Prospective Payment (DPC/PDPS, just DPC hereafter) in 2003. DPC has a declining per-diem prospective payment according to the length of stay (LOS), rather than a fixed payment for each hospital stay like under the PPS. The main reason why a declining per-diem payment was adopted by DPC is to prevent an abrupt change of medical treatment in hospitals as well as to encourage hospitals to contain healthcare costs. As explained later, some studies showed that a fixed payment for an episode in the US not only reduced medical expenditure but also worsened patients' health conditions because of its strong financial incentive. Therefore, it can be expected that a milder incentive such as declining per-diem payment could reduce medical expenditure without harming patient's health. Evaluating the policy impact of DPC can thus illuminate how the weaker financial incentive compared with PPS works for cost containment and patient's health. This has an important policy implication for the healthcare sector because policymakers care about both medical expenditure and people's health conditions. Also, because of its aging society and increasing medical expenditure, it is critical to understand whether the new payment system is effective for reducing excessive medical expenditure in Japan.

I investigate how the implementation of DPC influences LOS, medical treatments, and patient's health for patients with circulatory disease. The unique feature of DPC is that the per-diem prospective payment rapidly declines according to LOS. Thus hospitals would be motivated to reduce LOS in response to DPC. Although per-diem payments have weaker incentives for cost reduction compared with PPS, it can also reduce the quantity of medical input because of its partial prospective payment scheme. Another important feature of DPC is that prospective payment is only applied to some components of medical procedures and other procedures are still reimbursed by FFS payment. The prospective payment component covers the cost of basic hospital fees, pharmaceuticals, injections, examinations, diagnostic images, and procedures worth less than 10,000 yen. The FFS component covers surgical procedure, materials used for surgery, anesthesia, endoscopies, radioactive treatment pharmaceuticals, and medical teaching as well as procedures worth more than 10,000 yen. This payment structure might cause a heterogeneous impact on different medical procedures. To examine this possibility, I investigate the impact of DPC on the procedures covered by the prospective payment component and FFS component respectively. It is also critical to examine the adverse health effects of DPC to check whether the system could prevent an abrupt change of medical treatment which would harm patients' health. Therefore we examine the effect of DPC implementation on patient's health using the death rate in hospitals and whether patients' health condition at the time of discharge is better than admission time.

To identify the effect of DPC on medical treatment and patients' health, I exploit hospitals' different timing of adoption of DPC. Except for university hospitals and advanced treatment hospitals, whether and when to adopt DPC is left to the decision of each hospital. The number of hospitals that adopt DPC has increased rapidly and 29.6 percent of all hospitals (54.8 percent of all beds) have adopted the new payment system by 2018. In particular, many hospitals adopted DPC around 2009. Focusing on hospitals that adopted DPC in 2009 or 2010, I estimate the effect of the new payment system on medical treatments and patients' health which is related to inpatient care by difference-in-differences (DID) framework. In this way, I can compare hospitals that are willing to adopt DPC and whose timing of the adoption differs by only one year.

I utilize patient-level discharge data of circulatory disease between 2008 and 2009, which enables us to control detailed patient characteristics for avoiding estimation bias. For example, hospitals might select patients who are more profitable under DPC when the reimbursement system change. To exclude this sort of estimation bias, we can control the detailed diagnosis and severity condition of each patient which is contained in our data. Another advantage of this dataset is to contain "FFS-equivalent" reimbursement after adopting DPC. This allows me to compare the medical treatments under FFS and those under DPC with comparable monetary measures (reimbursement under FFS).

Using the same dataset, I also investigate how the reduction of per-diem payment affects hospitals' decisions on LOS. If hospitals respond to the reduction of per-diem payment, the number of patients who are discharged right before the reduction should be larger than the other day of LOS. To examine whether this is the case, I exploit the change from FFS to DPC. Under Per-diem payment under FFS decreases on 15th day and 31st day of LOS for all patients, while the timing of reduction of per-diem payment under DPC varies among specific diagnostic groups (DPC group). Therefore, the change from FFS to DPC eliminates the reduction of per-diem payment on the 15th and 31st day of LOS. I apply a "bunching" estimation proposed by Saez (2010) to examine how the number of patients who are discharged on the 14th and 30th day of LOS (just before the reduction of per-diem payment under FFS) changes due to the introduction of DPC.

The main findings of this paper are the followings; First, the change from FFS to DPC reduced LOS and average input of medication significantly. These results indicate that per-diem prospective payment of DPC have enough financial incentives for hospitals to reduce the cost of health care service concerning LOS and medication. Second, whether to conduct surgical procedures is not affected by DPC. Therefore, hospitals do not replace surgery with other medical treatments for exploiting fee-for-service payment of surgical procedures. Third, we do not observe the adverse health effects of DPC, which is favorable as an alternative payment system. Finally, the number of discharged patients decreases due to the elimination of the reduction of per-diem payment, especially for those whose LOS is relatively short. This suggests that hospitals respond to marginal payment by controlling LOS.

There are many studies on the impact of the introduction of PPS in the US. One of the influential studies is Randall and McGuire (1996), which examines moral hazard, selection and practice-style effects in the PPS. Cutler (1995) shows that the timing of adverse medical outcomes occurred earlier after the implementation of PPS. Dafny (2005) examines the price change after PPS implementation and shows that hospitals respond to the price change by "upcoding" rather than a change in medical practice. In terms of DPC in Japan, Wang (2010) and Nawata and Kawabuchi (2010) study the effect of DPC on LOS. Both studies are based on a comparison of before and after treatment so that estimates may be biased by time trend. Shigeoka and Fushimi (2014) avoid this problem by exploiting the different timing of DPC implementation and shows that there is clear evidence of supplier-induced demand for newborn treatment. Besstremyannaya (2014) examined the differential effects of declining per-diem reimbursement rates on LOS. Whereas most of these studies focus on the policy effects on LOS, DPC can cause more complicated incentive problems as we explained. Recently, Fu and Noguchi (2018) examined the effect of DPC on medical costs including outpatient care of university and advanced treatment hospitals. They show that outpatient care is not affected by DPC.

In the last decade, there has been a great deal of research exploiting a bunching in distribution. Saez (2010) proposed a method to estimate structural parameters such as labor supply elasticity for income tax from a bunching of the income distribution. Chetty et al. (2011) developed the method of Saez (2010) and estimate taxable income elasticity using Danish tax records. Similar approaches are also being used in the field of health economics. Einav et al. (2015) exploits nonlinear contract design in Medicare Part D and estimate the behavioral responses to prices from a bunching of distribution of drug expenditure. Einav et al. (2018) and Elliason et al. (2018) use nonlinear reimbursement of long-term care hospitals in the US and construct an economic model to quantify the welfare implication of the counterfactual reimbursement system. I investigate both the effect of DPC on detailed medical treatments and hospitals' discharge policies that discharge patients on a particular day of LOS. Therefore, this study can contribute to a more profound understanding of the financial incentive of DPC.

The rest of this paper is organized as follows. In Section 2, we explain how the reimbursement system has changed in Japan. In Section 3, we describe the structure of the dataset in this study. In Section 4, we set up econometric models for estimating policy impact. In Section 5, estimation results are presented. Section 6 concludes.

2.2 Institutional Background

2.2.1 Fee-for-service payment system

Traditionally, one of the main features of the Japanese health care system is fee-for-service (FFS) payment to medical service providers. In the case of inpatient care, hospitals are supposed to receive reimbursements based on per-diem basic hospital fees and the sum of the volume-based payment of each medical service they provided (see Figure 1). Hospitals might have an economic incentive to provide excessive health care services under this system, which is the so-called "physician's induced demand". For example, hospitals can make almost fully-healed patients stay longer so that they can generate additional revenue with few medical costs. To reduce needlessly long hospital stay under the FFS payment system, per-diem basic hospital fee is designed to drop at 15th and 31st day of LOS by around 13% and 11% respectively (see Figure 2). However, the effect of this schedule on reducing medical expenditure should be limited because the reduction of reimbursements rate is relatively modest and hospitals still have other medical practices, such as medication and diagnostic testing, that are subject to volume-based payment.

2.2.2 Newly introduced partial prospective payment system

In 2003, the Japanese government started an alternative partial prospective payment system called Diagnostic Procedure Combination / Per-Diem Payment System (DPC) for acute inpatient care. The introduction of DPC was motivated both by improving the quality of inpatient care and by controlling over medical expenditure. Under the previous healthcare system, it was difficult for hospitals to evaluate whether their treatment policy is adequate or not in an objective way because they do not have a statistical benchmark. DPC categorizes each patient into a diagnostic group (similar to DRG in the US) and collects treatment and outcome of these patients from hospitals under DPC. This database allows hospitals and medical researchers to quantitatively evaluate specific treatment, so it was considered to facilitate the improvement of inpatient care.

DPC also has an incentive scheme to contain growing medical expenditure of inpatient care. From this perspective, institutional characteristics of DPC can be summarized as a diagnostic group ("DPC group" hereafter) specific per-diem prospective payment. Under DPC, reimbursements to hospitals consist of both per-diem prospective payment and volume-based payment as in FFS. However, per-diem payment in DPC includes not only basic hospital fees but also medical services such as medication, diagnostic imaging and laboratory test ("PPS-component" hereafter). These PPS-component procedures are all paid through fixed per-diem reimbursement regardless of the cost of service. Surgical procedures and medicine and supplies used for surgery are paid through the volume-based payment as in FFS (see Figure 3).

Per-diem prospective payment in DPC is set according to the DPC-group-specific reimbursement schedule. Figure 4 illustrates the general per-diem payment schedule for DPC. As shown in the figure, the per-diem payment level is set to be high at the beginning of hospital stay and it is lowered in a stepwise fashion based on LOS. At the beginning of the hospital stay the payment level is the rate of 15% above the average cost of medical resources per admission of the DPC group, which is calculated from the previous year. When the LOS reaches 25 percentile of its DPC group, the payment level is dropped to the rate below the mean input level. Its range of reduction is approximately nearly 30%. When the LOS reaches the mean length of its DPC group, the payment level is dropped by 15% further. Finally, if LOS reaches mean + (2 × standard deviation), the patients are regarded as "outlier" and reimbursement is paid through FFS.

In the first year of implementation, DPC is applied to 82 universities and public hospitals on a mandatory basis. The number of acute hospitals that adopt DPC increased rapidly afterward. By 2018, about 1,730 hospitals (29.6% of all acute hospitals in Japan) adopt DPC. This means that the number of beds in DPC hospitals consists of roughly 54.8% of all beds for acute inpatient care. Note that DPC is not mandatory except for the initial 82 hospitals. If hospitals want to adopt DPC, it has to apply to the Ministry of Health, Labour, and Welfare (MHLW) and meet certain conditions of hospital quality. Also, two years-long "preparation period" is set before officially adopting DPC to make sure whether the hospitals applying to this new reimbursement system can submit a vast amount of medical records to MHLW correctly. It is important to keep in mind that hospitals submit treatment information but reimbursements are paid through FFS during the preparation period. Therefore, among hospitals that are willing to adopt DPC, there are both hospitals under DPC and hospitals under FFS payment in a certain year due to the difference in timing of application to DPC. This variation is used as a source of identification for estimating the effect of DPC.

2.2.3 Potential effects of DPC on medical procedure and health outcome

The potential impact of DPC on medical treatments and patients' health

The replacement of FFS with DPC can affect the treatment decision of hospitals in several dimensions. First, LOS can be shortened by DPC due to the rapid decline of reimbursement rate compared with FFS. Under DPC, many components of reimbursements are contained in per-diem prospective payment so that LOS gets economically more important for hospitals. Per-diem payment is reduced by nearly 30% at the start of Period Ⅱ and 15% at the start of Period Ⅲ, which are more rapid decline than hospital basic charge in FFS. This reimbursement schedule could motivate a hospital to discharge patients who stay relatively long in the hospital and admit new ones. As a result, the adoption of DPC likely shortens LOS. However, it is also possible that DPC does not have enough incentive to make hospitals shorten LOS for the following reasons. Prospective payment of DPC is paid by "per-diem" so that hospitals can get reimbursement for each additional overnight stay even if LOS gets longer. In the case of DRG/PPS in the US, hospitals get no reimbursement for an additional stay because reimbursements are fixed by each "admission". One of the reasons that policy-makers adopted the per-diem reimbursement schedule for DPC rather than DRG/PPS type schedule was to avoid drastic changes in reimbursement schemes and medical treatments for inpatient care. Therefore, whether and to what extent implementation of DPC affects LOS is an important empirical question.

Second, DPC could induce hospitals to reduce "treatment-in-ward" (treatment which is usually conducted in ward) compared with the FFS reimbursement system. Treatment-in-ward, such as medication and diagnostic imaging, is reimbursed through per-diem prospective payment *regardless of its actual medical costs* under DPC, while these treatments used to be reimbursed based on the amount of inputs under FFS. Thus the change of the reimbursement system from FFS to DPC possibly makes hospitals reduce treatment-in-ward.

Third, some treatment procedures would be substituted by other treatments that are more profitable for hospitals under DPC. In contrast with treatment-in-ward, surgical procedure is reimbursed based on fee-for-service payment even in DPC. Fee-for-service payment to the surgical procedure could make surgery relatively more profitable than treatment-in-ward. Therefore hospitals might have an incentive to perform more surgical procedures rather than medication to increase their profits. As a result, hospitals would substitute treatment-in-ward with surgical procedure and the share of patients who get surgery could increase. Finally, the implementation of DPC affects patients' health outcomes through changes in LOS and medical treatments described above.

The effect of nonlinear reimbursement schedule on LOS

The introduction of DPC could also affect LOS through the change of incentives that make hospitals discharge patients at particular day of LOS. Recall that both per-diem hospital basic charge under FFS and per-diem payment under DPC are set to drop at certain LOS (I call these points as "kink points"). Basic microeconomic theory predicts that economic agents respond to marginal price. When marginal price declines at a particular supply level, distribution of supply level should cluster at the level where the marginal price drops. Figure 3 illustrates the above rationale. LOS of patient L is l^* both with and without kink of reimbursement. In contrast, LOS of patient H is $l^* + \epsilon$ if reimbursement has no kink but it is l^* if reimbursement has a kink.

This theoretical prediction can be applied to the hospital decision under the payment system with reimbursement kink. If hospitals respond to marginal price (i.e. per-diem payment), distribution of LOS should be bunching at kink points. In the case of FFS, kink points of hospital basic charge schedules are set at the 15th and 31st day of LOS for all patients regardless of the type of disease and diagnosis. Although per-diem payment of DPC also drops at certain LOS, its kink points vary among DPC group that patients are categorized. Therefore, hospitals under FFS has strong incentives to discharge hospitals at 14th day LOS compared with DPC. Therefore, there should be bunching in the distribution of LOS on the 14th day of LOS under FFS, but the bunching should disappear under DPC because the kink point is set for each DPC group (and it is not necessarily 14th day of LOS).

2.3 Data

Medical records in the DPC database collected by the DPC working group in MHLW are the main data used in this research. I use discharge data of patients categorized as circulatory diseases in 2008 and 2009. I focus on circulatory diseases because of the following reasons; First, circulatory diseases are very common among adult people. For example, the number of potential patients of high blood pressure in Japan is estimated to be 43 million (MHLW). Also, cardiac disease is one of the main causes of death of elderly people along with cancer. In 2017, 15.3% of all deaths in Japan are attributed to circulatory diseases. Second, circulatory diseases are treated in many hospitals in the same way so that it is relatively easy to interpret the impact of change in reimbursement on medical treatments.

In each year, the DPC database contains patients' information from July 1st to December

31st. The database includes not only information of patients in DPC hospitals but also information of those in hospitals in "preparation period" (from now on we call these hospitals "preparation hospital"). Therefore the DPC database contains information about patients both under DPC and under the traditional FFS payment system. One of the main advantages of using data in this period is that the trend of adopting DPC hit the peak around 2007-9 and many big acute hospitals adopted DPC during this period. Also, after 2008 we can identify whether each hospital is DPC hospitals or preparation hospitals. This is crucial to analyze the introduction of DPC so that I focus on data after 2008.

Medical records in the DPC database include various information on patients stayed in hospitals. The dataset contains the date of admission and of discharge, which enables us to calculate LOS for each patient. In terms of pre-determined patient characteristics, it contains age, sex and zip code of living place. Diagnostic information of each patient such as DPC group code, name of diseases, and ICD10 code and severity of diseases are included. During the sample period, there exists 150 DPC groups in circulatory disease and each patient is categorized in one of them. The severity of diseases is recorded as the Charlson Comorbidity Index (CCI), which is a patient's severity index based on the degree of complication. Importantly, for hospitals under DPC, the database contains hypothetical FFS reimbursement (FFS-equivalent reimbursement) for each treatment. That is, we can observe the reimbursement to DPC hospitals if these were under FFS rather than DPC. Using this information, we can compare the input of medical procedures between hospitals under DPC and FFS with a unified monetary measure.

Table 1 shows summary statistics for two hospital groups. One is a group that changes from FFS to DPC between 2008 and 2009 (treatment group) and another is a group that keeps FFS during the same period (control group). A simple cross-sectional comparison between the two groups shows that mean LOS is longer in the treatment group than in the control group. The number of beds indicates that the average scale of hospitals is larger in the control group than the treatment group. The fraction of urgent admission and that of patients under severe conditions are larger in the treatment group. Comparison before and after adopting DPC in Treatment shows that LOS declined by 1.5 days on average. In the case of the control group, LOS slightly increased on average. Other key variables do not change so much between 2008 and 2009 in both groups.

2.4 Econometric Model

2.4.1 Difference-in-differences analysis

I estimate the impact of the implementation of DPC on hospitals' treatment decisions based on difference-in-differences (DID) exploiting a different timing of adoption of DPC among hospitals. Recall that the DPC database contains medical records of DPC hospitals and preparation hospitals (before adopting DPC). While reimbursement of preparation hospitals are paid by FFS, those hospitals are willing to adopt DPC (otherwise they did not become a preparation hospital in the first place) and are supposed to adopt it after two years of the "preparation period". Therefore we can reasonably assume that all of the hospitals in the sample want to adopt DPC even if some of the hospitals are still paid by FFS. In this sample, the unique cause of variation in payment system among hospitals is "timing" of applying to DPC. That is, hospitals in the treatment group apply to adopt DPC just one year earlier than those in the control group. It is reasonable to assume that this small difference in the adoption timing is exogenous.

Although the dataset has a panel structure on a hospital level, it is a repeated cross-section on a patient level. For basic specifications, we limit the sample to patients who were not dead during the hospital stay. Specification of the econometric model is such as follows:

$$Outcome_{iht} = t_{2009} + \alpha_h + \gamma(DPC_h \times t_{2009}) + \delta_d + X_{iht}\beta + \varepsilon_{iht}. \tag{2.1}$$

As $Outcome_{iht}$, we use a log of LOS, average input in ward, having surgery or not, and input of surgical procedure. One problem of our data is not to contain actual input of medical resources. Therefore we use FFS-equivalent reimbursement as a proxy of input measure. FFS-equivalent reimbursement is calculated by transforming DPC reimbursement of each episode into FFS as a treatment procedure is identical. Since FFS reimbursement is basically volume-based and strongly affected by medical input, FFS-equivalent reimbursement can provide comparable input measures between DPC hospital and FFS hospital. Then it enables us to examine how medical input changes when DPC is adopted. Year dummy of t_{2009} is equal to 1 if the observation is in 2009 and 0 otherwise. I control unobserved heterogeneity of hospitals by including hospital fixed effect α_h. DPC_h is a dummy variable that is equal to 1 if the hospital is in the treatment group and 0 otherwise. γ is a coefficient on interaction of DPC_h and t_{2009} and it is our target parameter.

It is important to control patient characteristics because hospitals might "select" patients who are more profitable under the new payment system. If it is the case, patient characteristics such as diagnosis and severity change differently between treatment and control group during 2008-2009. To avoid this problem, I control patient diagnosis by including the DPC group fixed effect δ_d. DPC groups are finely classified in terms of both diagnosis and treatment procedure, so the DPC group fixed effect can capture important information about patient conditions. When surgery dummy is used as a dependent variable, DPC group dummies are dropped from independent variables because the choice of surgery and DPC group have a deterministic relationship.

Besides, it is also important to control the "severity" condition of patients because the DPC group does not contain explicit severity information. Therefore I include the Charlson comorbidity index (CCI) in the regressors. Each patient is categorized in DPC group based on their one main diagnosis and complication is not considered. Therefore CCI contains additional information on patient characteristics about complication and severity. X_{iht} contains CCI and other demographic variables including age, sex, switch from outpatient or not, a referral from another hospital, urgent hospitalization, and discharge or changing hospital. I cluster the standard error on the hospital level.

Implementation of DPC can also affect patients' health. As explained in section 2, DPC is designed to make hospitals reduce the input of medical services such as medication. Because of this incentive, hospitals might reduce not only unnecessary treatment but also necessary ones and be at risk for worsening patient's health.

To check this possibility, I estimate the following regression model whose specification is identical to the above estimation:

$$Patient\ Health_{iht} = t_{2009} + \alpha_h + \gamma(DPC_h \times t_{2009}) + \delta_d + X_{iht}\beta + \varepsilon_{iht} \tag{2.2}$$

I use in-hospital death and whether health condition was improved during hospitalization as outcome variables.

2.4.2 Bunching analysis

In this section, I quantify the excess discharge of patients right before the reduction of reimbursement by estimating the “bunching" of distribution of LOS. I apply the estimation method proposed by Saez (2010) and got sophisticated by Chetty et al. (2011).

The basic idea of their method is to construct smooth counterfactual distribution without reduction of reimbursement and estimate how much more observed patients were discharged than counterfactual ones right before the reimbursement change. Hereafter, I define “kink points” as days of LOS one day before the reduction of reimbursement. That is, kink points under FFS are 14th and 30th day of LOS.

Estimation of the bunching is conducted as follows; First, we approximate the observed distribution by polynomial function while excluding kink points:

$$C_j = \sum_{i=0}^{q} \beta_i \cdot (Z_j)^i + \gamma_i \cdot \mathbf{1}[Z_j = k] + \epsilon_j \tag{2.3}$$

Notations are taken from Chetty et al. (2011). In our study, C_j is the number of patients discharged at jth day from admission, Z_j is relative LOS to kink point, k is the position of kink, and q is the order of the polynomial. I set $q = 9$ in the following estimation. $\mathbf{1}[Z_j = k]$ is an indicator function that is equal 1 if Z_j is at a kink point and 0 otherwise. Thus the coefficient γ_i represents the effect of kink point on the number of discharged patients. I define the estimate of the counterfactual distribution as predicted values from (2) setting the effects of kink point, γ_i, is zero:

$$\hat{C}_j = \sum_{i=0}^{q} \hat{\beta}_i \cdot (Z_j)^i \tag{2.4}$$

Then, the excess number of discharged patients who locate at the kink relative to this counterfactual distribution can be calculated as a difference between the observed number of discharged patients and the counterfactual number of the discharged patient at the kink point:

$$\hat{B}_k = C_k - \hat{C}_k \tag{2.5}$$

Finally we calculate the excess mast the kink relative to the density of the counterfactual distribution:

$$\hat{b}_k = \frac{\hat{B}_k}{\hat{C}_k} \tag{2.6}$$

To generate a standard deviation of $\hat{b}_k$, I conduct bootstrap.

Using this method, I estimate the excess discharge at kink points under FFS payment. While kink points are located on 14th and 30th day from admission for all patients under FFS, the location of kink points depends on the DPC group under DPC. That is, change from FFS to DPC eliminate the kink points at 14th and 30th day. I exploit this variation of kink points and check whether excess discharge under FFS is reduced by the elimination of the kink points under DPC.

2.5 Results

2.5.1 Difference-in-differences analysis

The estimation results of the effects of DPC on medical treatments are summarized in Table 2. Column (1) reports the estimates of coefficients of cross-term, γ, using all circulatory disease patients in the main sample. LOS is reduced by 5.8% and statistically significant. This result is consistent with the prediction that DPC would reduce LOS due to the rapid decline of per-diem prospective payment. Average medical input in ward is also reduced significantly by DPC implementation. This can be interpreted as hospitals reduce medical services to secure profit margin under per-diem prospective payment of DPC. In contrast, the probability of conducting surgical procedures is not affected by DPC. Although I pointed out the possibility that hospitals would substitute medication with surgery because surgery is paid through fee-for-service, they do not respond to financial incentives in this way. This result is consistent with previous literature showing that financial incentives to hospitals do not affect non-elective medical services such as surgery for serious disease. Circulatory diseases are often hazardous to patients' life, so DPC does not provide enough incentives for hospitals to change their decision whether they conduct surgery. For the same reason, the input of surgery is also not affected by DPC.

Column (2) to (6) in Table 2 reports the estimation results of patients who are categorized into five main diagnoses in circulatory disease. Angina is the most common cause of circulatory disease and the effect of DPC is similar to the previous results and LOS and average input in ward is reduced significantly, while a decision on surgery is not affected by DPC. The other four diagnoses do not have enough statistical power to detect significant effects of DPC. However, the signs of the effect of LOS are all negative, which suggests LOS might be easily affected by DPC implementation.

The estimation of adverse effects of DPC on patient health is reported in Table 3. "Improvement" is the indicator function which is equal to 1 if patient health condition at the time of discharge is better than the beginning of admission and 0 otherwise. "Mortality" indicates the in-hospital death of patients. The sample size of the estimates for mortality is larger than other estimation because it includes patients who passed away in hospital. I also estimate the effect of DPC on the readmission rate, using whether a patient readmitted to the same hospital for the same disease in 30, 60, and 90 days. Column (1) shows estimates for all circulatory disease patients and both improvement and mortality are not affected by the reimbursement change. Column (2) to (6) report estimates for five main diagnoses in circulatory disease. Note that the effect DPC on "improvement" and "mortality" is the same for all diagnoses in the adverse direction. This might suggest that DPC somewhat harms patient health. However, the effect is not statistically significant when using all circulatory patients. The effects on readmission are also not statistically significant. Therefore, we can conclude that the adverse effect on health is negligible.

This result suggests that DPC is succeeded in reducing medical expenditure while it does not harm patients' health, which is a desirable property of the payment system replacing FFS. In contrast, many studies show that PPS in the US had adverse health effects. For example, Cutler (1995) shows that the introduction of PPS increased both mortality and readmission rates. These contrasting results between DPC and PPS highlight the effectiveness of per-diem prospective payment, which achieves both economic efficiency and protecting patients' health.

2.5.2 Bunching analysis

I exploit the variation of change in payment system between 2008 and 2009 fiscal years and analyze how the change from FFS to DPC affect bunching at kink points. I divide hospitals into three groups; hospitals that change the payment system from FFS to DPC (treatment hospitals) and those that stay as FFS payment (control hospitals). In this analysis, I examine the change of bunching in the case of patients who did not receive surgery during a hospital stay. This is because LOS is crucial in determining reimbursement to hospitals for patients without surgery.

Figure 5 summarizes the observed and counterfactual distribution and estimated bunching in each group-year. There is no bunching on the 30th day so that I focus on the excess discharge on the 14th day in the following analysis. Panel (a) is the distribution for patients in treatment hospitals and Panel (b) is the one for those in control hospitals. In both figures, a vertical line is drawn on the 14th day. Note that there is no reimbursement drop on the 15th day for treatment hospitals in 2009. The percentage value reported on each figure is the estimated value of bunching on the 14th day. Table 5 presents the estimated value. The estimated bunching decline significantly from 26.4% to 8.9% in the treatment hospitals. In the case of control hospitals, the estimated bunching remains high (21.1% and 38.6%). Therefore it is reasonable to conclude that eliminating kink reduces bunching in the case of patients without surgery.

There are some possible explanations why bunching exists even in the hospital group under DPC that has no reimbursement drop on the 15th day. First, hospitals under DPC are paid through FFS before and the habit of treatment procedure under FFS remains even after the adoption of DPC. Second, the 14th day is equal to two weeks so that hospitals might decide LOS by “rule of thumb" based on a weekly basis.

I also analyze the difference of bunching estimate between FFS and DPC by two-sample t-test using “pseudo sample” generated by bootstrap.[1] The results indicate that two bunching estimates above are statistically different. That is, a kink in the payment schedule,

[1] To estimate the standard error of bunching estimate, I perform simulation 300 times, which generate 300 “sample" of bunching estimate. I use these values for simple two-sample t-test and check whether two bunching estimates are sufficiently different.

which generate discontinuous drop of per-diem hospital basic charge, increase the number of patients discharged at kink in cross-sectional analysis.

2.5.3 Kink of DPC Payment

As mentioned in section 2, the DPC schedule is also nonlinear and per-diem payment for each DPC group is supposed to drop at certain LOSs. It is important to note that the kink in the DPC schedule might have a stronger effect on LOS than FFS for two reasons. First, the drop in reimbursement rate is bigger in DPC(27% and 15%) than FFS(13% and 11%). Second, LOS is more important under DPC than under FFS in terms of reimbursement. Per-diem payment in DPC contains many medical practices such as medicine and supplies at wards, while these factors are paid based on the cost under FFS. To examine the effect of DPC kinks on the distribution of LOS, it is necessary to choose DPC groups that are suitable for this study. I chose four DPC groups that have a relatively large sample size and long LOS, which enable us to examine bunching of distribution.

I find the surprising result that there exists almost no bunching at both first and second kink in all DPC groups. Observed and counterfactual distributions are shown in Figure 7-8. Detailed results are in Table 6. All of the bunching estimates have either a small absolute value or a strongly negative value. A possible explanation for these results is that physicians might change treatment choice during a hospital stay and patients could end up being categorized into different DPC groups than the initial one. Therefore, it could be difficult for physicians to specifically control patients' LOS based on the reimbursement schedule of specific DPC group.

2.6 Conclusion

It is crucial to understand how medical providers respond to the financial incentive for health care policy. In this paper, we studied the effect of a per-diem prospective payment system called DPC on inpatient care in Japan. Previous research has focused on the effect of DPC on LOS. I examined not only the effects on LOS but also the effects on medication and substitution relationship among medical treatment. I utilize patient-level discharge data which enables us to control detailed patient characteristics and identify the policy effects.

Using circulatory disease patients data, I find that LOS and average input in ward is reduced significantly. In contrast, the probability of conducting surgical procedures is not affected by DPC implementation. This indicates that even if surgery is paid through fee-for-service under DPC, hospitals do not substitute medication with surgery. I also find that the adverse effects of DPC on patient health outcomes, such as in-hospital mortality and condition at the time of discharge, are negligible. This suggests that DPC is succeeded in reducing medical expenditure while it does not harm the patient's health.

I also show that the significance of bunching in the distribution of LOS decreases due to the elimination of the kink of per-diem payment. This result suggests that hospitals could respond to the change in marginal reimbursement, but the bunching does not exist in the DPC group level. One of the possible explanations is that physicians could change medical treatment in the middle of inpatient care, which leads to a change in the DPC group (and reimbursement schedule). Thus it is difficult for hospitals to specifically control LOS following reimbursement schedule of particular DPC group. However, it is difficult to examine whether this is the reason for the lack of bunching in the distribution of LOS in the DPC group. From the policy perspective, it is important to understand under what condition hospitals tend to respond to marginal reimbursement. I leave this issue for future research.

There exist other remaining issues in this study. First, we analyze with circulatory disease patients but the effects on relatively elective medical services are unknown. Especially, substituting medication with surgery might occur in the elective procedure. Second, the "long-run" effect of DPC cannot be studied in our analysis because we only have the two-year panel. For example, it is possible that hospitals gradually adjust their medical procedure to fit DPC so that the long-run effect of DPC can be larger than the short-run effect.

Bibliography

[1] **Bajari, Patrick Han Hong, Minjung Park, and Robert Town.** 2017. "Estimating Price Sensitivity of Economic Agents Using Discontinuity in Nonlinear Contracts." *Quantitative Economics*, 8(2), pp.397-433.

[2] **Besstremyannaya, Galina.** 2014. "Differential Effects of Declining Rates in a Per Diem Payment System." *Health Economics*, 25(12), pp.1599-1618.

[3] **Clemens, Jeffrey and Joshua D. Gottlieb.** 2014. "Do Physicians' Financial Incentives Affect Medical Treatment and Patient Health?" *American Economic Review*, 104(4), pp.1320-1349.

[4] **Chetty, Raj, John N. Friedman, Tore Olsen, and Luigi Pistaferri.** 2011. "Adjustment Cost, Firm Responses, and Micro vs Macro Labor Supply Elasticities: Evidence from Danish Tax Records" *Quarterly Journal of Economics*, 126, pp.749-804.

[5] **Coey, Dominic.** 2015. "Physicians' Financial Incentives and Treatment Choices in Heart Attack Management." *Quantitative Economics*, 6(3), pp.703-748.

[6] **Cutler, David.** 1995. "The incidence of adverse medical outcome under prospective payment." *Econometrica*, 63, pp.29-50.

[7] **Dafny, Leemore.** 2005. "How Do Hospitals Respond to Price Changes?" *American Economic Review*, 95(5), 1525-1547.

[8] **Ellis, Randall and Thomas McGuire.** 1986. "Provider Behavior under Prospective Reimbursement: Cost Sharing and Supply." *Journal of Health Economics*, 5(2), pp.129-151.

[9] **Einav, Liran, Amy Finkelstein, and Neale Mahoney.** 2018. "Provider Incentives and Healthcare Costs: Evidence from Long-Term Care Hospitals." *Econometrica*, 86(6), pp.2161-2219.

[10] **Einav, Liran, Amy Finkelstein, and Paul Schrimpf.** 2015. "The Response of Drug Expenditure to Nonlinear Contract Design: Evidence from Medicare Part D." *Quarterly Journal of Economics*, 130(2), pp.841-899.

[11] **Eliason, Paul J., Paul L. E. Grieco, Ryan C. McDevitt, and James W. Roberts.** 2018. "Strategic Patient Discharge: The Case of Long-Term Care Hospitals." *American Economic Review*, 108(11), pp.3232-3265.

[12] **Ellis, Randall and Thomas McGuire.** 1996. "Hospital Response to Prospective Payment: Moral hazard, Selection, and Practice-style Effects." *Journal of Health Economics*,15(3), pp.257-277.

[13] **Fu, Rong and Haruko Noguchi.** 2019. "How Did Hospitals Respond to Prospective Payment System under the Japanese Universal Healthcare System?" *Working Paper.*

[14] **Iizuka, Toshiaki.** 2012. "Physician Agency and Adoption of Generic Pharmaceuticals." *American Economic Review*, 102(6), pp.2826-2858.

[15] **Ikegami, Naoki.** 2005. "Medical Care in Japan." *New England Journal of Medicine*, 133(19), pp.1295-1299.

[16] **Ito, Koichiro.** 2014. "Do Consumer Respond to Marginal or Average Price?: Evidence from Nonlinear Electricity Pricing." *American Economic Review*, 104(2), pp.537-563.

[17] **Ikegami, Naoki, and John Campbell.** 2004. "Japan's Health Care System: Containing Costs and Attempting Reform." *Health Affairs*, 23(3), pp.26-36.

[18] **Marsh, Christina.** 2014. "Estimating Demand Elasticities Using Nonlinear Pricing." *International Journal of Industrial Organization*, 37, pp.178-191.

[19] **Nawata, Kazumitsu and Koichi Kawabuchi.** 2010. "Analysis of Length of Stay for Cataract Surgery before and after Introduction of Diagnosis Procedure Combination-based Inclusive Payment System." *Japanese Journal of Health Economic and Policy*, 21(3), pp.291-302.

[20] **OECD.** 2009. "OECD Economic Survey JAPAN." *OECD.*

[21] **Saez, Emmanuel.** 2010. "Do Taxpayers Bunch at Kink Points?" *American Economic Journal: Economic Policy*, 2(3), pp.180-212.

[22] **Shigeoka, Hitoshi and Kiyohide Fushimi.** 2014. "Supplier-induced Demand for Newborn Treatment: Evidence from Japan." *Journal of Health Economics*, 35, pp.162-178.

[23] **Wang, Kai, Ping Li, Ling Chen, Ken Kato, Makoto Kobayashi, and Kazunobu Yamauchi.** 2010. "Impact of the Japanese Diagnosis Procedure Combination-based Payment System in Japan." *Journal of Medical Systems*,34, pp.95-100.

A. Figure

Figure 2.1: Basic Payment Structure of FFS

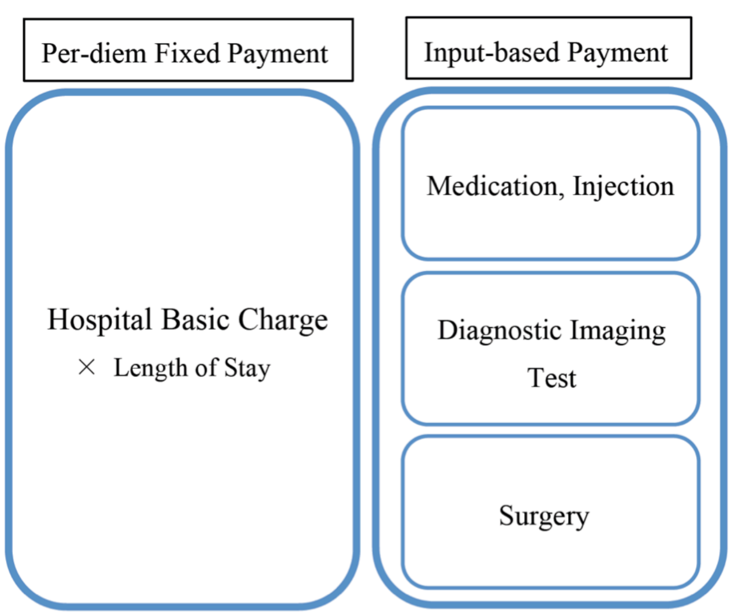


Figure 2.2: Per-diem Hospital Basic Charge of FFS

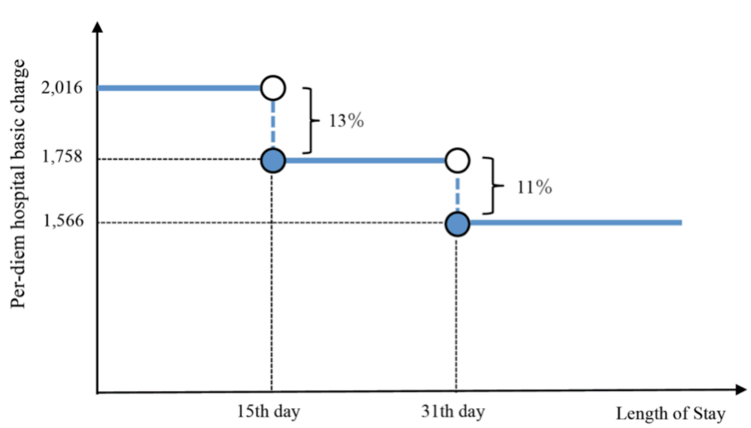


Figure 2.3: Basic Payment Structure of FFS

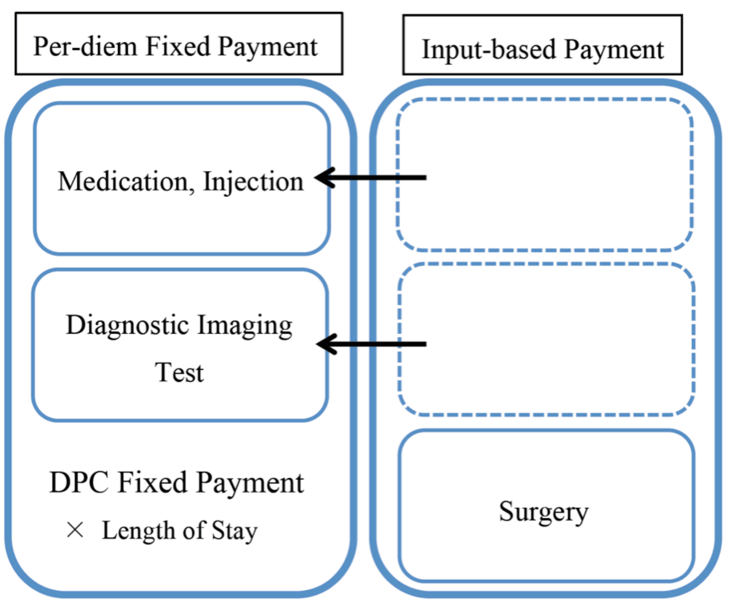


Figure 2.4: Per-diem Hospital Basic Charge of FFS

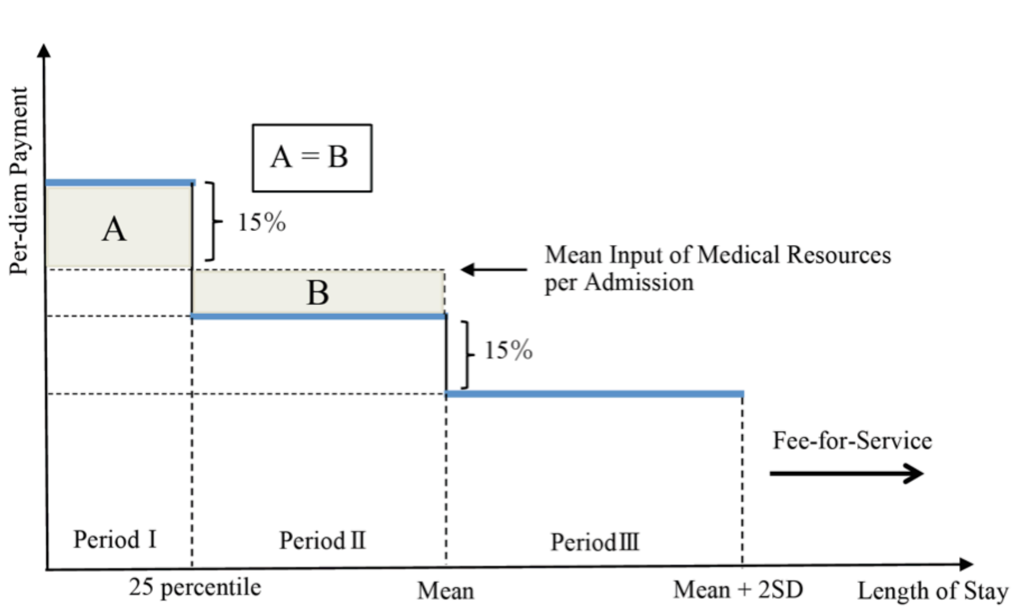
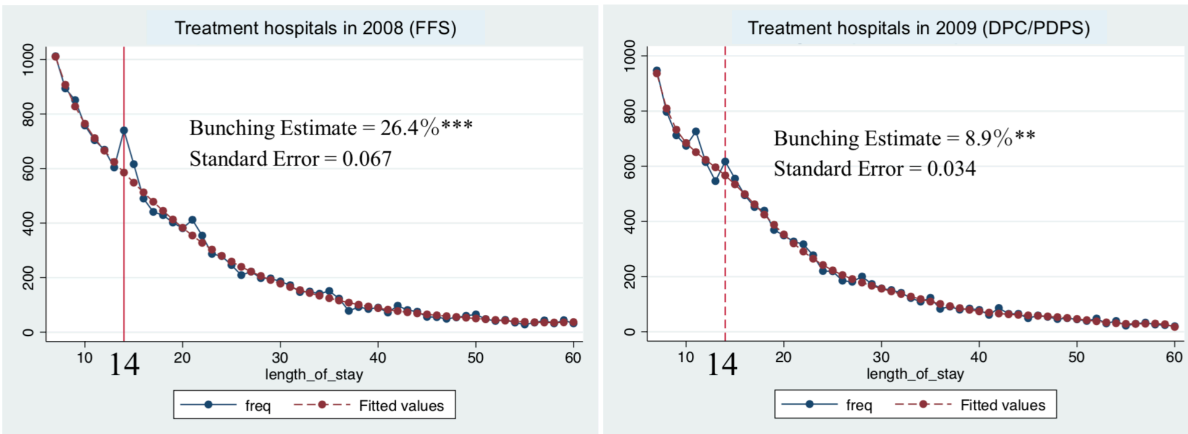


Figure 2.5: Change in Distribution of LOS between 2008 and 2009 (Patients without Surgery)

(a) FFS to DPC



(b) Stay FFS

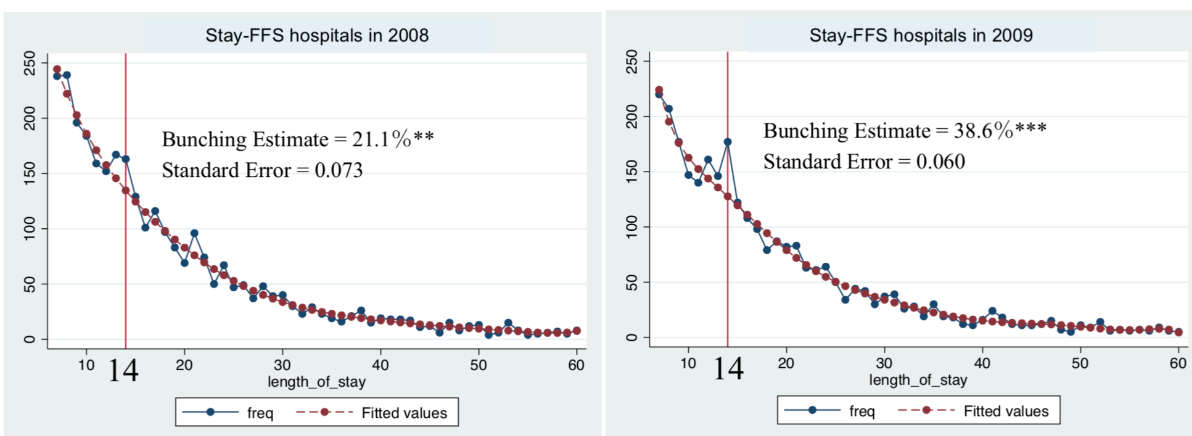
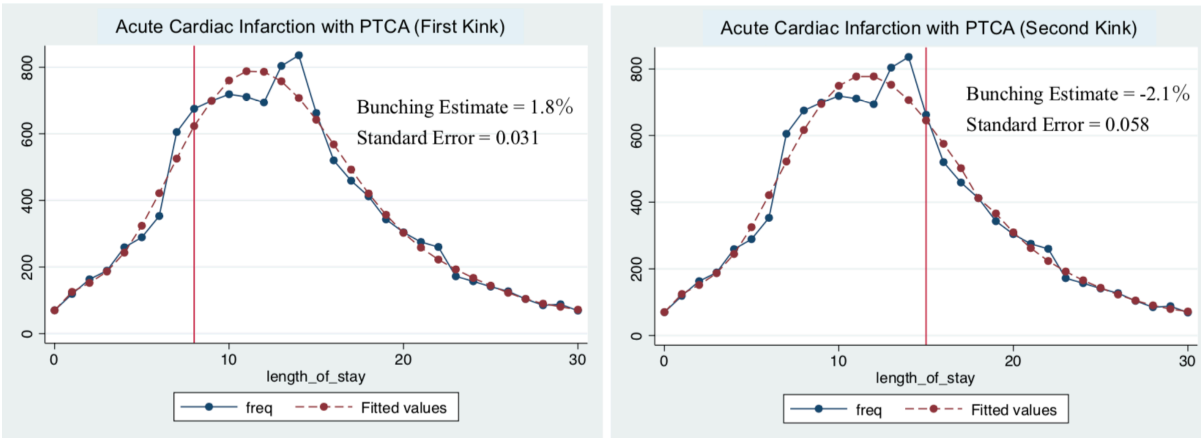


Figure 2.6: Distribution of LOS for DPC Group (1)

(a) Acute Cardiac Infarction with PTCA



(b) Cardiac Arrest without Surgery

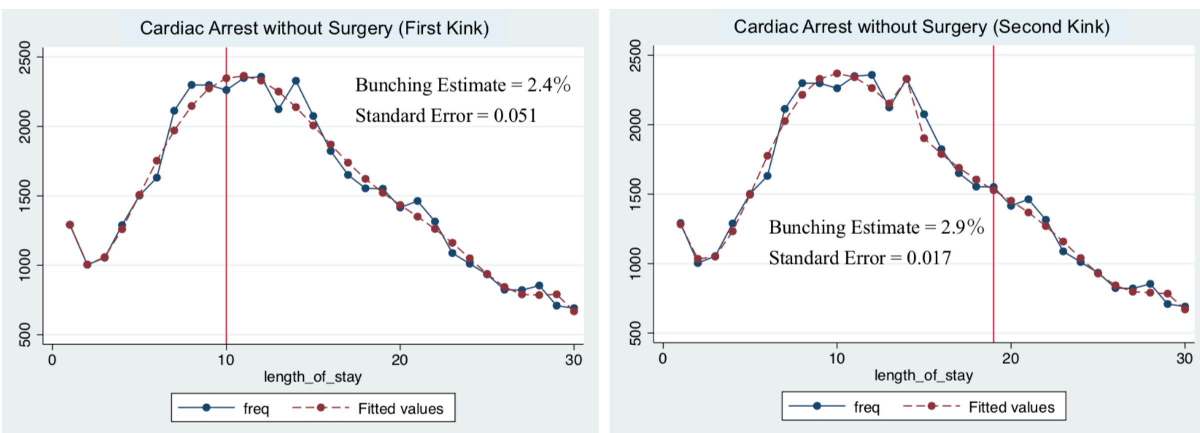
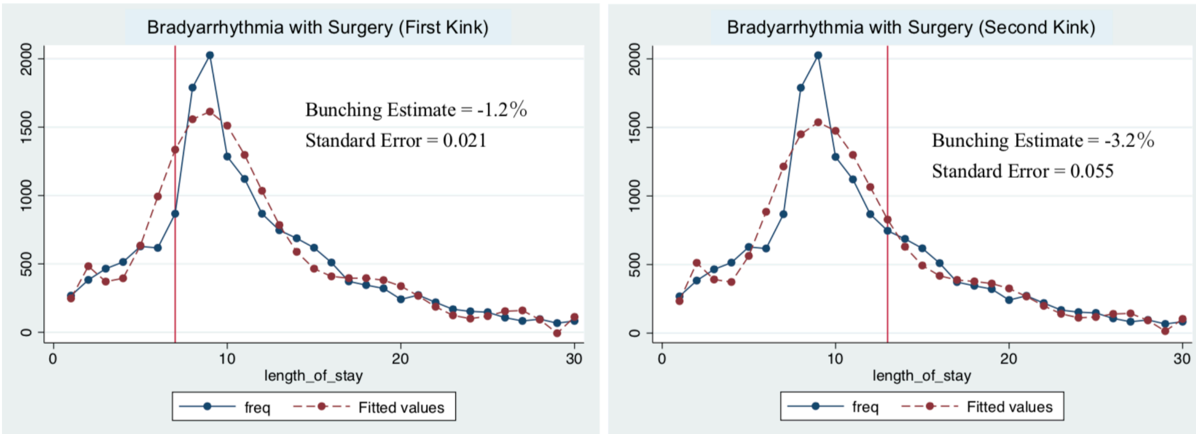
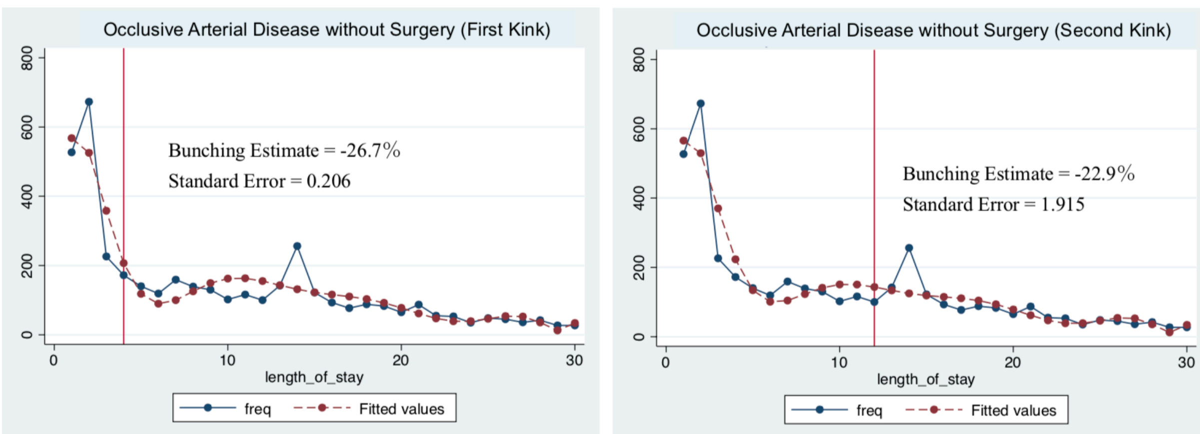


Figure 2.7: Distribution of LOS for DPC Group (2)

(a) Bradyarrhythmia with Surgery



(b) Occclusive Arterial Disease without Surgery



B. Table

Table 2.1: Summary Statistics

	Treatment	Control	Treatment 2008-2009		Control 2008-2009	
Length of Stay	12.3	9.8	12.7	11.9	9.6	9.8
Age	71.3	70.4	71.1	71.6	69.9	70.9
Referral	42%	51%	42%	42%	48%	54%
From outpatient	73%	75%	72%	73%	73%	78%
Urgent	45%	34%	45%	45%	33%	34%
Ambulance	16%	14%	16%	16%	13%	15%
Male	62%	64%	62%	62%	63%	64%
Discharge	65%	51%	64%	65%	55%	47%
Surgery	38%	43%	37%	38%	43%	45%
Serious	9%	5%	7%	10%	5%	6%
Number of beds	425	526	425	425	526	526
Hospital observation	268	68	268	268	68	68
Patient observation	102111	34639	52489	49622	18061	16578

Table 2.2: The Effect of Implementation of DPC (Medical Treatments)

	All	Angina	Cardiac Insufficiency	Cardiac Infarction	Tachyarrhythmia	Bradyarrhythmia
	(1)	(2)	(3)	(4)	(5)	(6)
Length of Stay	-0.058***	-0.074***	-0.029	-0.059	-0.094	-0.036
	(0.014)	(0.016)	(0.034)	(0.051)	(0.049)	(0.055)
Average Input in Ward	-0.062**	-0.069**	0.020	-0.083	-0.093	-0.064
	(0.023)	(0.026)	(0.024)	(0.042)	(0.060)	(0.038)
Surgery dummy	-0.004	-0.005	0.010	-0.004	-0.001	-0.050*
	(0.007)	(0.008)	(0.009)	(0.022)	(0.019)	(0.021)
Surgery spending	-0.049	-0.017	-0.361	-0.045	-0.056	-0.007
	(0.028)	(0.032)	(0.208)	(0.035)	(0.074)	(0.043)
Obs.	136,734	61,819	22,789	8,764	8,329	8,356

Covariates: Age, Male, Comorbidity, Switch from outpatient, Referral, Urgent, and Discharge.
Standard errors in parentheses
*$p < 0.05$ **$p < 0.01$ ***$p < 0.001$

Table 2.3: The Effect of Implementation of DPC (Health Outcomes)

	All	Angina	Cardiac Insufficiency	Cardiac Infarction	Tachyarrhythmia	Bradyarrhythmia
	(1)	(2)	(3)	(4)	(5)	(6)
Improvement	-0.007	-0.009	-0.006	-0.012	-0.005	-0.023
	(0.033)	(0.059)	(0.013)	(0.019)	(0.017)	(0.018)
Mortality	0.002	0.005	0.012	0.014	0.016	0.009
	(0.006)	(0.006)	(0.014)	(0.015)	(0.024)	(0.020)
Readmission (30 days)	-0.016	-0.023	0.000	0.000	0.001	-0.002
	(0.010)	(0.015)	(0.009)	(0.002)	(0.006)	(0.003)
Readmission (60 days)	-0.019	-0.025	-0.003	-0.006	0.000	-0.002
	(0.014)	(0.020)	(0.012)	(0.004)	(0.007)	(0.005)
Readmission (90 days)	-0.021	-0.028	-0.004	-0.007	0.001	-0.001
	(0.016)	(0.022)	(0.013)	(0.004)	(0.009)	(0.005)
Obs.	300,357	62,906	23,164	9,004	8,827	8,817

Notes: Covariates contain Age, Male, Comorbidity, Switch from outpatient, Referral, Urgent, and Discharge. Standard errors in parentheses. *$p < 0.05$ **$p < 0.01$ ***$p < 0.001$

Table 2.4: Change of Bunching Estimates of Patients without Operation

Hospital Group	FFS to DPC		Keep DPC		Keep FFS	
Year	2008	2009	2008	2009	2008	2009
Bunching Estimates	0.264***	0.0890**	0.134**	0.141***	0.211**	0.386***
	(0.0627)	(0.0345)	(0.0738)	(0.0603)	(0.0426)	(0.0274)

Standard errors in parentheses
*$p < 0.05$ **$p < 0.01$ ***$p < 0.001$

Table 2.5: Bunching Estimates of DPC Group

DPC Group	(a) 050030xx03x0xx		(b) 050130xxxx00xx		(c) 050210xx97x0xx		(d) 050170xx9900xx	
Kink	First	Second	First	Second	First	Second	First	Second
Bunching Estimates	0.0184	-0.0212	0.0238	0.0286	-0.0121	-0.0324	-0.267	-0.229
	(0.0311)	(0.0576)	(0.0507)	(0.0168)	(0.0124)	(0.0552)	(0.206)	(1.915)

Standard errors in parentheses

*$p < 0.05$ **$p < 0.01$ ***$p < 0.001$

(a) Acute Cardiac Infarction with PTCA operation

(b) Cardiac Arrest without Operation

(c) Bradyarrhythmia with Operation

(d) Occlusive Arterial Disease without Operation

Chapter 3

The Behavioral Effects of Insurance Coverage and Health Consequences: Evidence from Long-Term Care

3.1 Introduction

As the elderly population increases worldwide, providing accessible and cost-efficient healthcare is becoming increasingly important. The social insurance system plays a central role in providing affordable healthcare for the elderly in many countries. And in terms of designing insurance benefits, controlling a generosity of insurance coverage is one of the key policy tools to affect the demand for healthcare. Recipients can get subsidies for healthcare within a coverage, while they would face a full price otherwise. Due to the salient price changes, recipients inevitably must utilize healthcare taking a given coverage into account. Understanding the impact of insurance coverage on demand for healthcare and its health consequences is crucial in developing an optimal design of a social insurance system.

The basic concept of moral hazard in health insurance assumes that an insurance coverage affects recipient's demand for healthcare only through price.[1] It makes the strong prediction that only recipient facing price changes would respond to coverage changes. This prediction

[1] Here, price means not only a "spot" price but also a "future" price an individual would face in intertemporal dynamic decision-making. See Aron-Dine, Einav, Finkelstein, and Cullen (2015) for the importance of taking into account the effect of future price on individual decision-making in a nonlinear pricing setting. See Einav and Finkelstein (2018) for a comprehensive literature review on moral hazard in health insurance.

also implies that if most recipients' demand for healthcare is well within (or outside) a given coverage, the effect of modest coverage changes could be very limited because the associated price changes influence a small fraction of recipients. Demand for healthcare is generally so diverse that it is unlikely that a large percentage of recipients would simultaneously be affected by a particular coverage change directly through price. This being the case, the policy effects of coverage changes could be significantly small.

When individual's behavioral biases are taken into account, the impact of changes in insurance coverage could be much larger than standard economic theory predicts. Some behavioral models suggest that even recipients who do not face effective price changes could respond to coverage changes. There are a couple of well-established behavioral biases that can be related to this context: anchoring effects and heuristic thinking. If recipients regard insurance coverage as an initial cue in judgment (psychological anchor), they might respond to coverage changes even though it is irrelevant to rational decision-making. Similarly, if recipients make their demand level a certain ratio of coverage through heuristic thinking, coverage changes affect their demand without changing the price they face. The policy implication of insurance coverage depends crucially on whether these behavioral biases exist because they imply that any recipients could be affected by changes in coverage, not merely those who face price changes. Therefore, examining the behavioral biases associated with insurance coverage is essential to assess the potential impact of insurance coverage on demand for healthcare.

This paper investigates how the generosity of insurance coverage affects the demand for health-related services and health outcomes of the elderly, taking into account their behavioral biases. Despite the considerable implications for insurance design, little is known about the economic and health effects of behavioral biases associated with insurance coverage. To address this issue, I study behavioral responses to insurance coverage and its health consequences in the context of a large social insurance program: the long-term care insurance (LTCI) system in Japan. LTCI coverage is characterized by a monthly coverage limit which sets the monthly maximum spending on home-based long-term care services that can be covered by insurance. A unique feature of LTCI is that the coverage limit for recipients who use home-based care is determined by a single health index (standardized care time) which reflects how much long-term care the recipient needs. Every LTCI recipient must take a nationally-standardized health survey that determines their own standardized care time and the coverage limit they will receive. Recipients are required to regularly retake the survey and they receive a newly designated coverage limit each time.

Using newly available administrative data on LTCI, I implement a regression discontinuity (RD) design to estimate the impact of generosity of insurance coverage (monthly coverage limit) on monthly long-term care utilization. I take advantage of five thresholds of standardized care time which create discontinuous variation in coverage limits for home-based care. Although the RD estimates explicitly show that generous insurance coverage leads to an increase in long-term care utilization, this empirical strategy has a limitation on how to interpret the estimates. It can only provide estimates of "overall effects" of insurance coverage and cannot determine to what extent they are due to "behavioral effects" caused by behavioral biases or "price effects" resulting from rational responses to price changes.

To identify behavioral effects, it is necessary to examine whether recipients who do not face effective price changes respond to coverage changes. For instance, coverage expansion does not change the price for recipients who do not exhaust insurance coverage. The advantage of using LTCI in this analysis is that monthly utilization is subject to predetermined plan and can be regarded as static decision-making. Therefore coverage changes are less likely to affect recipients' utilization through their forward-looking behavior such as precautionary motive. I implement the RD design with a set of recipients whose utilization has been sufficiently lower than a given coverage limit to identify behavioral effects. This identification strategy is enabled by coverage changes resulting from a periodical re-examination of standardized care time and subsequent reclassification. Specifically, I first collect recipients whose monthly long-term care utilization is less than half the monthly coverage limit ("low-demand" recipients). This set of recipients is then further narrowed to those whose coverage limit in the next term is either unchanged or one step higher than the previous one. This procedure isolates recipients who face behavioral effects but not price effects, and allows for implementing an RD design to estimate behavioral effects of coverage expansion. Similarly, I also estimate the effects of coverage expansion on recipients who exhaust insurance coverage ("high-demand" recipients) and compare the impact on these two different groups.

The RD estimation focusing on low-demand recipients shows that recipients are significantly affected by behavioral effects of insurance coverage. Importantly, the magnitude of behavioral effects varies among recipients with different health conditions. Behavioral effects are statistically significant for recipients with relatively less severe health conditions: a one-unit expansion of insurance coverage increases long-term care utilization by 0.2 - 0.3 units for those affected by behavioral effects. In contrast, recipients with relatively severe health conditions are not influenced by behavioral effects, with the estimates for these recipients statistically insignificant. Although the RD estimates show that the effect of coverage expansion on low-demand recipients are smaller than high-demand ones, the overall effect of coverage expansion is mainly attributed to behavioral effects because a majority of recipients do not exhaust insurance coverage.

The welfare implications of the effect of insurance coverage depend crucially on its health consequences. Therefore, I examine the influence of home-based long-term care on health outcomes. I use the thresholds generating discontinuous changes in insurance coverage as instruments for service utilization. A recipient's standardized care time and service utilization during the next certification term are used as proxies for health outcomes, as this information should reflect his/her needs for long-term care. The RD estimates show that, for both preventive and regular care, utilizing more long-term care has little effect on health outcomes. This result suggests that generous insurance coverage could induce excessive utilization of long-term care from the perspective of recipients' health.

This paper is primarily related to a growing literature on individual decision-making under nonlinear price schedules of health insurance, especially Medicare Part D program for prescription drugs (see Aron-Dine, Einav, Finkelstein, and Cullen, 2015; Einav, Finkelstein, and Schrimpf, 2015, 2017; Kowalski, 2015; Abaluck, Gruber, and Swanson, 2018; Dalton, Gowrisankaran, and Town, 2019). Insurance coverage with a coverage limit such as LTCI in Japan can be interpreted as having a nonlinear price schedule because individuals face different prices for the same goods depending on whether they are within or outside a specific coverage. Baicker, Mullainathan, and Schwartzstein (2015) discuss the welfare consequences of "behavioral hazard" in health insurance, and recent studies of Medicare Part D also highlight the behavioral biases associated with nonlinear pricing. In particular, Abaluck, Gruber, and Swanson (2018) have found that low-spending beneficiaries, who have essentially zero probability of reaching the coverage gap, nonetheless respond to filling the coverage gap. It is noteworthy that the literature generally attributes behavioral biases to a complex and dynamic decision-making environment. In contrast, this paper finds that even in a relatively simple static decision-making environment, insurance coverage strongly induces behavioral biases. This result suggests that behavioral biases associated with insurance coverage are indeed prevalent, and highlighting the importance of taking non-standard decision-making into account to achieve optimal insurance design. Another important contribution to these studies is that I examine the health consequences of a nonlinear price schedule, while previous studies have exclusively focused on the effect on individual economic behavior.

This paper is also related to the literature on behavioral biases such as anchoring and heuristic decision-making. Anchoring effects have been shown to be prevalent in many decision-making contexts, including in annuity plans (Bernheim, Fradkin, and Popov, 2015), savings (Choi, Haisley, Kurkoski, and Massey, 2017), credit spreads (Dougal, Engelberg, Parsons, and Van Wesep, 2015), art auctions (Beggs and Graddy, 2009) and real estate transactions (Bucchianeri and Minson, 2013). The heuristic thinking has also been shown to exist in important decision-making contexts such as the purchase of cars (Lacetera, Pope, and Sydnor, 2012) and the repayment of debt (Gathergood, Mahoney, Stewart, and Weber, 2019). As for healthcare, Coussens (2018) shows that heuristics are used in physicians' diagnostic decision-making. The present study extends this literature to show that these behavioral biases also significantly affect decision-making in social insurance programs.

Additionally, this study contributes to the literature on the health effects of long-term care. While the impact of nursing home quality on health outcomes has been extensively studied in the health economics literature (see Lin, 2014; Foster and Lee, 2015; Kim and Lim, 2015; Friedrich and Hackmann, 2018), home-based care has not acquired much attention despite its growing importance. A notable exception is Kim and Lim (2015) which study the effect of subsidies for home-based care and institutional care on health outcomes using the long-term care insurance system in South Korea.[2] They find that subsidies for home-based

[2] Kim and Lim (2015) also examine the effect of subsidies for formal home and institutional care on long-term care utilization and informal care use. This paper studies the influence of insurance coverage for home-based care on long-term care utilization and find that the main driver of behavioral responses to

care do not affect the medical expenditure of the beneficiaries. Meanwhile, some medical studies have also explored the effect of home-based care on health outcomes. For example, Gill *et al.* (2008) run an experiment that randomly assigned the elderly to a home-based intervention program to examine the health consequences of the program. They find that those who are assigned to the program exhibit less functional decline over time. While the literature mainly focuses on the effect of the extensive margin of home-based care, this study sheds light on that of an intensive margin of home-based long-term care for recipients with various health conditions. I find that utilizing more long-term care has little effect on health outcomes regardless of the health condition of recipients.

The rest of the paper proceeds as follows. Section 3.2 provides the institutional background and discusses the expected effects of insurance coverage on long-term care utilization. Section 3.3 describes LTCI administrative data and presents summary statistics. Section 3.4 presents empirical strategies applying RD design. Section 3.5 reports the estimation results for the effect of insurance coverage on long-term care utilization and its health consequences. Section 3.6 discusses some policy-relevant topics of the behavioral response to insurance coverage. Section 3.7 concludes.

3.2 Background and Framework

3.2.1 Long-Term Care Insurance System in Japan

The social insurance system for long-term care in Japan was established in 2000, which was arguably the most radical change in the Japanese healthcare system in decades. The long-term care insurance (LTCI) allows eligible recipients to choose long-term care services from among various options and to use them with a moderate out-of-pocket payment within a specific coverage. Due to the rapid aging of the population, spending on LTCI has increased rapidly and has come to account for a substantial fraction of Japanese national finance. The total cost of LTCI, which was 3.6 trillion JPY (36 billion USD,[3] 0.7% of Japanese GDP) in 2000, amounted to 10.8 trillion JPY (108 billion USD, 2.0% of Japanese GDP) in 2017.[4]

insurance coverage is behavioral biases rather than price changes.

[3]For simplicity, I use an exchange rate of 100JPY = 1USD throughout this paper.

[4]Ministry of Finance. URL: https://www5.cao.go.jp/keizai-shimon/kaigi/special/reform/wg1/291108/shiryou1-8.pdf (Japanese)

The cost is expected to continue increasing in the years and decades to come.[5]

LTCI is a mandatory social insurance system in which people of a particular age group must participate. People aged 65 and older are classified as "first-insured" and those between ages 40 and 64 as "second-insured". Both types pay premiums set by their municipal government, but second-insured individuals must present with designated diseases to receive insurance benefits.[6] Because second-insured people are relatively young and have additional eligibility requirements, the majority of LTCI recipients are first-insured.[7]

Care-Needs Certification

There are several steps for receiving long-term care services under LTCI. Figure 3.1 illustrates the utilization process of LTCI. The first step is to apply to a municipal government for the "care-needs certification", which is a health survey for assessing to what extent the applicants need for long-term care.[8] The main purpose of care-needs certification is to classify each applicant into a specific "care-needs level" that determines the available services and insurance coverage. Care-needs certification is based on a nationally-standardized face-to-face survey that is carried out by a trained examiner at the applicants' home (or the hospital if applicants are hospitalized). The examiner first checks 74 items about the applicant's physical and mental condition. Based on the checkup, a special formula generates hypothetical care times for eight categories of assistance. Table 3.1 lists the possible time ranges for each category. The sum of these care times is the "standardized care time", which reflects how much long-term care is needed for the applicant. The longer the standardized care time is, the more an applicant is considered to need long-term care. This indicator plays a key role in determining insurance coverage for each applicant.

Next, based on the standardized care time, applicants are tentatively assigned to a cor-

[5] Based on the prediction made by the Cabinet Office, the total cost of LTCI is expected to reach 25 trillion JPY (250 billion USD) in 2040. URL: https://www5.cao.go.jp/keizai-shimon/kaigi/minutes/2018/0521/shiryo_04-1.pdf (Japanese)

[6] The designated diseases include terminal cancer, rheumatoid arthritis, ALS, ossification of posterior longitudinal ligament, osteoporosis, dementia, Parkinson's disease, spinocerebellar degeneration, spinal canal stenosis, progeria syndrome, multiple system atrophy, diabetes, cerebrovascular disease, arteriosclerosis obliterans, chronic obstructive lung disease and osteoarthritis.

[7] In 2017, first-insured recipients account for 96.4% of all recipients. URL: https://www.mhlw.go.jp/topics/kaigo/osirase/jigyo/m17/1712.html (Japanese)

[8] Family member or guardians can also apply for the care-needs certification on behalf of the recipients.

responding care-needs level. Table 3.2 presents a range of standardized care time and corresponding care-needs level. After the tentative assignment, the Certification Committee of Needed Long-Term Care (hereafter, the Certification Committee), which consists of physicians, nurses, and other health and social service experts, assesses whether the standardized care time appropriately reflects the applicant's needs for long-term care. If the care time is regarded as appropriate, the applicant is assigned to the relevant care-needs level; if not, then the Certification Committee reassigns the applicant to the proper care-needs level.

Available Services and Coverage Limits

Care-needs levels can be divided into two broad categories, which are "Care level" and "Support level". These broad categories determine long-term care services available to recipients. Those who are classified as the Care level are deemed to need long-term care for going about their daily lives, and so these recipients are allowed to use a wide range of long-term care services. In contrast, people classified as the Support level are considered able to perform normal daily activities on their own. Thus the available services under the Support level are aimed at preventing recipients from having an increased need for care in the future.[9] The prices of long-term care services are fixed by the government in both categories and adjusted every three years.

One of the distinct features of LTCI is that insurance coverage for recipients who select home-based care is determined by the recipient's assessed care-needs level. Recipients are therefore unable to choose their coverage on their own based on their preferences. The insurance coverage of LTCI is characterized by monthly coverage limits.[10] For long-term care up to the monthly coverage limit, recipients pay 10 or 20 percent of the total expenditure, depending on income, after which the recipient pays the full price.[11] Table 3.2 presents the coverage limit for each care-needs level expressed as a total unit value for long-term care services. Although the unit value varies slightly across services and municipalities,

[9] Long-term care under the Support level is designed to help recipients accomplish everyday activities independently. For example, if a recipient has a lack of mobility in her hands, the caretaker's role is to design a method for housework that the recipient can perform by herself.

[10] In this paper, the terms "insurance coverage" and "coverage limit" are used interchangeably.

[11] The 20 percent coinsurance, introduced in 2015, applies to those with total annual income of more than 1.6 million JPY (16K USD), and that of first-insured family members is more than 3.46 million JPY (34.6K USD), or 2.8 million JPY (28K USD) for a single-person household).

a reasonable approximation of unit value is 10 JPY or about 0.1 USD. Appendix Figure 3A.1 plots out-of-pocket expenditures as a function of total expenditures on long-term care services when the coinsurance rate is 10 percent. Table 3.2 also shows that recipients who belong to the higher care-needs level (i.e. those who have a relatively severe condition) are entitled to more generous coverage. If applicants' standardized care time is between 35 and 49.9 minutes, they are classified into either Support level 2 or Care level 1. This allocation procedure draws on specific items of standardized care time representing the applicants' cognitive ability and variability in health status.[12]

Long-Term Care Utilization

After available services and insurance coverage are determined, recipients who select home-based care create a monthly plan ("care plan") indicating when and what services are to be provided. In most cases, recipients make their detailed care plan with the help of specialist called care manager.[13] The care manager is also responsible for monitoring the recipient's living conditions to check whether the care plan continues to fulfill the recipients' needs. If the care plan becomes unsuitable due to, for example, changes in health condition, recipients may make changes of the plan on a monthly basis.[14] This type of long-term care utilization scheme based on a care plan thus requires recipients to make a static decision each month on the long-term care services they will receive. Another important thing to note is that any unused portion of the monthly coverage limit cannot be carried over to the next month. That is, recipients cannot expand future coverage by underutilizing current services. This institutional feature also guarantees that long-term care utilization can be interpreted as static decision-making rather than dynamic forward-looking one.

To accommodate changes in long-term care needs, recipients are required to take care-

[12] Recipients are sorted into Care level 1 if both of the following requirements are satisfied: (1) It is difficult for the recipient to understand how to appropriately utilize care prevention services due to mental disability. (2) The physical and mental condition of the recipient is likely going to get worse in a short period.

[13] According to the Long-Term Insurance Act, a care manager is defined as an expert who has specialized knowledge about long-term care who helps recipients draw up the best care plan based on their needs, in coordination with long-term care providers and the municipal government.

[14] In practice, long-term care utilization should be regarded as joint decision-making by the recipient, family members, and care manager. For simplicity, I call "recipient" for decision-maker throughout this paper.

needs certification regularly and are reclassified into a different care-needs level if necessary. In principle, the first care-needs certification is valid for a half year and the following certification is valid for one year. Recipients need to retake the care-needs certification before the term expires to continue using long-term care services under LTCI. Hereafter, I call "certification term" or just "term" for representing each valid term of the care-needs certification.

3.2.2 Mechanisms of Recipients' Responses to Insurance Coverage

Discussions of moral hazard in health insurance assume that price is the only channel through which insurance coverage affects the demand for healthcare. Figure 3.2 depicts a simple mechanism of recipients' responses to insurance coverage under the typical moral hazard framework. In keeping with the structure of decision-making under LTCI, I consider the case in which recipients make a static decision on long-term care utilization with given insurance coverage. Insurance coverage is expressed as a nonlinear budget set because a discount for long-term care services is available only within the coverage region. I suppose that there are two insurance plans: normal coverage (blue line) and generous coverage (red line). While recipients enjoy a discounted out-of-pocket price ("oop price") for long-term care services within their coverage, they face the full price otherwise. Based on the moral hazard, changes in coverage generosity (coverage limit) affect the demand of recipient H, while the demand of recipient L is not affected. Thus the traditional concept of moral hazard predicts that these low-demand recipients who do not exhaust insurance coverage do not respond to coverage changes because they do not change any economic factors, including price, for these recipients.

In contrast to the prediction of moral hazard discussed above, key concepts of behavioral economics suggest that individuals who do not face effective price changes (recipient L in Figure 3.2) might still respond to changes in insurance coverage. The research question of this study is whether an expansion (reduction) of insurance coverage increases (decreases) recipients' demand through behavioral biases in decision-making. Many empirical studies of behavioral biases have demonstrated that an individual's decisions are frequently influenced by various factors that are irrelevant to a rational decision-making framework. Although not excluding other possibilities, this study focuses on the following well-established behavioral biases: anchoring and heuristic thinking. Other potential explanations are discussed in section 3.6.

The anchoring effect is a phenomenon that an individual's decision is influenced by inherently irrelevant information such as unrelated random numbers (Tversky and Kahneman, 1974). The psychological anchor is considered to implicitly provide "suggestion" for its taker. In this way, the coverage limit of LTCI may serve as a salient "starting point", or anchor, that frames people's thinking about their demand for long term care services. It may also give an official "stamp of approval" (Bernheim, Fradkin, and Popov, 2015) to use services to the extent the coverage limit allows. The concept of heuristic thinking suggests that people often use a shortcut to quickly make a decision rather than thinking about the problem through in detail. With LTCI, recipients might use a heuristic (such as a percentage of given coverage limit) for determining their demand for long-term care rather than making it perfectly fit their own needs. If the decision-making of recipients is driven by these psychological biases, their long-term care utilization could be influenced by changes in insurance coverage even when they do not face effective price changes. This study does not attempt to identify a specific behavioral mechanism underlying the relationship between insurance coverage and long-term care utilization but instead interprets any evidence of behavioral biases as being caused by one or more psychological factors.

3.3 Data

LTCI Administrative Data

I use two sets of LTCI administrative data obtained from an anonymous local metropolitan government near Tokyo. The first set is LTCI claims data, which contains monthly information on eligibility, long-term care utilization, and demographic characteristics for all LTCI recipients in the city. The available sample period of this data is from June 2012 to March 2018. Eligibility information contains the care-needs level, start and end dates of each certification term, coinsurance rate, and public subsidy eligibility. For service utilization, the claims data provides information on how much each recipient uses and spends on a monthly basis for each type of long-term care service. For this study, home-based care is grouped into five categories: home care, day care, home-visit nursing care, rehabilitation, and others.

The claims data contains relatively limited information on the demographic characteristics of recipients, providing age and gender but no information on income and family structure.

The second dataset is newly available LTCI administrative data on care-needs certification. The most important information in this dataset is the assessed standardized care time, which is used for assigning recipients to a care-needs level. This data is available for each certification term and contains a breakdown of how the final standardized care time was calculated (that is, it provides a hypothetical care time for each category of assistance) as well as other information related to care-needs certification such as the start and end dates of each certification term. The sample period of LTCI certification data is the same as LTCI claims data.

The analysis sample is created by linking the LTCI certification data and claims data via a unique ID number. This dataset allows me to associate recipients' long-term care utilization and their standardized care time. From this preliminary dataset, several types of recipients were excluded from the baseline analysis sample. First, I omit recipients who live in a nursing home, because the coverage limit is applied only to those who use home-based care. Second, I omit recipients who receive public subsidies for LTCI, because these individuals face a different incentive scheme than usual recipients. The remaining recipients constitute a baseline sample.

Summary Statistics

Table 3.3 presents the summary statistics for the baseline sample and each care-needs level. Panel A shows the demographic information of LTCI recipients.[15] The age of the baseline sample is around 81.2 years old, with recipients in high care-needs levels tending to be slightly older than those in low care-needs levels. More than half of recipients are women, which is natural because women tend to live longer than men and thus are more likely to suffer from a disability due to their older age.[16] Approximately 13% of recipients face a higher coinsurance rate (20%) due to high income, and this rate is uniform across different care-needs levels. Change of care-needs level presents a fraction of recipients whose care-needs level is changed

[15] As some recipients were allocated into different care-needs levels for several care-needs certifications, the sum of recipients for each care-needs level is not equal to the number of total recipients in the first column.

[16] In Japan, the average length of life for men and women in 2017 was 81.09 and 87.26 years old respectively.

by an assessment of the Certification Committee. In the baseline sample, only 5% of all recipients receive different coverage limit than standardized care time indicates. This means that standardized care time determines coverage limit in most cases.

Panel B shows information on the standardized care time calculated during care-needs certification as well as a breakdown of the hypothetical care time for each category of assistance. It can be seen that with higher care-needs levels, the burden of assistance related to eating, transferring, toileting and hygiene increases sharply while other categories of assistance are relatively constant across care-needs levels.

Panel C presents information on long-term care utilization. It shows that day care is the most used long-term care services, followed by home care. Home care and home-visit nursing care utilization monotonically increases as the care-needs level becomes higher, while demand for day care and rehabilitation decreases at higher care-needs levels. In the baseline sample, only 9% of recipients exceed the monthly coverage limit, with the rate of exceeding the coverage limit higher for recipients with high care-needs. As most recipients use long-term care services within their allocated coverage limit, this highlights the importance of investigating the behavior of recipients who do not fully exhaust their insurance coverage.

Who Faces Price Changes?

To study behavioral biases associated with insurance coverage, it is necessary to determine who would face effective price changes due to coverage changes. Appendix Figure 3A.2 shows the distribution of monthly long-term care utilization for each care-needs level, and the vertical red line indicates the coverage limit for each level. There exist a clear "bunching" around coverage limit in the distributions of Care level 1 - 5, which suggests that recipients recognize given coverage limits and try to avoid paying higher prices outside the coverage. As the bunching has some width, the coverage limit appears to restrict the demand of recipients whose utilization level is sufficiently close to it. One possible reason for recipients not completely exhausting their insurance coverage is that they may be saving some of their coverage for any unexpected long-term care expenditure. Although an unexpected expenditure for long-term care is rare and small compared with that for medical care, this precautionary motive could prevent recipients from exhausting insurance coverage.[17] Another possibility is that the demand for long-term care services is discrete rather than continuous in practice and this is preventing recipients from fully exhausting their coverage limit. If this is the case, then the long-term care utilization of recipients who do not completely exhaust their insurance may also be constrained by the coverage level, and these recipients would face effective price changes when the coverage expands.

To address this concern, I set conservative criteria that recipients whose monthly utilization is lower than 50% of the coverage limit are not affected by price changes due to coverage expansion. I also adopt criteria in assuming that recipients whose monthly utilization is higher than 80% of the coverage limit are affected by price changes. Similarly, I assume that recipients whose monthly utilization is lower than 80% of the one-stage lower coverage limit are assumed to be not affected by price changes when they face a one-stage reduction in coverage.[18] I set 80% rather than 50% to ensure statistical power for analyzing behavioral biases due to coverage reduction.[19] Recipients whose monthly utilization is higher than 80% of the one-stage lower coverage limit would be affected by price changes. The following section describes the empirical strategy for estimating the effect of insurance coverage through behavioral biases, using those recipients not affected by price changes.

3.4 Empirical Strategy

3.4.1 The Overall Effects of Insurance Coverage

The empirical strategy throughout this study is based on an RD design. In this section, a simple RD design is used to initially estimate overall effects of insurance coverage on long-term care utilization, which includes both responses to price changes ("price effects") and behavioral biases associated with insurance coverage ("behavioral effects"). Section 4.2 then describes how a "switch" in insurance coverage is used to isolate behavioral effects.

First, to estimate recipients' responses to various generosity of insurance coverage, I

[17] For example, family members of the recipient have to leave home urgently for a couple of days, and the recipient utilizes additional short-stay services.

[18] For example, one-stage lower coverage of Care level 3 is Care level 2.

[19] Recipients' needs for long-term care tend to increase as they get older, so there are a relatively small number of recipients whose care-needs level is changed to lower levels.

implement an RD design exploiting thresholds of the standardized care time that generate discontinuous changes in the monthly coverage limits. Figure 3.3 illustrates the relationship between standardized care time and coverage limits. There are five thresholds at 32, 50, 70, 90, and 110 minutes, and these thresholds change the coverage limit by $5,470$ (+109.3%), $2,924$ (+17.5%), $7,315$ (+37.3%), $3,875$ (+14.4%), and $5,259$ (+17.1%) units respectively. The relationship between standardized care time and coverage limit is not deterministic because recipients' care-needs level could be altered by the Certification Committee. Thus I exploit these thresholds as instruments for the generosity of insurance coverage.

Given a particular threshold, I only use recipients whose standardized care time is in one of the two neighboring care-needs levels separated by the threshold. The effect of insurance coverage on long-term care utilization at each of the five different thresholds is:

$$Utilization_{it} = \alpha^c + \beta^c Coverage_{it} + f^c(Caretime_{it}) + X_{it}\gamma^c + \varepsilon_{it}, \tag{3.1}$$

where $Utilization_{it}$ is the long-term care utilization of recipient i in year-month t measured by monthly total units. [20] $Coverage_{it}$ is the generosity of insurance coverage (coverage limit) and $Caretime_{it}$ is a standardized care time, which is a running variable. The covariates in X_{it} include age, gender, coinsurance rate, and the hypothetical care times of each category of assistance which together are used to calculate standardized care time. $f^c(\cdot)$ is a set of functions of standardized care time specified below. The parameter of interest is β^c, which indicates the effect of a one-unit increase in insurance coverage on long-term care utilization.[21]

[20] An adjustment is required to accurately reflect service utilization, as the number of units of same day care service could be higher at upper care-needs levels because of the higher price of these services. To address this, units of day care services are normalized using a price of Care level 1 as a baseline. Hereafter, the measure of long-term care utilization refers to this adjusted data. The effect of price changes in day care on long-term care utilization is assumed to be negligible because price changes across care-needs levels is fairly small (approximately 10%).

[21] For the 32 and 50 minute thresholds, an ambiguity in the sorting procedure of recipients into either Support level 2 or Care level 1 could present a challenge in the interpretation of the estimated responses to insurance coverage. If standardized care time is between 32 and 50 minutes, relatively healthy recipients are allocated to Support level 2 while relatively unhealthy recipients are assigned to Care level 1. In the estimation for the threshold at 32 minutes, the demand of Support level 1 is compared with that of Support level 2 around the threshold. Therefore, recipients just above the threshold might not be comparable to those just below the threshold because the recipients just above the threshold might be "overly healthy" due to the sorting procedure. The summary statistics show that healthy recipients tend to require less long-term care than unhealthy recipients, and so the estimated responses could be biased downward at the 32 minutes threshold. Similarly, recipients just below the threshold at 50 minutes could be "overly unhealthy" compared

All parameters are indexed by a threshold c because they are estimated separately for each threshold.

The first-stage regression is:

$$Coverage_{it} = \alpha_0^c + \beta_0^c \mathbb{1}\{Caretime_{it} \geq Cutoff\} + f_0^c(Caretime_{it}) + X_{it}\gamma_0^c + \varepsilon_{it}, \tag{3.2}$$

where *Cutoff* denotes one of the cutoff values of standardized care time which separates neighboring care-needs level; that is, $Cutoff \in \{32, 50, 70, 90, 110\}$ depending on the threshold exploited for the estimation. $\mathbb{1}\{Caretime_{it} \geq Cutoff\}$ is a dummy variable that takes a value 1 if standardized care time is greater than or equal to a given cutoff value. $f_0^c(\cdot)$ (and $f^c(\cdot)$ in equation 3.1) is a set of functions of running variable, whose parameters are free to vary on either side of a given threshold. The function is specified as a linear, quadratic, or cubic function for parametric estimation.[22] In addition, a nonparametric local polynomial regression with robust confidence intervals proposed by Calonico, Cattaneo, and Titiunik (2014) is also estimated. Standard errors are clustered at recipient level in equation 3.1 and 3.2.

3.4.2 Coverage Changes and Identification of Behavioral Effects

The limitation of the above RD design is that it cannot disentangle whether the overall effects of insurance coverage are due to behavioral effects or price effects. To examine the existence of behavioral effects, it is necessary to see whether recipients who do not exhaust insurance coverage respond to coverage changes as described in section 3.2 (see Figure 3.2).

To identify behavioral effects, I make use of a "switch" in the coverage generosity (including both an expansion and a reduction in coverage) caused by a reclassification of recipients' care-needs levels. Specifically, I use a two-step procedure to narrow down the recipients and then implement the RD design. First, I focus on recipients whose average monthly utilization during any certification term s is lower than 50% of the coverage limit. As discussed in

with those just above the threshold, and so those estimates could be biased downward. However, in both cases, this would cause the estimates to be conservative and the main implication of the analysis is not affected by this sorting procedure.

[22]The same parametric specification is used for both sides of the threshold.

the section 3.3, this step excludes recipients who exhaust the coverage limit and thus could be influenced by price effects caused by changes in insurance coverage. Recipients utilizing services less than 50% of the coverage limit are defined as “low-demand”, while those whose service utilization is higher than 80% are “high-demand”. I assume that low-demand recipients are affected by behavioral effects of coverage changes, but not price effects.

In the second step, I further narrow down the low-demand recipients to those whose standardized care time for the next term $s + 1$ is in either the same or one stage higher care-needs level compared with that of term s. Recipients who are switched to an upper care-needs level and more likely to face a higher coverage limit can be considered a “treatment group”, while those who remain in the same care-needs level can be considered a “control group”. With the selected low-demand recipients, the RD design is implemented to estimate the behavioral effect of an increase in coverage limit (coverage expansion) on long-term care utilization.

Using low-demand recipients selected through the above procedure, behavioral effects of coverage expansion is estimated through leveraging each of the five thresholds separately as in the previous RD design. The causal relationship of interest can be written as:

$$\begin{aligned}\Delta Utilization_{it}^{s,s+1} = \alpha^c + \beta^c \Delta Coverage_i^{s,s+1} &+ f^c(Caretime_{is+1}) \\ &+ \gamma^c Caretime_{is} + X_{it}^{s,s+1}\eta^c + \varepsilon_{it},\end{aligned} \tag{3.3}$$

where $\Delta Utilization_{it}^{s,s+1}$ is a change in long-term care utilization of recipient i, which is calculated as a difference between the monthly utilization at year-month t during the term $s + 1$ and mean monthly utilization during the previous term s (before the reclassification). $\Delta Coverage_i^{s,s+1}$ is a change in insurance coverage between s and $s + 1$ recipient i faces. β^c is the target parameter and indicates behavioral effects of one-unit coverage expansion on long-term care utilization.

The first-stage regression is:

$$\Delta Coverage_i^{s,s+1} = \alpha_0^c + \beta_0^c \mathbb{1}\{Caretime_{is+1} \geq Cutoff\} + f_0^c(Caretime_{is+1}) \\ + \gamma_0^c Caretime_{is} + X_{it}^{s,s+1}\eta_0^c + \varepsilon_{it}. \tag{3.4}$$

In this model, standardized care time during $s+1$, $Caretime_{is+1}$, is a running variable. The functional specification of $f_0^c(\cdot)$ (and $f^c(\cdot)$) is the same as the previous model. I control the standardized care time during previous term s, $Caretime_{is}$, in order to compare recipients who are also similar in past health conditions. The covariates in $X_{it}^{s,s+1}$ include age, gender, coinsurance rate, and the hypothetical care times of each category of assistance.

Similarly, by conditioning on high-demand rather than low-demand recipients, I estimate the effect of coverage expansion on recipients who have exhausted insurance coverage using the same strategy. High-demand recipients could be affected both behavioral and price effects, and it is not possible to disentangle these effects among these recipients. Therefore, the effect on high-demand recipients should be interpreted as a combination of these two effects.

Another advantage of exploiting coverage changes is that it also allows for estimating the effect of a "decrease" in coverage limit (coverage reduction). In this case, I focus on recipients whose standardized care time in the following term $s+1$ is in the same or one stage lower care-needs level compared with that of term s. The estimation for coverage reduction is very similar to the case of coverage expansion so that I describe the procedure in Appendix B.

Why Not Difference-in-Differences?

Another possible approach to estimating behavioral effects is to exploit an exogenous timing of coverage changes. The timing of next care-needs certification is predetermined and most recipients follow the schedule. I could regard low-demand recipients who face coverage changes at a given timing as a treatment group and those who do not take the care-needs certification at the same time as a control group and estimate behavioral effects by difference-in-differences (DID) approach.

The DID approach, however, has a serious problem for identifying behavioral effects in this context. Recipients tend to change monthly plans for long-term care utilization at the timing of care-needs certification even if their insurance coverage does not change. This is probably because it is a good opportunity for recipients to rethink their monthly utilization plan. Therefore, changes in utilization at the timing of coverage changes could be due to not only coverage changes but also a renewed care-needs certification per se. This means that an exogenous variation of the timing of care-needs certification cannot be used to identify the effect of coverage changes, needless to say, behavioral effects. In contrast, the above RD design focuses on recipients after a renewed care-needs certification, which offset its effect on long-term care utilization and identify behavioral effects due to coverage changes.

3.4.3 Estimation of the Health Consequences of Long-Term Care

Using recipients' responses to insurance coverage, I estimate the effect of long-term care utilization on their health outcomes through a 2SLS estimation. Recipients in Support level can only utilize preventive long-term care while those in Care level utilize regular long-term care services. The causal relationship of interest can be written as follows:

$$Health_{is+1} = \alpha^c + \beta^c Utilization_{is} + f^c(Caretime_{is}) + X_{is,s+1}\gamma^c + \varepsilon_{is}, \tag{3.5}$$

where $Health_{is+1}$ is a measure of the health status of recipient i at term $s+1$. I use three different variables as recipient's health measures: (i) standardized care time calculated at the beginning of term $s+1$, (ii) long-term care utilization during the term $s+1$, and (iii) whether recipients end up entering a nursing home during $s+1$. $Utilization_{is}$ in equation 3.5 is the mean monthly long-term care utilization of recipient i during term s and $f^c(\cdot)$ is a function of standardized care time. $X_{is,s+1}$ contains the length of time of term s as well as the recipient's characteristics including hypothetical care time of each category of assistance as in the regressions discussed above. If utilization information is used as outcomes, standardized care time at $s+1$ is included in $X_{is,s+1}$ because long-term care utilization is affected by insurance coverage (care-needs level). The parameter of interest is β^c, which indicates the health effects of long-term care utilization.

Because long-term care utilization is an endogenous variable, I use each threshold generating discontinuous variation in coverage limit as an instrument for it. The first-stage of 2SLS is estimated by the following regression at each of the five different thresholds:

$$Utilization_{is} = \alpha_0^c + \beta_0^c \mathbb{1}\{Caretime_{is} \geq Cutoff\} + f_0^c(Caretime_{is}) + X_{is,s+1}\gamma_0^c + \varepsilon_{is}, \quad (3.6)$$

where notations are the same as the previous regressions and $\mathbb{1}\{Caretime_{is} \geq Cutoff\}$ is an instrument. The functional specification of $f_0^c(\cdot)$ in first-stage is same as $f^c(\cdot)$ in the second-stage.

It is noteworthy that the estimates represent the short-run health effect of long-term care utilization. The above 2SLS estimation uses long-term care utilization only during the first term, $s = 1$, to focus on the direct effect of long-term care utilization rather than the accumulated effect from utilization in the past.[23] The average length of the first term is 6.8 months so that the estimated health effect is of different long-term care utilization for half a year.

3.4.4 Validity Tests for the RD Design

Distribution of the Running Variable and McCrary Test

In this section, I check the validity of the local randomness around thresholds. First, I examine the smoothness of the distribution of the running variable around thresholds. Figure 3.4 presents the distribution of standardized care time for the baseline sample. Panel (a) of Figure 3.4 is the distribution for the Support level, which appears fairly smooth around the thresholds. Panel (b) shows the distribution for the Care level. While there exists a noticeable jump at just above the thresholds of 50 and 70 minutes, it is relatively smooth around other thresholds. Appendix Figure 3A.3 shows the distribution for recipients who take their first care-needs certification and are used to estimate health effects of long-term care utilization. The distribution is similar to Figure 3.4 and smooth except the threshold of 50 minutes. Figure 3.5 shows the distribution after reclassification conditioning on low-demand

[23] Note that it is impossible to distinguish whether responses to insurance coverage are due to behavioral effects or price effects in this estimation, because I use long-term care utilization of the first term and cannot condition on the utilization during the previous term.

recipients who are used to identify behavioral effects of coverage expansion. Compared with Figure 3.4, all distributions do not have a visible discontinuity around the thresholds. This validation test strongly suggests that the local randomness is satisfied in the case of coverage expansion.

In contrast, the local randomness assumption appears to be violated in the case of coverage reduction. Appendix Figure 3A.4 shows the distribution of running variable conditional on the (modified) low-demand recipients who are used to identify behavioral effects of coverage reduction. The figures clearly show discontinuities at all thresholds. This suggests that care-needs certification examiners avoid reducing recipients' insurance coverage, and the assignment of recipients into either side of the thresholds is not random.

To test the smoothness of the distribution statistically, I conduct the McCrary test (McCrary, 2008) at each threshold. Panel A and B in Appendix Table 3A.1 present the estimates of the test with all recipients in the baseline sample and those who take the first care-needs certification respectively. Panel C shows the estimates of the same test with coverage expansion, which is estimated separately for low-demand, high-demand, and all recipients. Panel D presents the results for coverage reduction. The test indicates that although some estimates are statistically significant, the discontinuity of the distribution is reasonably small especially the case where coverage expansion is considered. As Appendix Figure 3A.4 suggests, however, the discontinuity is more significant in the case of coverage reduction.

Covariates Balance Test

Furthermore, I examine whether predetermined recipient characteristics are balanced around the thresholds. For this analysis, the following parametric equation is estimated:

$$Y_{is} = \alpha^c + \beta^c \mathbb{1}\{Caretime_{is} \geq Cutoff\} + f^c(Caretime_{is}) + \varepsilon_{is}, \tag{3.7}$$

where Y_{is} is the covariate of recipient i during certification term s. Linear and quadratic specifications are used for $f^c(\cdot)$.

Table 3.4 reports the estimates for β^c with recipients used for the estimation which exploits a coverage expansion. The results are presented separately for low-demand, high-demand, and all recipients. With a few exception, the estimates are mostly statistically insignificant, which indicates that the local randomness assumption is reasonably satisfied. Appendix Table 3A.2 presents β^c for all recipients in the baseline sample. Except for the quadratic specification for Support level 1 - 2, the variations of covariates around thresholds are mostly statistically insignificant. Appendix Table 3A.3 presents the estimates for recipients who take first care-needs certification. These estimates are also statistically insignificant in most cases. These results indicate that the covariates are well balanced around each threshold. Appendix Table 3A.4 presents the estimates for the recipients used for analyzing the response to a coverage reduction. Despite the sharp discontinuity in the distribution of the running variable around thresholds, the covariates are relatively balanced, which suggests that the discontinuity could be attributed to unobservable factors.

Overall, the validity test suggests that the RD identification assumption is reasonably satisfied for estimates of coverage expansion. As far as coverage reduction is concerned, while observable characteristics are balanced around the thresholds, unobservable factors could be causing the sharp discontinuities in the distribution of running variable. I thus regard estimates of coverage expansion as reliable main results for behavioral effects.

3.5 Results

3.5.1 Overall Effects of Insurance Coverage

Before turning to the estimation results, Figure 3.6 plots the relationship between standardized care time, monthly coverage limits, and monthly long-term care utilization. Standardized care time is divided into 1 minute-wide interval (bin) and each hollow circle and triangle represent the local average of coverage limits within bins for the Care level and the Support level respectively. Each filled circle and triangle represent the local average of the monthly long-term care utilization within bins, and a quadratic prediction is fitted on the plots for each care-needs level. The first thing to notice is that coverage limits discontinuously increase at all thresholds. This indicates that the thresholds are valid instruments for the generosity of insurance coverage. In addition, long-term care utilization also increases at all thresholds. This strongly suggests that insurance coverage significantly affects long-term care utilization through price and/or behavioral effects.

Table 3.5 presents the first-stage estimates (β_0 in equation 3.2) and the second-stage estimates (β in equation 3.1) for the overall effect of insurance coverage at each threshold. Each column presents the results for a specific polynomial order as well as nonparametric estimates. The first-stage estimates are positive and statistically significant at all thresholds regardless of specifications, which confirms the validity of instruments.

The second-stage estimate represents the effect of having more generous insurance coverage by one unit on long-term care utilization. The estimates are positive and statistically significant at almost all thresholds and specifications. These estimates unambiguously show that generous insurance coverage increases long-term care utilization. Importantly, despite only a small fraction of Support level recipients exhausting their insurance coverage, more generous insurance coverage leads to higher long-term care utilization. This result suggests that behavioral biases could play a major role in recipients' responses to changes in insurance coverage.

The above estimates show the existence of important heterogeneity among recipients who have different needs for long-term care. First, Support level recipients are less sensitive to insurance coverage than those in the Care level. One unit increase in the coverage limit for the Support level increases utilization by around 0.2 units, while that for the Care level increases utilization by 0.5 to 0.8 units. This difference could be related to the fact that there are few recipients in the Support level who exhaust their insurance coverage, and thus these recipients are not affected by price effects. Second, even among Care level recipients, the effect of insurance coverage tends to grow larger as the care-needs level becomes higher. This trend supports the prediction that the more recipients who exhaust their insurance coverage, the larger the price effect will be.

3.5.2 Behavioral Effects of Coverage Expansion

The estimation results in the previous section suggest that behavioral biases could play a key role in recipients' responses to changes in insurance coverage. The results also make clear the necessity of excluding price effects in order to identify behavioral biases, because the strength of price effects significantly affects long-term care utilization. In this section, the effect of a coverage expansion on long-term care utilization is estimated for three types of recipients: low-demand, high-demand, and all recipients. As in the previous section, the main outcome variable is monthly long-term care utilization.

Figure 3.7 plots the change of the utilization for low-demand recipients to see whether coverage expansion affects long-term care utilization through behavioral effects. As explained in section 3.4.2, the estimation for each threshold uses a different set of recipients. Thus the plots around each threshold are presented by separate panels. Note that recipients to the right of the threshold are more likely to face coverage expansion. This figure presents significant behavioral effects and also their notable heterogeneity among recipients with different health status. Panel (a) - (c) in the figure indicate that the utilization of low-demand recipients clearly responds to the expansion of insurance coverage. The salience of this response strongly suggests that recipients with relatively low needs for long-term care are affected by behavioral effects. In contrast, Panel (d) and (e) indicates that recipients with severe health conditions show little or no visible behavioral biases associated with insurance coverage. Appendix Figure 3A.5 presents the plots of the changes in utilization for high-demand recipients. Although the plots tend to be sparse because of the relatively small sample size, the long-term care utilization by high-demand recipients is seen to be affected by the expansion of insurance coverage. These responses are consistent with the prediction of the moral hazard. Appendix Figure 3A.6 presents the plots using all recipients regardless of previous utilization. It shows that coverage expansion increases long-term care utilization at all thresholds, reflecting both behavioral and price effects.

Table 3.6 presents estimates of the effect of one-unit coverage expansion on long-term care utilization. For each threshold, the estimates of low-demand recipients and that of high-demand recipients are presented separately. The (1) - (8) columns show the estimates and standard errors of each specific polynomial order as well as nonparametric estimates. The first-stage estimates for coverage expansion are given in Appendix Table 3A.5, which shows the validity of instruments.

The estimation results shed light on some important aspects of behavioral biases associated with coverage expansion. First, a coverage expansion significantly increases long-term care utilization through behavioral effects for those with relatively low demand for long-term care. The estimates for low-demand recipients in Support level 1- 2, Care level 1 - 2, and 2 - 3 are around 0.2 - 0.3 and statistically significant. This result shows that these recipients are affected by the coverage expansion even when they do not face effective price changes. The interesting thing is that the estimates are almost identical among the three groups, which implies that the larger the coverage expansion is, the more low-demand recipients increase their long-term care utilization. This is consistent with the prediction of anchoring effects and heuristic thinking which are introduced as potential sources of behavioral effects.

Second, the magnitude of behavioral effects gradually decreases as recipients' care-needs level gets higher. The behavioral effects for recipients in Care level 3 - 4 is positive but not statistically significant. In the case of recipients with the highest need for care (Care level 4 - 5), the estimates are relatively noisy and insignificant. Although the insignificance could be due to less statistical power, these results may suggest that recipients with a high need for care are not affected by behavioral effects. For instance, recipients could be more serious about making care plan when their needs for long-term care are high. In this case, recipients are already using long-term care necessary for their lives, and they have little incentives to increase the services unless their utilization is constrained by a coverage limit. However, it is beyond the scope of this study to pin down the exact mechanisms of the heterogeneity of behavioral effects.

Third, estimates for high-demand recipients are statistically significant in almost all cases and the magnitudes are generally larger than those of behavioral effects. This strong reaction of high-demand recipients makes sense because the demand of these recipients is constrained by the full price for services beyond the insurance coverage. As oppose to low-demand recipients, estimates for high-demand recipients vary among different care-needs levels, with estimates at low care-needs levels tending to be larger than those at high care-needs levels. A one-unit expansion in insurance coverage increases long-term care utilization of high-demand recipients by around 0.6 - 0.8 units at low care-needs levels (Support level 1 - 2 or Care level 1 - 2) while it increases only 0.3 - 0.6 units at high care-needs levels. One interpretation of these results is that recipients who exhaust their insurance coverage at low care-needs levels are very rare and these exceptional recipients could have demand for the services well beyond the insurance coverage.

Fourth, except for Care level 3 - 4 and 4 - 5, estimates for all recipients are close to those of low-demand recipients. As shown in the summary statistics as well as the sample size of this estimation, a large portion of recipients uses long-term care services at a sufficiently lower level than the coverage limit, especially those with low care-needs levels. Therefore, the overall effect of insurance coverage can be mainly attributed to behavioral effects. By contrast, the estimates of Care level 3 - 4 and 4 - 5 indicate that the overall effect of insurance coverage for these recipients can be attributable to both behavioral and price effects.

3.5.3 The Health Consequences of Long-Term Care Utilization

For estimating health effects of long-term care, I use the thresholds of standardized care time which create discontinuous variation in insurance coverage as instruments. Appendix Table 3A.6 presents estimates of the first stage effect of insurance coverage on long-term care utilization. The focus of this analysis is health effects of long-term care utilization during the first certification term. The estimates show that the instruments have statistically significant effects on long-term care utilization for recipients with relatively low needs for long-term care (Support level 1 - 2, Care level 1 - 2 and 2 - 3), while the effects are statistically insignificant for those with high needs for long-term care (Care level 3 - 4 and 4 - 5). These results are consistent with the estimates for behavioral effects as shown in Table 3.6. Therefore, in terms of the health effects, I focus on recipients in Support level 1 - 2, Care level 1 - 2 and 2 - 3. Note that recipients in Support level utilize preventive care while those in Care level utilize regular long-term care.

Table 3.7 presents OLS and 2SLS estimates of the health consequences of long-term care utilization using standardized care time as a health outcome. Note that the positive values of the estimates imply a negative health effect of long-term care because the standardized care time reflects the recipient's greater need for long-term care. OLS estimates suggest that both preventive care (Support level) and regular care (Care level) could have a slightly negative effect on a recipient's health. The 2SLS estimates also tend to have a positive value (a negative effect), with estimates even larger than those of OLS for those with low needs for care. However, the significance of these 2SLS estimates depends on the specification, and the estimates are insignificant in many cases. Therefore, it is appropriate to conclude that, for both preventive and regular care, there are few short-run effects of long-term care on health outcomes.

Validity tests of the RD design suggest a manipulation that care-needs certification examiners avoid assigning recipients to lower care-needs level. It is then possible that recipients just above the threshold could be more likely to get higher standardized care time at next term than those just below the threshold, regardless of long-term care utilization. In other words, the exclusion restriction of the instrument might be violated when standardized care time is used as a health outcome. It is therefore important to check the health effects using different outcome variables. Table 3.8 presents OLS and 2SLS estimates of the health effects of long-term care, using long-term care utilization during the term $s+1$ ($Utilization_{s+1}$) and whether recipients end up entering a nursing home during $s+1$ as health outcomes. The results are consistent with those using standardized care time as health outcomes. Although OLS estimates indicate significant negative health effects of long-term care, the 2SLS estimates are mostly insignificant except the case where long-term care utilization is used as outcome for Care level 2 - 3. These results reinforce the argument that, for both preventive and regular care, there are little health effects of long-term care in the short run.

There are some empirical limitations on estimating the health consequences of long-term care in this setting. First, the effect of different types of long-term care such as home care and day care cannot be identified because there is only one instrument. Hence the estimates should be interpreted as the collective effect of all different types of long-term care services on health outcomes. Second, due to data limitations, it is beyond the scope of this study to estimate health effects of total long-term care that includes both long-term care services and informal care. Some previous studies have argued that informal care can be considered as a substitute for formal long-term care service (see Mommaerts, 2018 for example). These studies suggest that the change in long-term care utilization might be larger than that in total care because the change in utilization could be partially made up by informal care. Therefore, the above estimates should be interpreted as an upper bound of the health effects attributed to the total amount of care a recipient receives.

3.6 Discussion

3.6.1 Asymmetric Response to Insurance Coverage Expansion and Reduction

From a policy perspective, it is important to understand recipients' responses to both coverage expansion and reduction, because either policy could be implemented depending on the policy objective. If the effect of coverage reduction is not exactly the opposite to that of coverage expansion, then it would be misleading to simply extrapolate the estimates of coverage expansion to the case of coverage reduction.[24] The asymmetric response can potentially become an issue especially when behavioral effects are considered. For example, the psychological concept of the endowment effect predicts that individuals are less likely to reduce services they are using than increase them.[25] As shown by the validity test in section 3.4.4, the estimates for coverage reduction are likely to be confounded by unobservable factors. Nevertheless, given the important policy implication of this issue, I discuss behavioral responses to a coverage reduction and their differences from those of coverage expansion in this section. Because covariates balance test shows that observed covariates are fairly balanced around thresholds, I believe that the estimated impact of coverage reduction can be trustworthy to some extent.

Appendix Table 3A.7 and 3A.8 reports the first-stage and the second-stage estimates for the coverage reduction respectively. Appendix Table 3A.7 shows that the instruments are statistically significant and valid. I find that the coverage reduction affects long-term care utilization asymmetrically with coverage expansion. Most importantly, behavioral effects of coverage reduction are far smaller than those of coverage expansion, and the estimates are not statistically significant in most cases. In contrast, a coverage reduction significantly decreases the service utilization of high-demand recipients through price effects, because the coverage reduction would expose these recipients to a much higher price of services outside their coverage region.

This asymmetric response to coverage expansion and reduction leads to important policy implications. Since low-demand recipients do not respond to coverage reduction, the overall effect of a reduction in coverage tends to be smaller than that of coverage expansion.

[24]For example, Iizuka and Shigeoka (2019) uses variation of patient cost-sharing for children and shows that the response to price changes in opposite directions is not asymmetric.

[25]See Kahneman, Knetsch, and Thaler (1991) for details of the endowment effect.

Therefore, even if policymakers attempted to constrain service utilization through coverage reduction, the impact might be smaller than expected from the estimates of coverage expansion.

3.6.2 The Influence of Care Manager

Anchoring effects and heuristic thinking have been discussed as potential sources of behavioral biases due to coverage changes. However, in the context of LTCI, the practice of the care manager could also affect the utilization of long-term care, including that of low-demand recipients. This section discusses the influence of care managers and argues that the anchoring effect and heuristic thinking are more plausible causes of the response of low-demand recipients.

The interpretation of behavioral effects is not reliable if care managers propose a predetermined selection of services to recipients based on care-needs levels, so insurance coverage itself does not affect long-term care utilization. Care managers have specialized knowledge about long-term care and play an important role when recipients make a care plan. Therefore, low-demand recipients might be affected more by the predetermined care plan proposed by a care manager than by insurance coverage. Although it is impossible to rule out the influence of the care manager, the responses of low-demand recipients strongly suggest that insurance coverage itself has a significant effect. The estimation results in Table 3.6 show that the larger the coverage expansion is, the more low-demand recipients increase their long-term care utilization as long as behavioral effects are statistically significant. The consistency between recipients' responses and the degree of coverage expansion demonstrates the influence of insurance coverage. If only care managers affected service utilization, then a larger expansion in insurance coverage would not necessarily lead to more significant responses.

3.6.3 Aggregate Impact of Behavioral Effects

The significance of behavioral effects and the fact that a large proportion of recipients do not exhaust insurance coverage suggest the importance of taking into account behavioral biases in the design of insurance. To quantify the aggregate impact of behavioral effects, I calculate the impact of a hypothetical policy on long-term care costs with and without behavioral effects based on a "back-of-the-envelope" calculation.

One of the policy issues of LTCI is that some recipients exhaust insurance coverage and are considered to not being able to receive long-term care services enough to fulfill their needs. Therefore, I consider the policy of expanding all recipients' insurance coverage by raising their care-needs levels by one-stage (except Support level 2 and Care level 5). Following the estimation procedure so far, I divide recipients by whether their utilization is more than 80% of coverage limits. [26] The estimates from Table 3.6 are used as the effect of coverage expansion on low-demand and high-demand recipients. Because the estimates vary with specifications, I calculate both the upper and lower bounds of aggregate impact by using the largest and smallest estimates. In choosing the estimates of behavioral effects, those exhibiting negative values are excluded because they could be affected by the inappropriate functional specification. Recipients as of January 2017 are used for the calculation.

Appendix Table 3A.9 presents how much the coverage expansion would increase long-term care costs both in the municipality and national level.[27] When only high-demand recipients are considered, the coverage expansion increases total long-term care costs by 2 - 4%. However, the impact on low-demand recipients are much larger than high-demand ones: the same coverage expansion increases the total costs by 7 - 11% through behavioral effects. This means that if policymakers make a prediction of the policy impact based only on those who could be affected by price effects, the increase in long-term care costs would be predicted as less than one-third of "real" increases which include behavioral effects. Given the total cost of home-based long-term care at the national level amount to 371 billion JPY (3.71 billion USD), the consequences of misguided prediction would lead to sizable social costs.

[26] Although I define low-demand recipients as those whose utilization is lower than 50% of coverage limits in the estimation, I change the criterion to 80% for completeness of this analysis.

[27] To calculate the policy impact at the national level, I assume that the distribution of long-term care utilization in each care-needs level is the same as those of the municipality I used for the estimation. The total cost of home-based long-term care at the national level comes from the estimate of the Ministry of Health, Labour and Welfare.

3.7 Conclusion

This paper explores the effect of insurance coverage on long-term care utilization and its health consequences. In particular, I estimate the impact of behavioral effects that might induce recipients to respond to changes in insurance coverage even when the price they face does not change. The institutional setting of LTCI in Japan permits the implementation of an RD design to estimate the effect of changes in insurance coverage on the long-term care utilization of recipients with different health conditions. To identify behavioral effects, this study focuses on recipients who do not exhaust their insurance coverage and estimate the effect of coverage expansion on the service utilization of these low-demand recipients. Also, the health consequences of long-term care utilization are estimated by using the thresholds as instruments.

The estimation results indicate that behavioral effects are the main causes of recipients' responses to changes in insurance coverage. For recipients with relatively low needs for long-term care, a one-unit expansion in coverage increases long-term care utilization by about 0.3 units through behavioral effects, with price effects for these recipients about twice as large. Overall effects of coverage expansion strongly reflect the magnitude of behavioral effects because a majority of recipients do not exhaust their insurance coverage and are only influenced by behavioral effects rather than price effects. For recipients with high needs for long-term care, however, behavioral effects are not statistically significant and only price effects affect their long-term care utilization. This heterogeneity in behavioral effects suggests that the magnitude of psychological biases could depend on an individual's health condition. The estimates for health effects indicate that long-term care utilization has little effect on health, at least in the short run.

The findings of this paper highlight the importance of taking into account non-standard decision-making frameworks in the analysis of social insurance policies. An optimal social insurance policy should reflect the total cause including both price and behavioral effects of the responses to changes in insurance coverage. I show that decision-makers are significantly influenced by behavioral biases even when they face relatively simple static decision-making. This suggests that behavioral biases might be prevalent in many public policy settings. To predict the impact of behavioral biases on public policy, it is useful to find specific mechanisms for these biases. Although I presented some potential mechanisms, it is beyond the scope of this study to identify exact mechanisms and quantify their relative importance. In terms of health effects, it is also important to quantify the effect of long-term care utilization on the health of family members, as the utilization could affect the provision of informal care by the family. These can be interesting avenues for future research.

Bibliography

[1] **Abaluck, Jason, Jonathan Gruber, and Ashley Swanson.** 2018. "Prescription Drug Use under Medicare Part D: A Linear Model of Nonlinear Budget Sets." *Journal of Public Economics*, 164, pp. 106-138.

[2] **Ariely, Dan, George Loewenstein, and Drazen Prelec.** 2003. " "Coherent Arbitrariness": Stable Demand Curves without Stable Preferences." *Quarterly Journal of Economics*, 118(1), pp. 73-106.

[3] **Aron-Dine, Aviva, Liran Einav, Amy Finkelstein, and Mark Cullen.** 2015. "Moral Hazard in Health Insurance: Do Dynamic Incentives Matter?" *Review of Economics and Statistics*, 97(4), pp. 725-741.

[4] **Baicker, Katherine, Sendhil Mullainathan, and Joshua Schwartzstein.** 2015. "Behavioral Hazard in Health Insurance." *Quarterly Journal of Economics*, 130(4), pp. 1623-1667.

[5] **Beggs, Alan and Kathryn Graddy.** 2009. "Anchoring Effects: Evidence from Art Auctions." *American Economic Review*, 99(3), pp. 1027-1039.

[6] **Bernheim, B. Douglas, Andrey Fradkin, and Igor Popov.** 2015. "The Welfare Economics of Default Options in 401(k) Plans." *American Economics Review*, 105(9), pp. 2798-2837.

[7] **Brot-Goldberg, Zarek C., Amitabh Chandra, Benjamin R. Handel, and Jonathan T. Kolstad.** 2017. "What Does a Deductible Do? The Impact of Cost-Charing on Health Care Prices, Quantities, and Spending Dynamics." *Quarterly Journal of Economics*, 132(3), pp. 1261-1318.

[8] **Bucchianeri, Grace W. and Julia A. Minson.** 2013. "A Homeowner's Dilemma: Anchoring in Residential Real Estate." *Journal of Economics Behavior & Organization*, 89, pp. 76-92.

[9] **Cabral, Marika.** 2017. "Claim Timing and *Ex Post* Adverse Selection." *Review of Economic Studies*, 84(1), pp. 1-44.

[10] **Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik.** 2014. "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs." *Econometrica*, 82(6), pp. 2295-2326.

[11] **Campbell, John Creighton, and Naoki Ikegami.** 2000. "Long-Term Care Insurance Comes To Japan." *Health Affairs*, 19(3), pp. 26-39.

[12] **Chang, Eric C., Tse-Chun Lin, Yan Lup, and Jinjuan Ren.** 2019. "Ex-Day Returns of Stock Distributions: An Anchoring Explanation." *Managenent Science*, 65(3), pp. 1076-1095.

[13] **Choi, James J., Emily Haisley, Jennifer Kurkoski, and Cade Massey.** 2017. "Small Cues Change Saving Choices." *Journal of Economic Behavior & Organization*, 142, pp. 378-395.

[14] **Coussens, Stephen.** 2018. "Behaving Discretely: Heuristic Thinking in the Emergency Department." *Working Paper.*

[15] **Dalton, Christina M., Gautam Gowrisankaran, and Robert Town.** 2019. "Salience, Myopia, and Complex Dynamic Incentives: Evidence from Medicare Part D." *Review of Economic Studies*, forthcoming.

[16] **Dougal, Casey, Joseph Engelberg, Christopher A. Parsons, and Edward D. Van Wesep.** 2015. "Anchoring on Credit Spreads." *Journal of Finance*, 70(3), pp. 1039-1080.

[17] **Einav, Liran, and Amy Finkelstein.** 2018. "Moral Hazard in Health Insurance: What We Know and How We Know It." *Journal of the European Economic Association*, 16(4), pp. 957-982.

[18] **Einav, Liran, Amy Finkelstein, and Paul Schrimpf.** 2015. "The Response of Drug Expenditure to Nonlinear Contract Design: Evidence from Medicare Part D." *Quarterly Journal of Economics*, 130(2), pp. 841-899.

[19] **Einav, Liran, Amy Finkelstein, and Paul Schrimpf.** 2017. "Bunching at the Kink: Implications for Spending Responses to Health Insurance Contracts." *Journal of Public Economics*, 146, pp. 27-40.

[20] **Feldman, Naomi E., Peter Katuščák, and Laura Kawano.** 2016. "Taxpayer Confusion: Evidence from the Child Tax Credit." *American Economic Review*, 106(3), pp. 807-835.

[21] **Foster, Andrew D. and Yong Suk Lee.** 2015. "Staffing Subsidies and the Quality of Care in Nursing Homes." *Journal of Health Economics*, 41, pp. 133-147.

[22] **Friedrich, Benjamin U. and Martin B. Hackmann.** 2018. "The Returns to Nursing: Evidence from a Parental-Leave Program." *Working Paper.*

[23] **Gathergood, John, Neale Mahoney, Neil Stewart, and Jörg Weber.** 2019. "How Do Individuals Repay Their Debt? The Balance-Matching Heuristic." *American Economic Review*, 109(3), pp. 844-875.

[24] **Gill, Thomas M., Dorothy I. Baker, Margaret Gottschalk, Peter N. Peduzzi, Heather Allore, and Amy Byers.** 2008. "A Program to Prevent Functional Decline in Physically Frail, Elderly Persons Who Live at Home." *New England Journal of Medicine*, 347(14), pp. 1068-1074.

[25] **Guo, Audrey and Jonathan Zhang.** 2019. "What to Expect When You Are Expecting: Are Health Care Consumers Forward-Looking?" *Journal of Health Economics*, forthcoming.

[26] **Iizuka, Toshiaki and Hitoshi Shigeoka.** 2019. "Free for Children? Patient Cost-Sharing and Health Care Utilization." *Working Paper*.

[27] **Imbens, Guido and Karthik Kalyanaraman.** 2012. "Optimal Bandwidth Choice for the Regression Discontinuity Estimator." *Review of Economic Studies*, 79(3), pp. 933-959.

[28] **Jetter, Michael and Jay K. Walker.** 2017. "Anchoring in Financial decision-making: Evidence from Jeopardy!" *Journal of Economic Behavior & Organization*, 141, pp. 164-176.

[29] **Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler.** 1991. "The Endowment Effect, Loss Aversion, and Status Quo Bias." *Journal of Economic Perspectives*, 5(1), pp. 193-206.

[30] **Kim, Hyuncheol Bryant and Wilfredo Lim.** 2015. "Long-Term Care Insurance, Informal Care, and Medical Expenditures." *Journal of Public Economics*, 125, pp. 128-142.

[31] **Kowalski, Amanda.** 2015. "Estimating the Tradeoff Between Risk Protection and Moral Hazard with a Nonlinear Budget Set Model of Health Insurance." *International Journal of Industrial Organization*, 43, pp. 122-135.

[32] **Lin, Haizhen.** 2014. "Revisiting the Relationship between Nurse Staffing and Quality of Care in Nursing Homes: An Instruments Approach." *Journal of Health Economics*, 37, pp. 13-24.

[33] **Lacetera, Nicola, Devin G. Pope, and Justin R. Sydnor.** 2012. "Heuristic Thinking and Limited Attention in the Car Market." *American Economic Review*, 102(5), pp. 2206-2236.

[34] **Lee, David S., and Thomas Lemieux.** 2010. "Regression Discontinuity Designs in Economics." *Journal of Economic Literature*, 48(2), pp. 281-355.

[35] **McCrary, Justin.** 2008. "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test." *Journal of Econometrics*, 142(2), pp. 698-714.

[36] **Mommaerts, Corina.** 2018. "Are Coresidence and Nursing Homes Substitute? Evidence from Medicaid Spend-down Provisions." *Journal of Health Economics*, 59, pp. 125-138.

[37] **Tamiya, Nanako, Haruko Noguchi, Akihiro Nishi, Michael R. Reich, Naoki Ikegami, Hideki Hashimoto, Kenji Shibuya, Ichiro Kawachi, and John Creighton Campbell.** 2011. "Population Ageing and Wellbeing: Lessons from Japan's Long-Term Care Insurance Policy." *Lancet*, 378, pp. 1183-1192.

[38] **Tsutsui, Takako and Naoko Muramatsu.** 2005. "Care-Needs Certification in the Long-Term Care Insurance System of Japan." *International Health Affairs*, 53(3), pp. 522-527.

[39] **Tversky, Amos, and Daniel Kahneman.** 1974. "Judgment under Uncertainty: Heuristics and Biases." *Science*, 185(4157), pp. 1124-1131.

[40] **Wansink, Brian, Robert J. Kent, and Stephen J. Hoch.** 1998. "An Anchoring and Adjustment Model of Purchase Quantity Decisions." *Journal of Marketing Research*, 35(1), pp. 71-81.

Figure 3.1: The Utilization Process of Long-Term Care Services under the LTCI

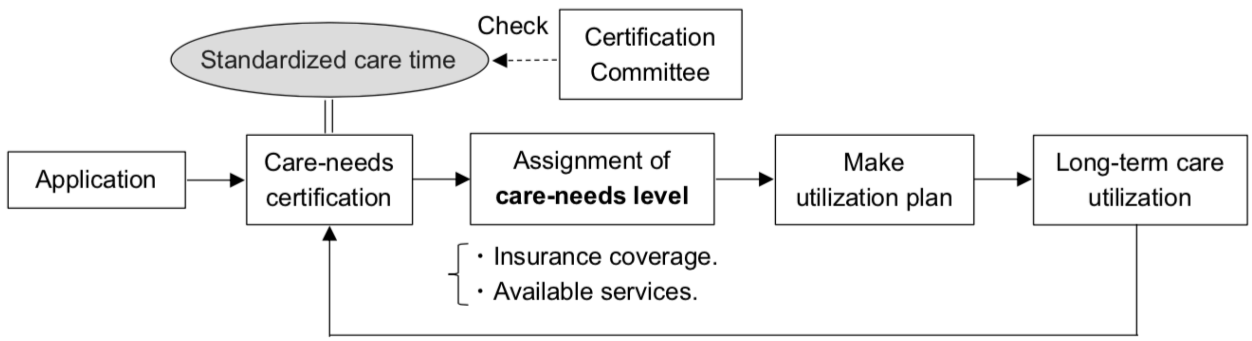
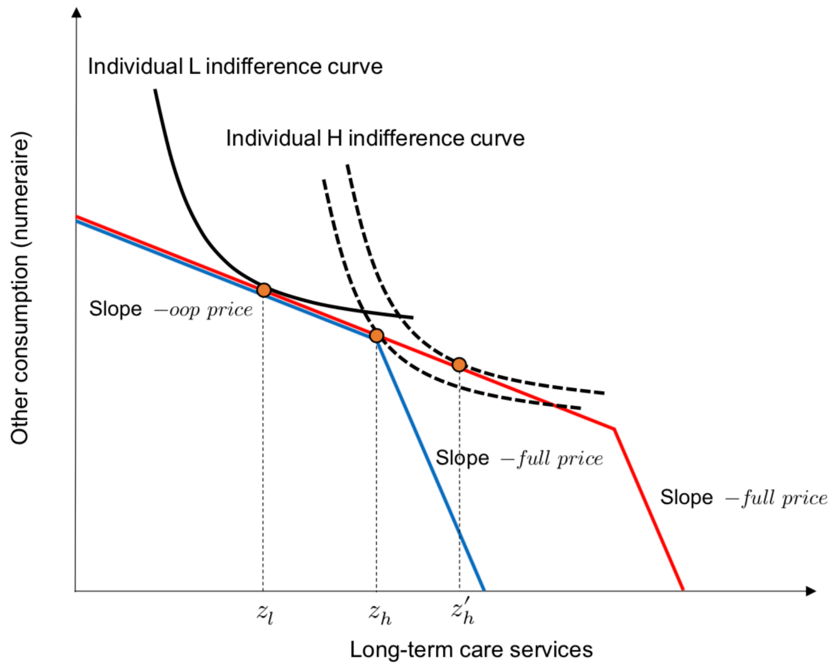
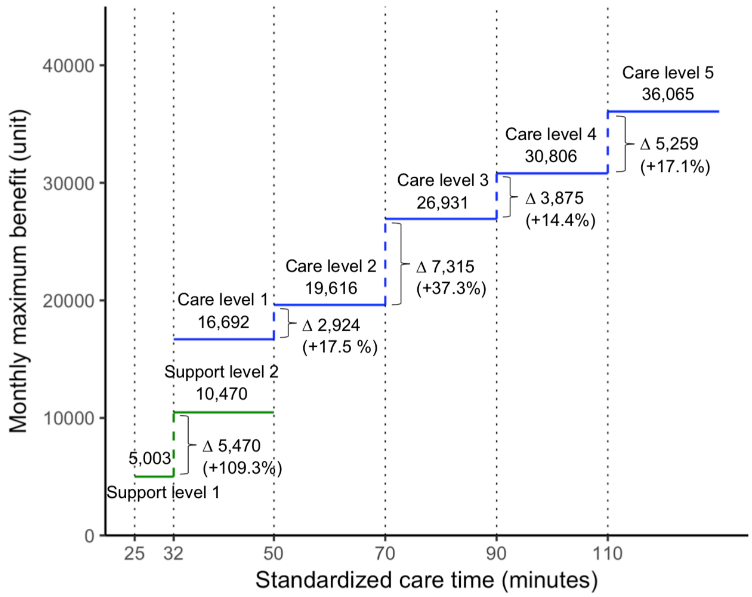


Figure 3.2: The Response to Insurance Coverage due to Moral Hazard



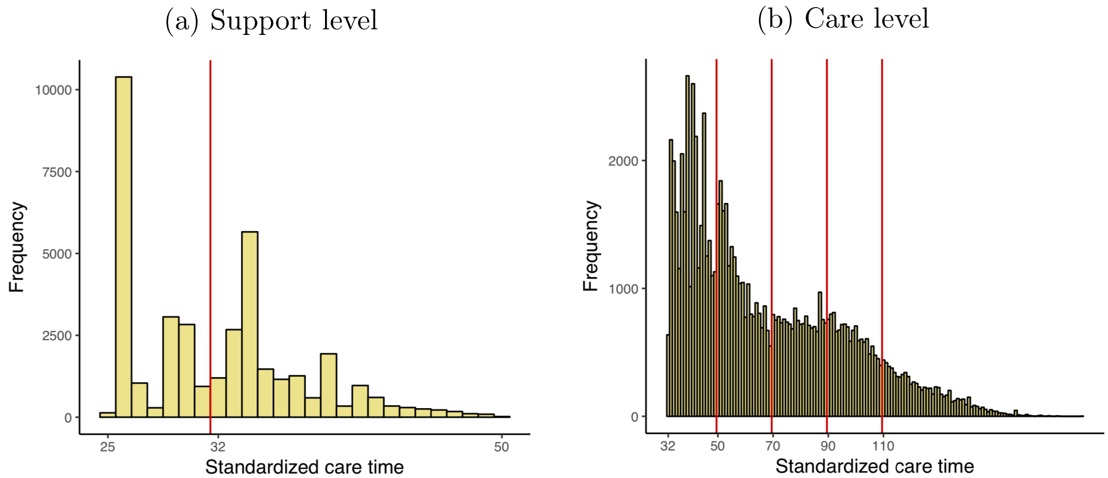
Notes: This figure illustrates a recipient's response to changes in insurance coverage due to moral hazard. Blue and red lines are budget constraints with different coverage region. Individual H alters her demand for long-term care (z_h and z_h') according to a switch of coverage generosity. In contrast, Individual L does not respond to the change in coverage, because she does not face any change in economic environments including price. Thus, based on standard theory, individuals whose long-term care utilization is sufficiently lower than coverage limit should not respond to changes in insurance coverage.

Figure 3.3: Standardized Care Time and Monthly Coverage Limit



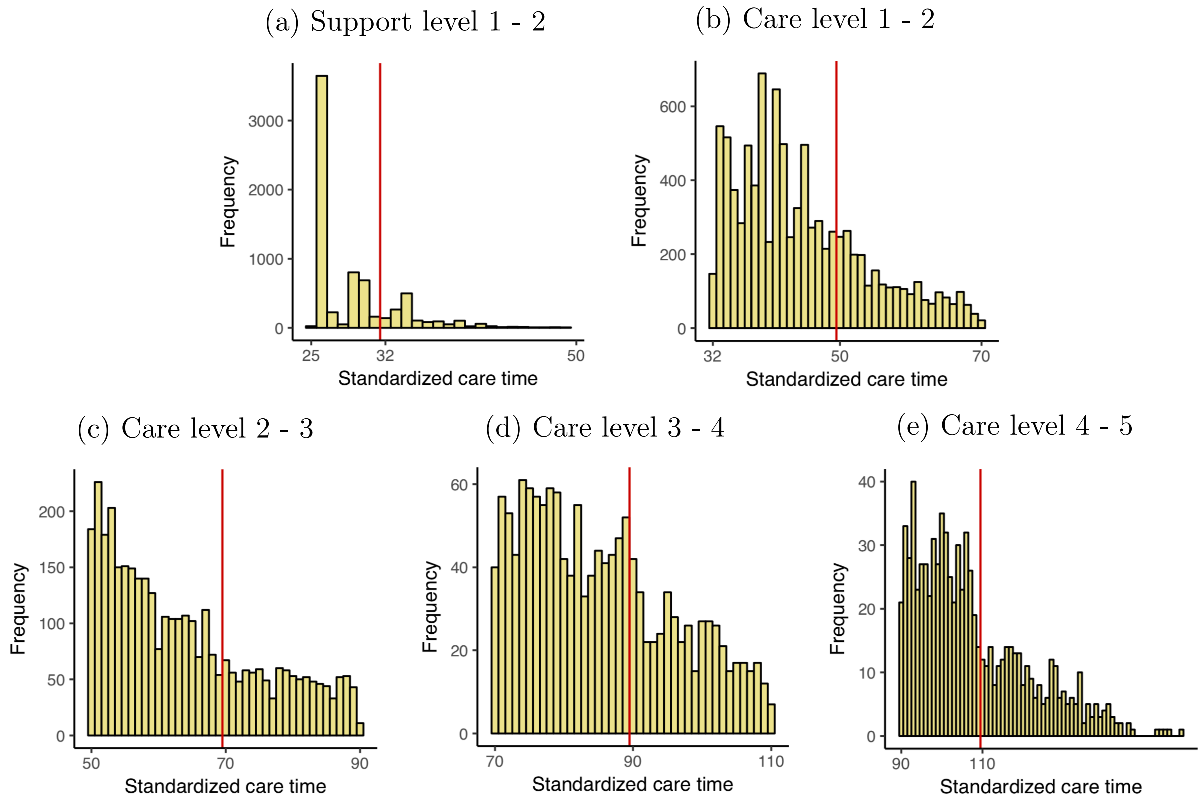
Notes: This figure illustrates the relationship between standardized care time and monthly coverage limit. The green line represents the coverage limits for the Support level, and the blue line represents those for the Care level.

Figure 3.4: Distribution of Standardized Care Time (All Recipients)



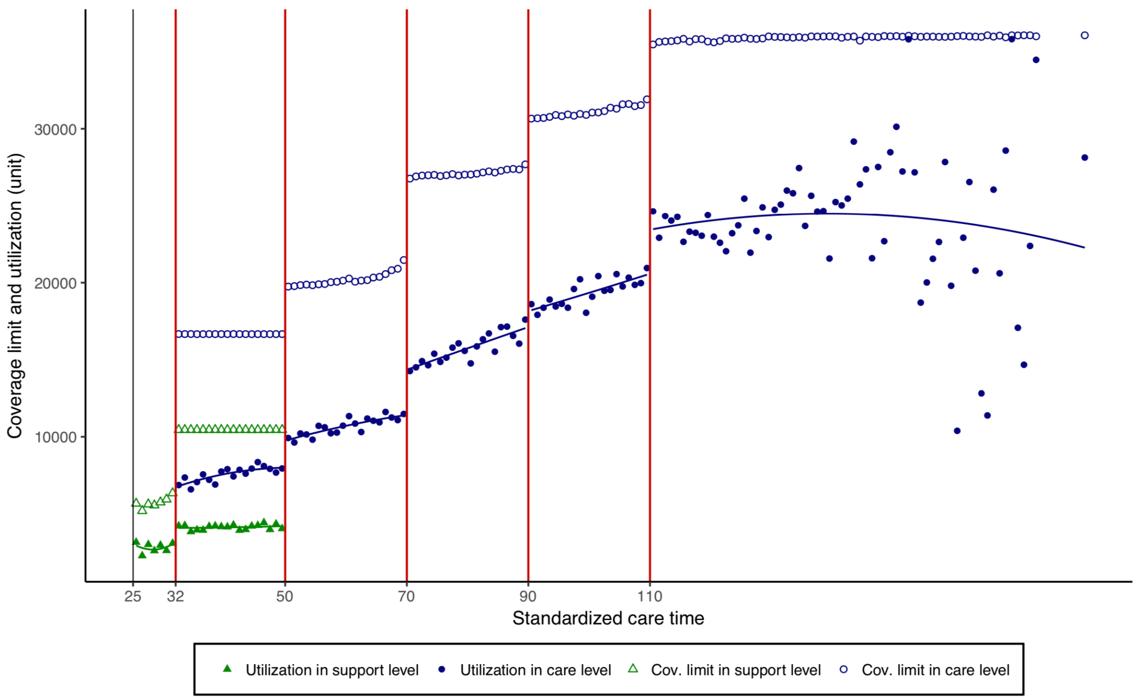
Notes: This figure presents the distribution of standardized care time for Support level (Panel (a)) and Care level (Panel (b)). The distribution is constructed using all recipients in the baseline sample. The bin width is 1 minute of standardized care time.

Figure 3.5: Distribution of Standardized Care Time (Coverage Expansion, Low-Demand)



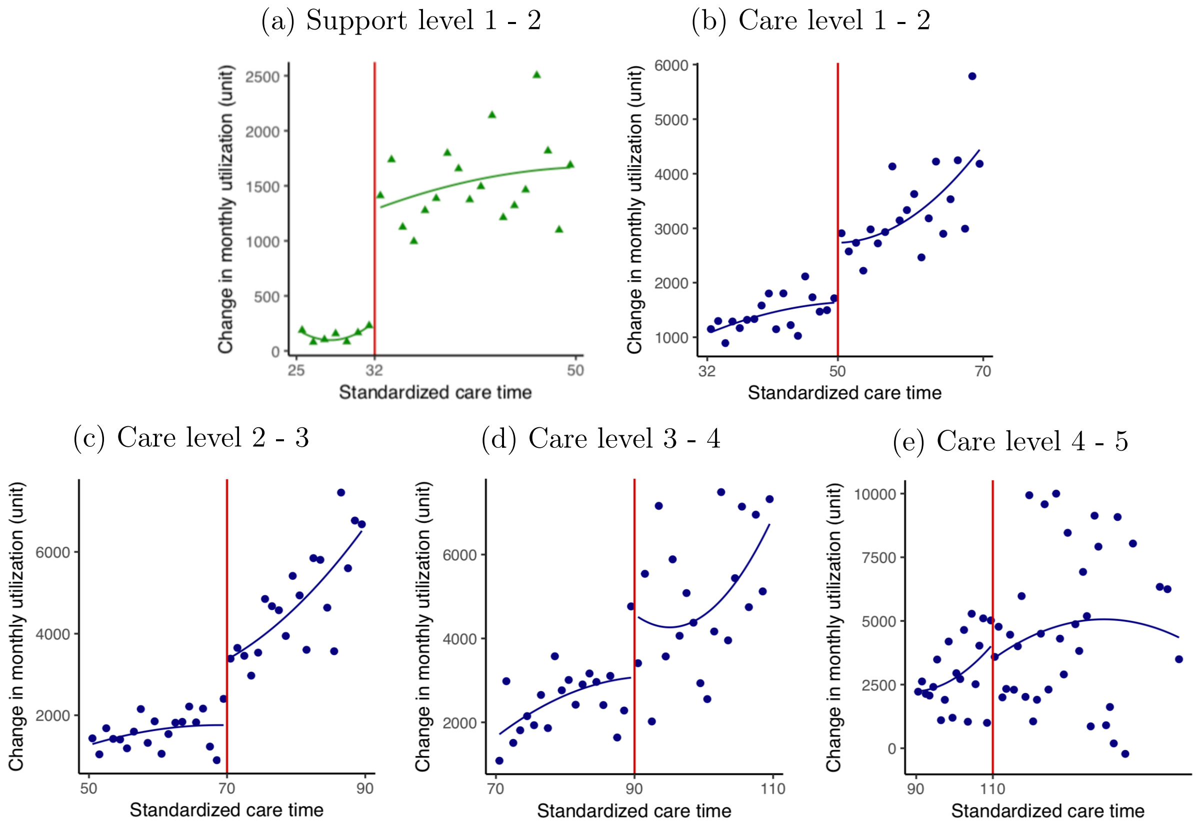
Notes: This figure presents the distribution of standardized care time for each care-needs level. The distribution is constructed using the low-demand recipients after the reclassification of care-needs levels (either the same or one-stage higher level). The bin width is 1 minute of standardized care time.

Figure 3.6: Coverage Limit and Long-Term Care Utilization



Notes: This figure presents the relationship between coverage limit and long-term care utilization using the baseline sample. I divide standardized care time into 1 minute-wide interval (bins) and each hollow circles and triangle represents a local average of coverage limit for Care level and Support level respectively. Each filled circle and triangle represents a local average of a monthly total unit of long-term care utilization within bins and a quadratic prediction is fitted on the plots for each care-needs level.

Figure 3.7: Changes in Long-Term Care Utilization (Coverage Expansion, Low-Demand)



Notes: This figure presents the effect of expanding insurance coverage on long-term care utilization using low-demand recipients who end up being either the same or one stage worse after the reclassification. I divide standardized care time into 1 minute-wide interval (bins) and each plot represents a local average of changes in monthly long-term care utilization before and after the reclassification. I plot the local average for both the Support level (Panel (a)) and the Care level (Panel (b)-(e)) and a quadratic prediction is fitted on the plots for each care-needs level.

Table 3.1: Category of Assistance and Range of Time Length

Category of assistance	Range of time length (minutes)
Eating	1.1 – 71.4
Transferring	0.4 – 21.4
Toileting	0.2 – 28.0
Hygiene	1.2 – 24.3
Housework	0.4 – 11.3
Dementia	5.8 – 21.2
Exercise	0.5 – 15.4
Medical care	1.0 – 37.2
Standardized care time	10.6 – 230.6

Table 3.2: Monthly Coverage Limits for Each Care-needs Level

Care-needs level	Standardized care time	Coverage cap (unit)
Support level 1	25.0 – 31.9	5,003
Support level 2	32.0 – 49.9	10,473
Care level 1	32.0 – 49.9	16,692
Care level 2	50.0 – 69.9	19,616
Care level 3	70.0 – 89.9	26,931
Care level 4	90.0 – 109.9	30,806
Care level 5	$\geq$110.0	36,065

Table 3.3: Summary Statistics

		Support level		Care level				
	Baseline sample	level 1	level 2	level 1	level 2	level 3	level 4	level 5
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Demographics								
Age	81.2	81.4	81.4	81.7	81.5	82.1	82.2	82.0
Woman	0.60	0.67	0.70	0.61	0.60	0.58	0.59	0.59
20% coinsurance	0.13	0.11	0.10	0.12	0.11	0.11	0.10	0.09
Change of care-needs level	0.05	0.00	0.06	0.07	0.01	0.09	0.09	0.07
Obs. (Recipient)	49,248	9,204	10,294	20,121	16,395	11,858	10,026	6,412
B. Care-needs certification								
Standardized care time	55.7	27.1	35.8	39.3	55.6	77.7	97.9	123.5
Eating	7.5	3.5	3.5	5.2	6.7	8.4	11.5	28.5
Transferring	6.1	0.7	2.1	2.4	5.1	11.0	17.1	17.6
Toileting	7.0	0.3	1.7	1.8	6.0	14.3	21.1	22.5
Hygiene	8.9	2.2	5.8	6.3	9.8	13.6	16.7	18.2
Housework	6.8	4.6	5.5	7.2	7.9	8.2	8.4	6.1
Dementia	6.8	5.8	5.9	6.5	7.5	7.9	7.0	6.9
Exercise	6.3	6.0	6.7	5.7	6.4	5.8	6.7	7.6
Medical care	5.8	4.0	4.6	4.2	5.1	5.9	9.1	16.1
Obs. (Recipient $\times$ Term)	132,881	18,605	20,845	32,552	23,530	15,808	12,807	8,734
C. Long-term care utilization								
Total expenditure per month (JPY)	105,120	25,078	43,259	73,615	105,681	167,705	207,286	262,789
Total unit per month	10,657	2,345	4,068	7,443	10,803	17,165	21,191	26,977
Home care	2,738	714	881	1,647	2,301	3,694	6,037	10,393
Day care	3,927	920	1,660	3,273	4,050	6,967	7,084	6,731
Home-visit nursing care	639	60	177	396	598	769	1,331	2,763
Rehabilitation	1,032	212	638	1,079	1,410	1,399	1,381	918
Exceed coverage limit	0.09	0.05	0.03	0.05	0.11	0.14	0.18	0.24
Monthly coverage limit (unit)		5,003	10,473	16,692	19,616	26,931	30,806	36,065
Obs. (Recipient $\times$ Month)	1,313,343	163,854	187,264	335,340	261,823	166,438	119,183	79,441

Notes: This table shows summary statistics for LTCI recipients who are analyzed in this study: recipients who utilize long-term care services between June 2012 and March 2018. The first column shows statistics for all recipients regardless of care-needs levels. The rest of the columns present statistics for recipients who belong to specific care-needs level separately. Note that recipients could be categorized into different care-needs levels for several care-needs certifications. Hence, the sum of recipients of each care-needs level is not equal to the number of total recipients (first column).

Table 3.4: Covariates Balance Tests (Coverage Expansion)

	All recipients				Low-demand recipients				High-demand recipients			
	Linear		Quadratic		Linear		Quadratic		Linear		Quadratic	
	Coef. (1)	SE (2)	Coef. (3)	SE (4)	Coef. (5)	SE (6)	Coef. (7)	SE (8)	Coef. (9)	SE (10)	Coef. (11)	SE (12)
Support level 1 - 2												
Age	0.069	(0.330)	1.627***	(0.613)	-0.140	(0.383)	1.072	(0.712)	-0.814	(1.106)	-0.198	(1.741)
Female	0.000	(0.020)	-0.093**	(0.040)	0.006	(0.025)	-0.074	(0.049)	0.056	(0.069)	-0.159	(0.123)
20% coinsurance	-0.001	(0.011)	0.056**	(0.022)	-0.008	(0.015)	0.030	(0.028)	-0.020	(0.043)	0.172**	(0.075)
Observation	10,903				7,112				847			
Cluster	5,631				4,039				538			
Care level 1 - 2												
Age	1.046***	(0.271)	1.530***	(0.397)	0.769**	(0.347)	1.381***	(0.505)	0.812	(0.833)	0.959	(1.254)
Female	-0.021	(0.016)	-0.002	(0.024)	-0.037*	(0.021)	0.001	(0.031)	-0.028	(0.045)	0.058	(0.067)
20% coinsurance	0.011	(0.009)	0.026*	(0.014)	0.004	(0.013)	0.016	(0.019)	0.018	(0.021)	0.063**	(0.032)
Observation	14,806				9,366				1,730			
Cluster	10,454				7,136				1,354			
Care level 2 - 3												
Age	0.515	(0.434)	0.535	(0.666)	0.638	(0.726)	0.154	(1.079)	0.375	(0.778)	2.506**	(1.247)
Female	0.031	(0.024)	0.018	(0.036)	0.037	(0.038)	0.064	(0.058)	0.026	(0.043)	0.041	(0.066)
20% coinsurance	-0.024*	(0.012)	-0.044**	(0.020)	-0.052***	(0.020)	-0.014	(0.023)	-0.016	(0.026)	-0.039	(0.036)
Observation	8,240				3,586				2,310			
Cluster	6,439				2,934				1,920			
Care level 3 - 4												
Age	-0.572	(0.541)	-0.803	(0.796)	2.687**	(1.044)	2.462	(1.508)	-1.707**	(0.732)	-1.964*	(1.087)
Female	0.019	(0.028)	-0.004	(0.040)	0.005	(0.055)	0.037	(0.078)	-0.006	(0.040)	-0.012	(0.057)
20% coinsurance	-0.005	(0.014)	-0.008	(0.020)	0.047	(0.030)	0.030	(0.040)	-0.036*	(0.020)	-0.037	(0.029)
Observation	4,813				1,445				2,075			
Cluster	3,867				1,217				1,725			
Care level 4 - 5												
Age	-0.376	(0.659)	-0.026	(0.924)	-1.384	(1.580)	-1.254	(2.137)	-0.848	(0.869)	-0.042	(1.230)
Female	0.015	(0.032)	0.015	(0.047)	0.036	(0.068)	0.103	(0.102)	-0.021	(0.043)	-0.014	(0.063)
20% coinsurance	0.014	(0.016)	0.019	(0.024)	-0.010	(0.033)	-0.036	(0.054)	0 .038*	(0.021)	0.040	(0.028)
Observation	3,435				815				1,785			
Cluster	2,700				686				1,433			

Notes: This table shows the estimates of covariate balance tests for the linear and quadratic RD specification from equation 3.7. These estimation use recipients who are used for analyzing coverage expansion after the reclassification of care-needs levels. Low-demand recipients are those whose average monthly long-term care utilization during the previous certification term is less than 50% of a given coverage limit. High-demand recipients are those whose average monthly long-term care utilization during the previous certification term is higher than 80% of a given coverage limit. *** p<0.01, ** p<0.05, * p<0.10.

Table 3.5: The Overall Effects of Insurance Coverage

	Linear		Quadratic		Cubic		LPR			
	Coef. (1)	SE (2)	Coef. (3)	SE (4)	Coef. (5)	SE (6)	Coef. (7)	SE (8)	Obs. (9)	Cluster (10)
Support level 1 - 2										
First-Stage	4,183.6***	(49.9)	3,859.1***	(107.2)	3,491.7***	(196.7)	3,291.5***	(218.5)	339,271	15,646
Second-Stage	0.260***	(0.014)	0.290***	(0.027)	0.176***	(0.049)	0.122**	(0.058)		
Care level 1 - 2										
First-Stage	2,957.7***	(22.5)	3,135.6***	(28.8)	2,990.2***	(34.3)	3,038.4***	(38.7)	549,574	28,607
Second-Stage	0.550***	(0.043)	0.524***	(0.057)	0.566***	(0.080)	0.614***	(0.109)		
Care level 2 - 3										
First-Stage	5,947.5***	(54.7)	5,729.2***	(92.8)	5,360.7***	(133.7)	5,193.8***	(161.7)	398,238	22,848
Second-Stage	0.459***	(0.037)	0.533***	(0.056)	0.477***	(0.082)	0.447***	(0.094)		
Care level 3 - 4										
First-Stage	3,101.0***	(39.5)	3,080.6***	(56.8)	2,965.0***	(75.4)	2,856.6***	(104.1)	274,045	18,801
Second-Stage	0.522***	(0.097)	0.455***	(0.141)	0.542***	(0.189)	0.427	(0.263)		
Care level 4 - 5										
First-Stage	3,942.4***	(57.5)	3,687.9***	(87.3)	3,630.3***	(120.2)	3,619.0***	(123.2)	189,281	14,371
Second-Stage	0.607***	(0.108)	0.726***	(0.156)	1.051***	(0.207)	1.003***	(0.224)		

Notes: This table presents the first-stage estimates of β_0^c in equation 3.2 and the second-stage estimates of β^c in equation 3.1. The first to sixth columns represent the estimates for different specification of $f_0(Caretime_{it})$ and $f(Caretime_{it})$: linear, quadratic and cubic respectively. The seventh and eighth columns shows the nonparametric local polynomial regression (LPR) estimates using the robust confidence intervals proposed by Calonico, Cattaneo, and Titiunik (2014). *** p<0.01, ** p<0.05, * p<0.10.

Table 3.6: The Effects of Coverage Expansion on Long-Term Care Utilization

	Linear		Quadratic		Cubic		LPR			
	Coef. (1)	SE (2)	Coef. (3)	SE (4)	Coef. (5)	SE (6)	Coef. (7)	SE (8)	Obs. (9)	Cluster (10)
<u>Support level 1 - 2</u>										
All recipients	0.235***	(0.013)	0.242***	(0.021)	0.289***	(0.031)	0.221***	(0.062)	110,120	5,631
Low-demand	0.202***	(0.014)	0.205***	(0.023)	0.237***	(0.035)	0.171***	(0.057)	71,747	4,039
High-demand	0.528***	(0.060)	0.574***	(0.087)	0.678***	(0.122)	0.732**	(0.300)	8,548	538
<u>Care level 1 - 2</u>										
All recipients	0.321***	(0.046)	0.345***	(0.065)	0.328***	(0.083)	0.339***	(0.106)	192,276	10,454
Low-demand	0.252***	(0.056)	0.328***	(0.079)	0.316***	(0.102)	0.407***	(0.125)	121,190	7,136
High-demand	0.682***	(0.182)	0.596**	(0.263)	0.413	(0.321)	0.247	(0.397)	22,236	1,354
<u>Care level 2 - 3</u>										
All recipients	0.250***	(0.036)	0.211***	(0.054)	0.237***	(0.075)	0.307***	(0.093)	110,245	6,439
Low-demand	0.184***	(0.054)	0.260***	(0.080)	0.241**	(0.110)	0.273**	(0.138)	55,524	3,231
High-demand	0.323***	(0.079)	0.186	(0.121)	0.319*	(0.176)	0.370	(0.230)	28,225	1,920
<u>Care level 3 - 4</u>										
All recipients	0.504***	(0.106)	0.517***	(0.147)	0.494*	(0.190)	0.198	(0.241)	63,109	3,867
Low-demand	0.230	(0.196)	0.453	(0.279)	0.218	(0.363)	-0.460	(0.470)	20,085	1,217
High-demand	0.700***	(0.161)	0.561***	(0.216)	0.488*	(0.279)	0.281	(0.269)	25,902	1,725
<u>Care level 4 - 5</u>										
All recipients	0.170*	(0.098)	0.223	(0.152)	0.297	(0.208)	0.441*	(0.266)	43,881	2,700
Low-demand	0.005	(0.235)	-0.377	(0.369)	-0.617	(0.495)	-0.659	(0.665)	10,760	686
High-demand	0.355***	(0.129)	0.487**	(0.202)	0.609**	(0.282)	0.690**	(0.308)	22,016	1,433

Notes: This table presents the second-stage estimates of β^c in equation 3.3. The first to sixth columns represent the estimates for different specifications of $f(Caretime_{it})$: linear, quadratic and cubic respectively. The seventh and eighth columns show the nonparametric local polynomial regression (LPR) estimates using the robust confidence intervals proposed by Calonico, Cattaneo, and Titiunik (2014). Low-demand recipients are those whose average monthly long-term care utilization during the previous certification term is less than 50% of a given coverage limit. High-demand recipients are those whose average monthly long-term care utilization during the previous certification term is more than 80% of a given coverage limit. *** p<0.01, ** p<0.05, * p<0.10.

Table 3.7: The Health Effect of Long-Term Care Utilization (1)

		2SLS				
	OLS (1)	Linear (2)	Quadratic (3)	Cubic (4)	LPR (5)	Obs. (6)
Preventive care						
Support level 1 - 2	0.0000	0.0001	0.0010	0.0027	0.0028	4,879
	(0.0001)	(0.0008)	(0.0014)	(0.0026)	(0.0026)	
Regular care						
Care level 1 - 2	0.0004***	0.0013	-0.0050	-0.0018	0.0000	5,515
	(0.0001)	(0.0016)	(0.0189)	(0.0027)	(0.0060)	
Care level 2 - 3	0.0006***	0.0010	0.0015	0.0006	0.0019	3,140
	(0.0001)	(0.0012)	(0.0010)	(0.0009)	(0.0017)	

Notes: This table presents the estimates of β^c in equation 3.5 using standardized care time at the beginning of the term $s+1$ as a health outcome. The first column represents the OLS estimates and second to fifth columns represent the 2SLS estimates. The 2SLS estimates are separately presented for different specifications of $f(Caretime_{it})$: linear, quadratic, cubic and LPR representing nonparametric local polynomial regression estimates using the robust confidence intervals proposed by Calonico, Cattaneo, and Titiunik (2014). In the case of OLS, $f(Caretime_{it})$ is linear. Standard errors in parentheses under each estimate. *** p<0.01, ** p<0.05, * p<0.10.

Table 3.8: The Health Effect of Long-Term Care Utilization (2)

	Utilization					Nursing home					
		2SLS					2SLS ($\times 10^3$)				
	OLS (1)	Linear (2)	Quadratic (3)	Cubic (4)	LPR (5)	OLS ($\times 10^3$) (6)	Linear (7)	Quadratic (8)	Cubic (9)	LPR (10)	Obs. (11)
<u>Preventive care</u>											
Support level 1 - 2	0.882***	-0.060	0.261	0.463	-0.192	0.014***	-0.002	-0.008	0.020	0.015	4,879
	(0.053)	(0.244)	(0.420)	(0.719)	(0.823)	(0.003)	(0.010)	(0.017)	(0.030)	(0.036)	
<u>Regular care</u>											
Care level 1 - 2	0.877***	0.509	-0.052	0.321	-0.894	0.012***	0.016	0.003	-0.022	-0.086	5,515
	(0.017)	(0.342)	(1.590)	(0.533)	(2.054)	(0.001)	(0.018)	(0.069)	(0.031)	(0.110)	
Care level 2 - 3	0.816***	0.550*	1.052***	0.773***	0.840***	0.011***	-0.001	0.005	0.002	0.005	3,140
	(0.019)	(0.294)	(0.229)	(0.214)	(0.320)	(0.001)	(0.017)	(0.013)	(0.012)	(0.018)	

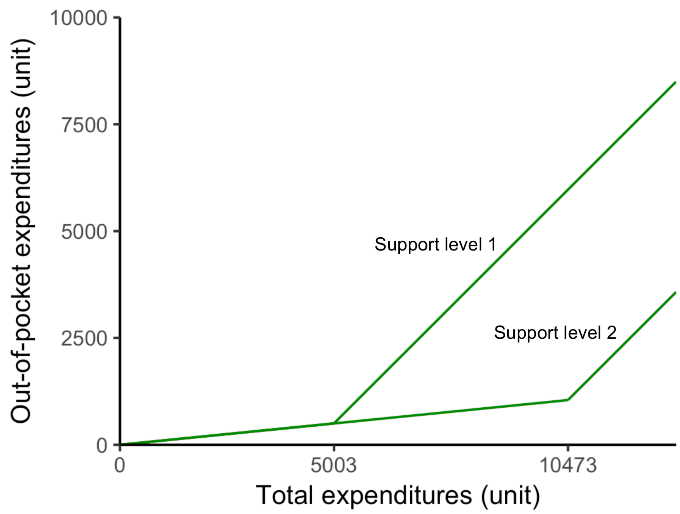
Notes: This table presents the estimates of β^c in equation 3.5 using utilization or whether to enter a nursing home as a health outcome. The first to fifth columns represent the estimates for the case of using long-term care utilization as a health outcome. The sixth to tenth columns represent the case of whether to enter a nursing home. The first and sixth columns represent the OLS estimates and other columns represent the 2SLS estimates. The 2SLS estimates are separately presented for different specifications of $f(Caretime_{it})$: linear, quadratic, cubic and LPR representing nonparametric local polynomial regression estimates using the robust confidence intervals proposed by Calonico, Cattaneo, and Titiunik (2014). In the case of OLS, $f(Caretime_{it})$ is linear. Standard errors in parentheses under each estimate. *** p<0.01, ** p<0.05, * p<0.10.

Appendix for Chapter 3

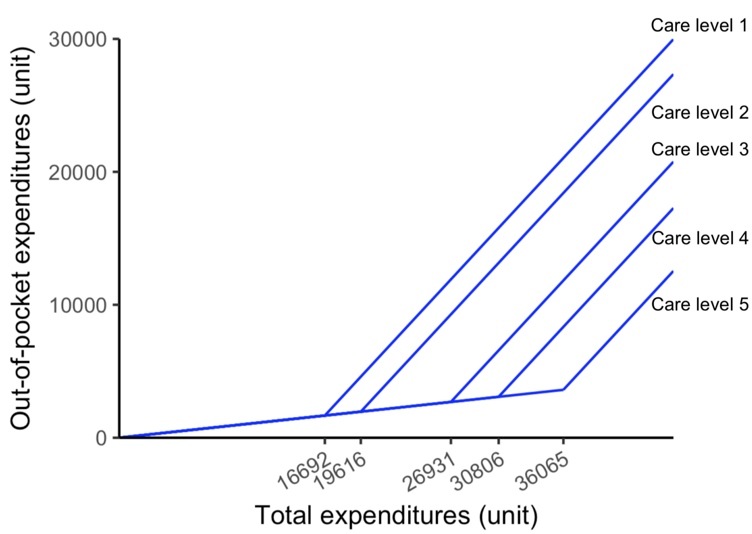
Appendix 3A: Figures and Tables

Figure 3A.1: Out-of-Pocket Expenditures as a Function of Total Expenditures

(a) Support level

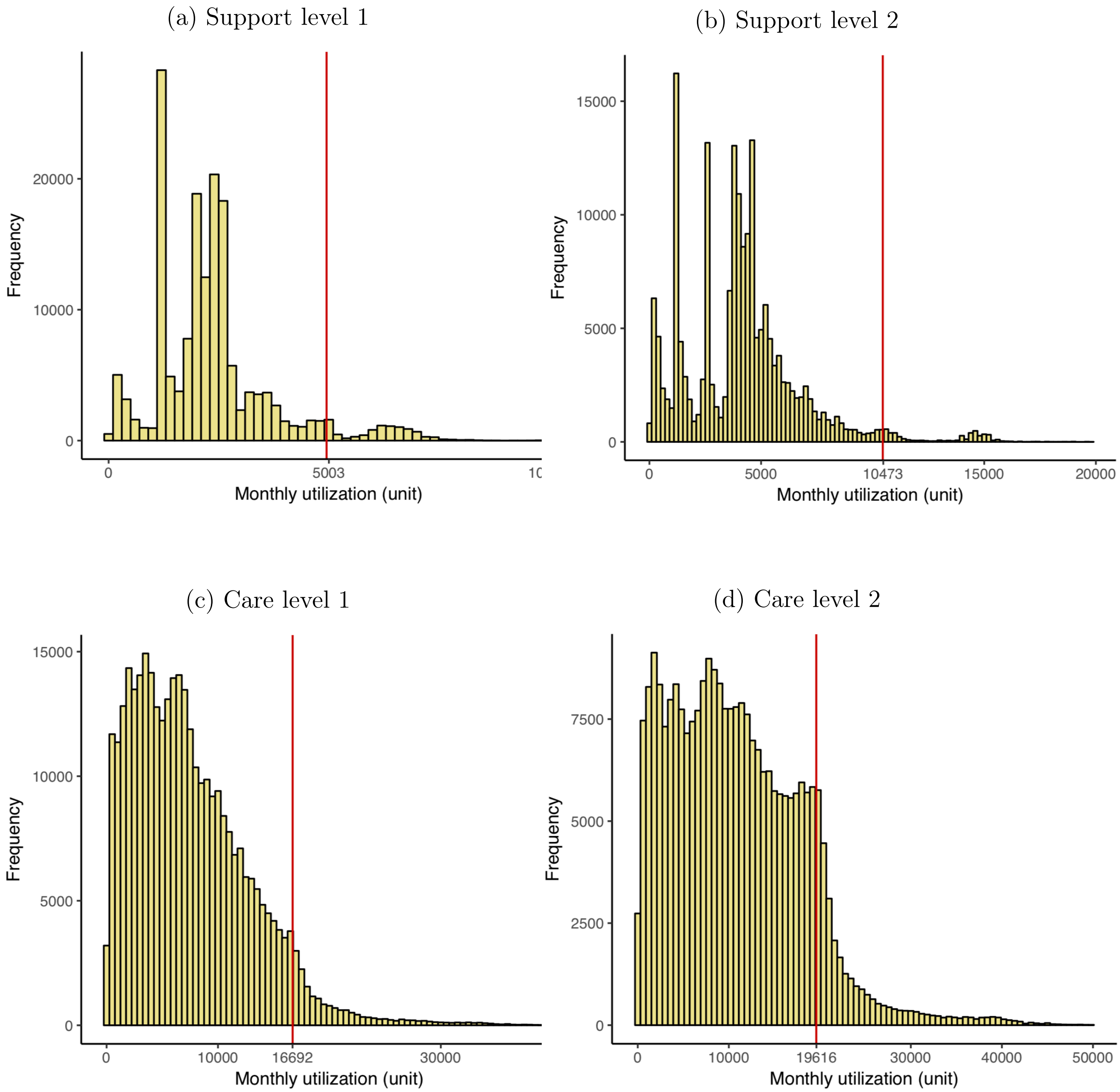


(b) Care level

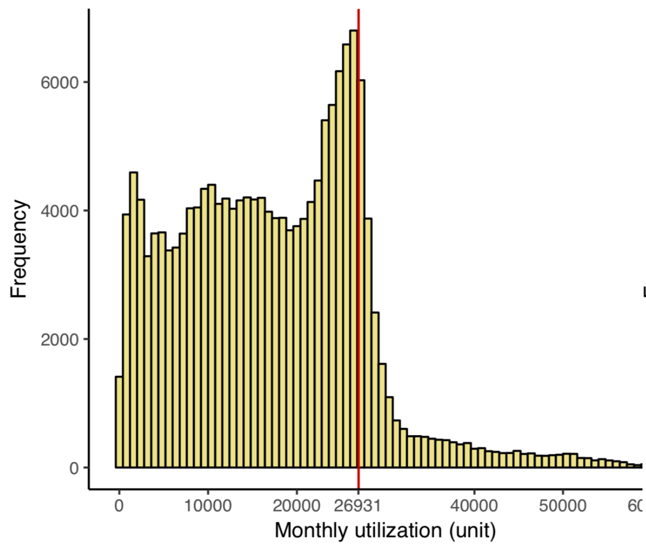


Notes: This figure presents monthly out-of-pocket expenditures as a function of monthly total expenditures in the case where the coinsurance rate is 10 percent. One unit of long-term care services is approximately 10 JPY (0.1 USD).

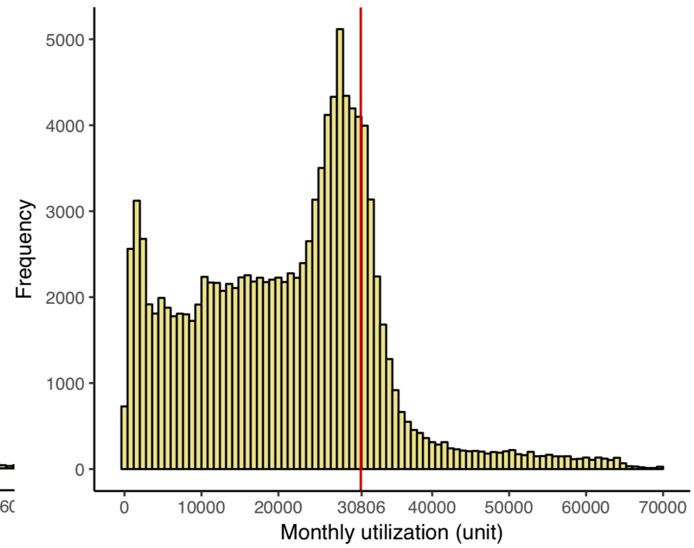
Figure 3A.2: Distribution of Monthly Long-Term Care Utilization



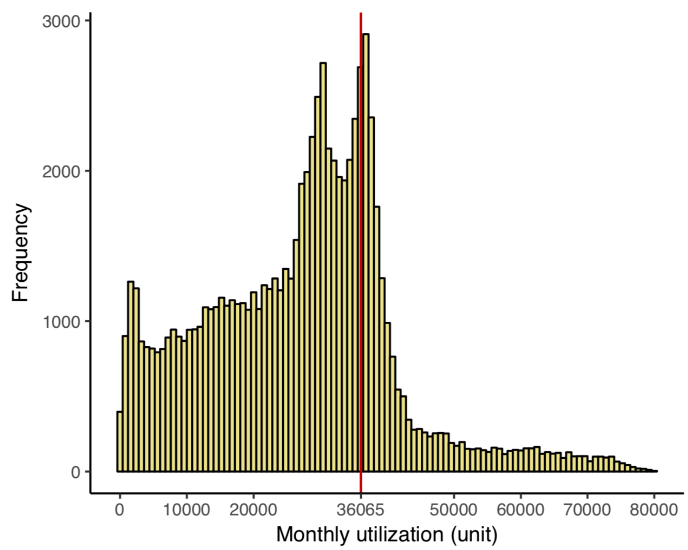
(e) Care level 3



(f) Care level 4

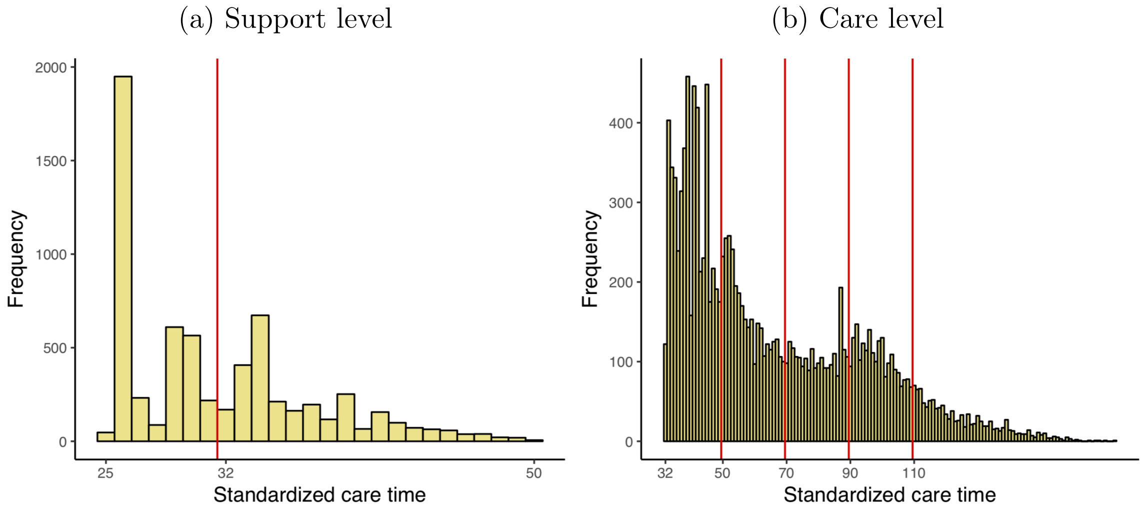


(g) Care level 5



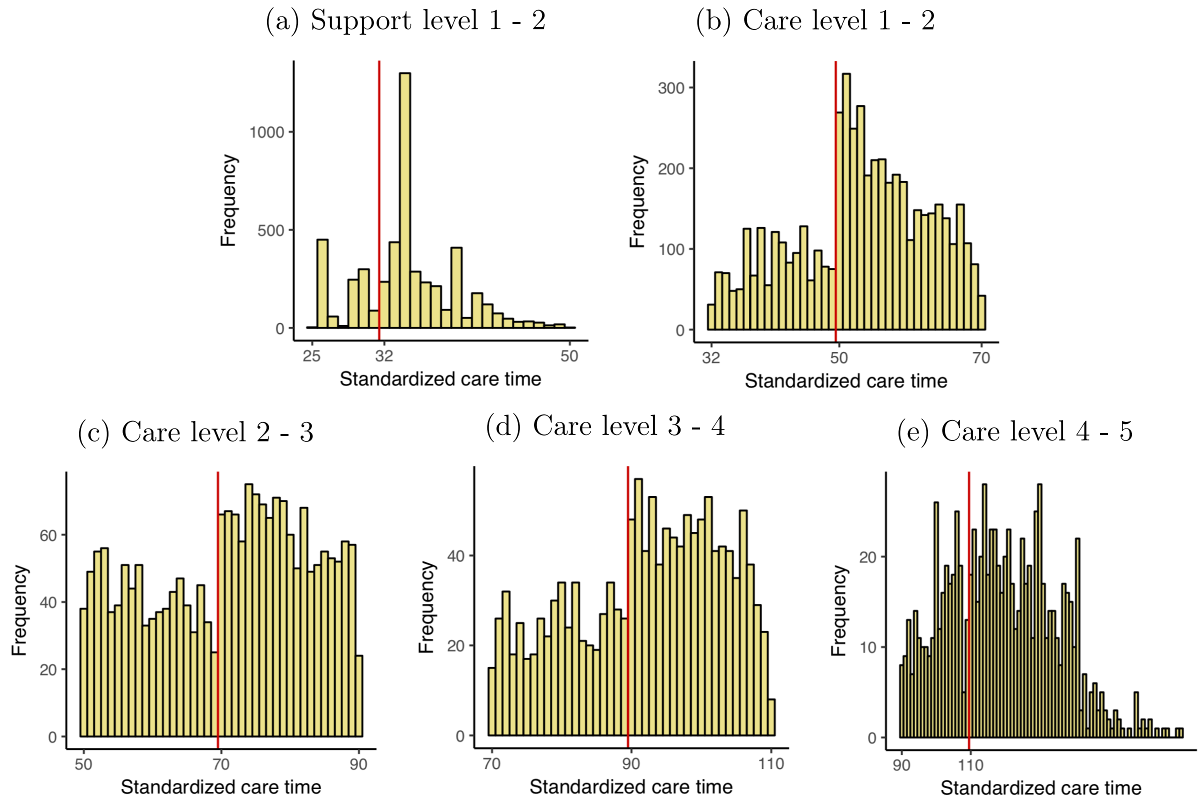
Notes: This figure shows the distribution of monthly long-term care utilization for each care-needs level. The red vertical lines indicate the monthly coverage limits.

Figure 3A.3: Distribution of Standardized Care Time (First Certification)



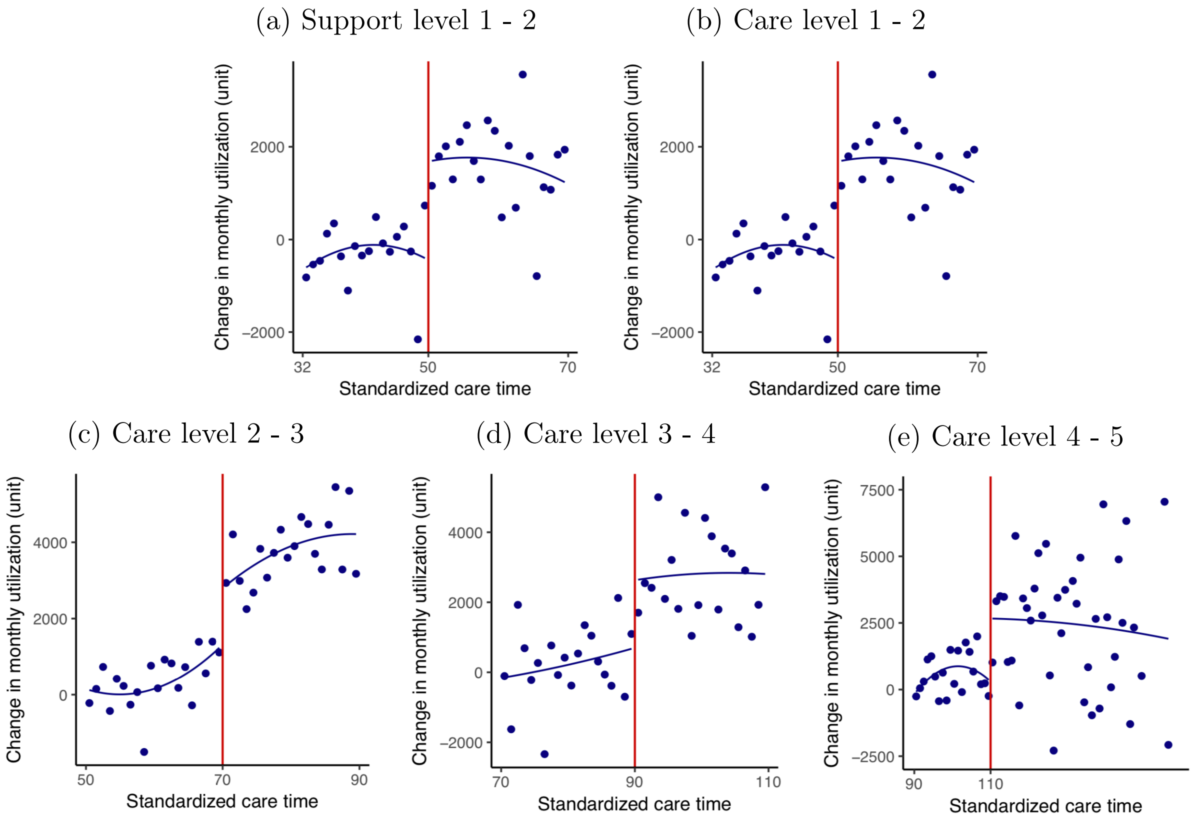
Notes: This figure presents the distribution of standardized care time for Support level (Panel (a)) and Care level (Panel (b)). The distribution is constructed using recipients who take their first care-needs certification. The bin width is 1 minute of standardized care time.

Figure 3A.4: Distribution of Standardized Care Time (Coverage Reduction, Low-Demand)



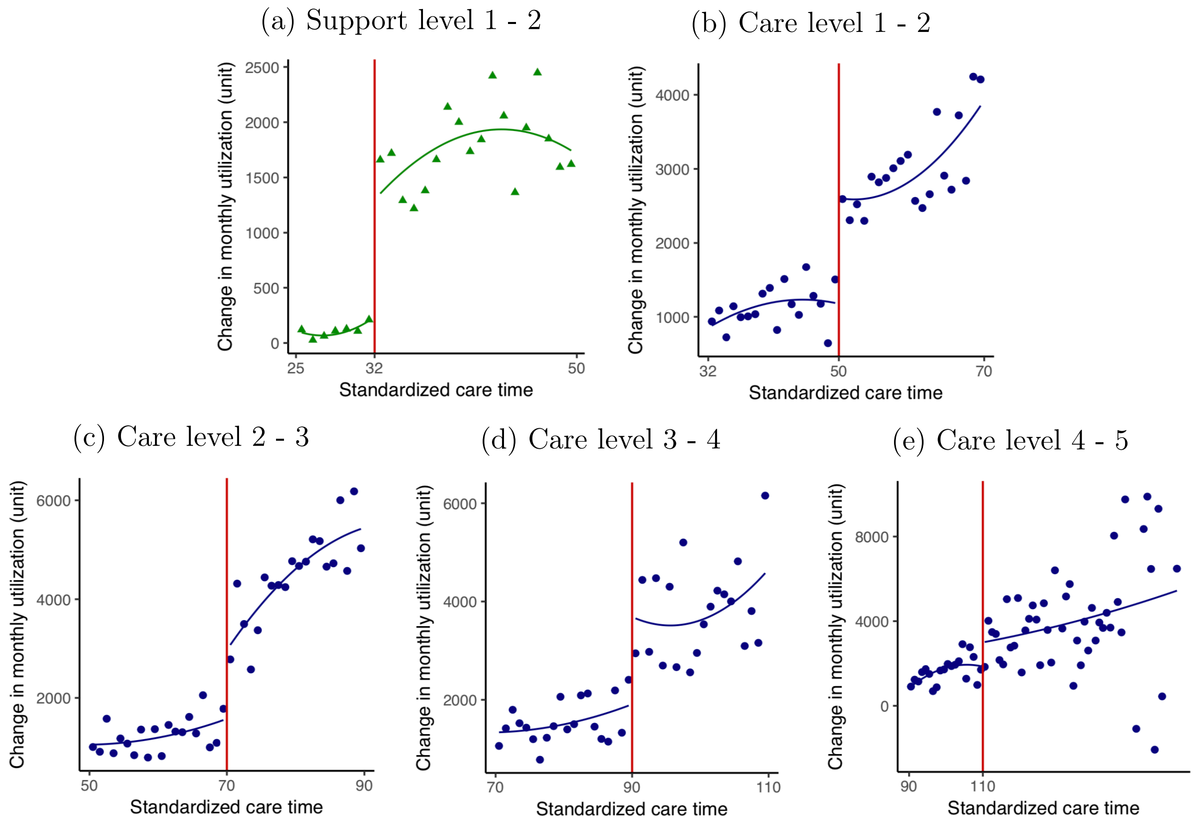
Notes: This figure presents the distribution of standardized care time for each care-needs level. The distribution is constructed using the (modified) low-demand recipients after the reclassification of care-needs levels (either the same or one-stage lower level). The bin width is 1 minute of standardized care time.

Figure 3A.5: Changes in Long-Term Care Utilization (High-demand)



Notes: This figure presents the effect of expanding insurance coverage on long-term care utilization using high-demand recipients who end up being either the same or one stage worse after the reclassification. I divide standardized care time into 1 minute-wide interval (bins) and each plot represents a local average of changes of monthly total units of long-term care utilization within bins. I plot the local average for both the Support level (Panel (a)) and the Care level (Panel (b)-(e)) and a quadratic prediction is fitted on the plots for each care-needs level.

Figure 3A.6: Changes in Long-Term Care Utilization (All Recipients)



Notes: This figure presents the effect of expanding insurance coverage on long-term care utilization using all recipients who end up being either the same or one stage worse after the reclassification. I divide standardized care time into 1 minute-wide interval (bins) and each plot represents a local average of changes of monthly total units of long-term care utilization within bins. I plot the local average for both the Support level (Panel (a)) and the Care level (Panel (b)-(e)) and a quadratic prediction is fitted on the plots for each care-needs level.

Table 3A.1: McCrary Test

	All recipients		Low-demand		High-demand	
	Est. (1)	SE (2)	Est. (3)	SE (4)	Est. (5)	SE (6)
A. Baseline sample						
Support level 1 - 2	1.374***	(0.030)				
Care level 1 - 2	0.462***	(0.019)				
Care level 2 - 3	0.314***	(0.027)				
Care level 3 - 4	0.018	(0.025)				
Care level 4 - 5	-0.055	(0.034)				
B. First certification						
Support level 1 - 2	0.863***	(0.067)				
Care level 1 - 2	0.366***	(0.049)				
Care level 2 - 3	0.166**	(0.069)				
Care level 3 - 4	0.060	(0.062)				
Care level 4 - 5	-0.058	(0.085)				
C. Coverage expansion						
Support level 1 - 2	1.232***	(0.071)	1.398***	(0.091)	1.737***	(0.296)
Care level 1 - 2	0.035	(0.036)	0.036	(0.047)	0.460***	(0.140)
Care level 2 - 3	-0.098*	(0.056)	-0.043	(0.090)	0.383**	(0.153)
Care level 3 - 4	-0.293***	(0.060)	-0.328***	(0.116)	-0.352**	(0.145)
Care level 4 - 5	-0.220***	(0.082)	-0.508***	(0.176)	0.065	(0.138)
D. Coverage reduction						
Support level 1 - 2	1.682***	(0.051)	1.488***	(0.071)	1.859***	(0.073)
Care level 1 - 2	1.474***	(0.056)	1.489***	(0.072)	1.919***	(0.122)
Care level 2 - 3	0.891***	(0.071)	0.757***	(0.102)	1.012***	(0.099)
Care level 3 - 4	0.695***	(0.073)	0.612***	(0.115)	0.750***	(0.094)
Care level 4 - 5	0.413***	(0.091)	0.267*	(0.160)	0.480***	(0.110)

Notes: This table presents the results of the McCrary test. Panel A and B shows the results for the distribution of standardized care time using all recipients-term and first care-needs certification respectively. Panel C and D show the results for the distribution after a reclassification of care-needs levels because this is the timing of estimation for behavioral effects. The first and second columns show the results for entire recipients. The rest of the columns show the estimates for low-demand and high-demand recipients separately. Low-demand recipients are those whose average monthly long-term care utilization during the previous certification term is less than 50% of a given coverage limit. High-demand recipients are those whose average monthly long-term care utilization during the previous certification term is higher than 80% of a given coverage limit. *** p<0.01, ** p<0.05, * p<0.10.

Table 3A.2: Covariates Balance Tests (All Baseline Samples)

	Linear		Quadratic			
	Coef. (1)	SE (2)	Coef. (3)	SE (4)	Obs. (5)	Cluster (6)
Support level 1 - 2						
Age	-0.234	(0.193)	1.007***	(0.356)		
Female	0.023	(0.011)	-0.070***	(0.023)	38,025	15,646
20% coinsurance	0.001	(0.006)	0.042***	(0.011)		
Care level 1 - 2						
Age	-0.411	(0.163)	-0.335	(0.239)		
Female	-0.031***	(0.008)	-0.002	(0.013)	51,367	28,607
20% coinsurance	0.001	(0.005)	-0.003	(0.007)		
Care level 2 - 3						
Age	-0.078	(0.220)	-0.346	(0.337)		
Female	0.019	(0.012)	0.011	(0.017)	36,613	22,848
20% coinsurance	-0.005	(0.006)	-0.019**	(0.009)		
Care level 3 - 4						
Age	-0.667***	(0.232)	-0.675**	(0.336)		
Female	0.014	(0.012)	0.021	(0.017)	27,469	18,801
20% coinsurance	-0.006	(0.006)	-0.005	(0.009)		
Care level 4 - 5						
Age	-0.381	(0.277)	-0.152	(0.389)		
Female	0.001	(0.013)	-0.001	(0.019)	20,689	14,371
20% coinsurance	-0.005	(0.007)	0.004	(0.010)		

Notes: This table shows the estimates of covariate balance tests for the linear and quadratic RD specification from equation 3.7, using all recipients in the baseline sample. *** p<0.01, ** p<0.05, * p<0.10.

Table 3A.3: Covariates Balance Tests (First Certification)

	Linear		Quadratic		
	Coef. (1)	SE (2)	Coef. (3)	SE (4)	Obs. (5)
Support level 1 - 2					
Age	0.239	(0.368)	1.629**	(0.693)	
Female	0.037	(0.025)	-0.037	(0.048)	6,535
20% coinsurance	0.020	(0.013)	0.050**	(0.025)	
Care level 1 - 2					
Age	-1.857***	(0.391)	-1.973***	(0.586)	
Female	-0.039	(0.022)	0.036	(0.033)	8,465
20% coinsurance	0.013	(0.013)	-0.029	(0.020)	
Care level 2 - 3					
Age	0.187	(0.568)	0.748	(0.862)	
Female	-0.022	(0.029)	-0.051	(0.044)	5,312
20% coinsurance	-0.008	(0.016)	-0.011	(0.023)	
Care level 3 - 4					
Age	-0.465	(0.584)	-0.090	(0.870)	
Female	-0.036	(0.030)	-0.035	(0.044)	4,201
20% coinsurance	0.013	(0.017)	0.011	(0.025)	
Care level 4 - 5					
Age	-0.823	(0.691)	-1.363	(1.006)	
Female	-0.065*	(0.034)	-0.043	(0.049)	3,199
20% coinsurance	0.024	(0.019)	0.033	(0.027)	

Notes: This table shows the estimates of covariate balance tests for the linear and quadratic RD specification from equation 3.7, using recipients who take their first care-needs certification. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

Table 3A.4: Covariates Balance Tests (Coverage Reduction)

	All recipients				Low-demand recipients				High-demand recipients			
	Linear		Quadratic		Linear		Quadratic		Linear		Quadratic	
	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
<u>Support level 1 - 2</u>												
Age	-0.846**	(0.379)	0.874	(0.732)	-0.295	(0.530)	1.466	(0.926)	-1.494***	(0.526)	0.247	(1.021)
Female	-0.018	(0.022)	-0.117***	(0.045)	-0.029	(0.029)	-0.152***	(0.060)	0.001	(0.034)	-0.077	(0.066)
20% coinsurance	0.021*	(0.011)	0.033	(0.022)	0.023	(0.016)	0.023	(0.033)	0.019	(0.015)	0.045	(0.029)
Observation	10,846				4,951				5,895			
Cluster	5,576				3,063				3,181			
<u>Care level 1 - 2</u>												
Age	0.138	(0.474)	0.188	(0.736)	-0.409	(0.524)	-0.039	(0.798)	1.584	(1.134)	0.854	(1.879)
Female	-0.003	(0.026)	0.001	(0.040)	-0.018	(0.029)	-0.050	(0.044)	0.035	(0.058)	0.206**	(0.093)
20% coinsurance	0.004	(0.014)	0.009	(0.021)	-0.001	(0.016)	0.018	(0.024)	0.022	(0.031)	-0.034	(0.042)
Observation	7,248				5,100				2,148			
Cluster	5,816				4,194				1,779			
<u>Care level 2 - 3</u>												
Age	0.065	(0.674)	-0.029	(1.043)	-1.037	(0.941)	-1.815	(1.486)	1.130	(0.975)	1.710	(1.449)
Female	0.055*	(0.032)	0.093*	(0.048)	0.047	(0.046)	0.080	(0.070)	0.043	(0.045)	0.090	(0.067)
20% coinsurance	-0.010	(0.018)	-0.039	(0.026)	-0.027	(0.027)	-0.039	(0.043)	0.017	(0.022)	-0.030	(0.031)
Observation	4,697				2,083				2,614			
Cluster	3,759				1,766				2,114			
<u>Care level 3 - 4</u>												
Age	-0.572	(0.541)	-0.804	(0.796)	2.078**	(0.923)	1.507	(1.344)	-1.970***	(0.663)	-1.883*	(0.981)
Female	0.019	(0.028)	-0.004	(0.040)	0.013	(0.049)	0.020	(0.071)	0.022	(0.034)	-0.018	(0.049)
20% coinsurance	-0.005	(0.014)	-0.008	(0.020)	0.046*	(0.025)	-0.032	(0.017)	-0.041**	(0.017)	-0.012	(0.024)
Observation	4,813				1,790				3,023			
Cluster	3,867				1,499				2,479			
<u>Care level 4 - 5</u>												
Age	-0.654	(0.862)	1.326	(1.242)	-2.531	(1.692)	-0.282	(2.583)	0.135	(0.969)	1.937	(1.340)
Female	0.025	(0.039)	0.089	(0.056)	-0.034	(0.067)	-0.110	(0.097)	0.052	(0.048)	0.190***	(0.069)
20% coinsurance	-0.014	(0.019)	-0.021	(0.030)	-0.017	(0.032)	0.025	(0.044)	-0.012	(0.024)	-0.045	(0.040)
Observation	2,701				897				1,804			
Cluster	2,013				706				1,370			

Notes: This table shows the estimates of covariate balance tests for the linear and quadratic RD specification from equation 3.7. These estimation use recipients who are used for analyzing coverage reduction after the reclassification of care-needs levels. Low-demand recipients are those whose average monthly long-term care utilization during a previous certification term is less than 50% of a coverage limit of the one-stage lower care-needs level. High-demand recipients are those whose average monthly long-term care utilization during a previous certification term is higher than 80% of a coverage limit of the one-stage lower care-needs level. *** p<0.01, ** p<0.05, * p<0.10.

Table 3A.5: First-Stage Estimates for Coverage Expansion

	Linear		Quadratic		Cubic		LPR			
	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE	Obs.	Cluster
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Support level 1 - 2										
All recipients	5,098.4***	(43.0)	4,931.8***	(108.2)	4,854.9***	(220.1)	4,602.6***	(409.8)	110,120	5,631
Low-demand	5,107.0***	(55.6)	4,915.1***	(42.8)	4,833.9***	(306.6)	4,415.9***	(327.9)	71,747	4,039
High-demand	4,939.3***	(159.1)	5,018.9***	(292.1)	5,407.0***	(311.5)	5,730.6***	(199.7)	8,548	538
Care level 1 - 2										
All recipients	2,869.4***	(17.3)	2,901.9***	(20.2)	2,858.8***	(24.7)	2,842.9***	(29.7)	192,276	10,454
Low-demand	2,856.4***	(23.2)	2,878.9***	(28.9)	2,819.4***	(37.4)	2,801.4***	(45.7)	121,190	7,136
High-demand	2,797.2***	(57.7)	2,878.0***	(66.2)	2,884.2***	(57.5)	2,902.8***	(12.1)	22,236	1,354
Care level 2 - 3										
All recipients	6,375.9***	(77.3)	6,118.4***	(139.6)	5,726.9***	(212.1)	5,483.4***	(319.3)	110,245	6,439
Low-demand	6,532.6***	(123.0)	6,368.9***	(216.4)	6,121.5***	(321.0)	5,966.3***	(479.3)	50,457	2,934
High-demand	5,985.0***	(165.6)	5,747.7***	(288.7)	5,351.4***	(421.9)	5,202.5***	(611.7)	28,225	1,920
Care level 3 - 4										
All recipients	3,327.6***	(68.9)	3,285.3***	(102.4)	3,188.9***	(136.0)	3,117.7***	(199.8)	63,109	3,867
Low-demand	3,401.7***	(109.3)	3,449.2***	(160.9)	3,432.9***	(223.2)	3,232.5***	(282.7)	20,085	1,217
High-demand	3,257.9***	(108.9)	3,204.1***	(152.8)	3,056.0***	(194.2)	3,095.3***	(210.1)	25,902	1,725
Care level 4 - 5										
All recipients	4,457.5***	(91.5)	4,252.5***	(145.8)	4,176.3***	(202.1)	4,199.7***	(258.3)	43,881	2,700
Low-demand	4,243.4***	(219.2)	4,006.3***	(343.6)	3,869.5***	(490.4)	3,428.2***	(731.7)	10,760	686
High-demand	4,434.0***	(128.9)	4,220.4***	(197.0)	4,178.9***	(262.7)	4,312.8***	(277.1)	22,016	1,433

Notes: This table presents the first-stage estimates of β_0^c in equation 3.4. The first to sixth columns represent the estimates for different specifications of $f_0(Caretime_{it})$: linear, quadratic and cubic respectively. The seventh and eighth columns show the nonparametric local polynomial regression (LPR) estimates using the robust confidence intervals proposed by Calonico, Cattaneo, and Titiunik (2014). *** p<0.01, ** p<0.05, * p<0.10.

Table 3A.6: First-Stage Estimates for the Health Effects of Long-Term Care Utilization

	Standardized care time				Utilization / Nursing home					
	Linear	Quadratic	Cubic	LPR	Linear	Quadratic	Cubic	LPR (Util)	LPR (NH)	Obs.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Support level 1 - 2	1,104.9***	1,134.7***	1,049.8***	760.7***	1,105.0***	1,135.1***	1,050.9***	706.6**	714.0**	4,879
	(100.4)	(184.6)	(297.5)	(275.6)	(100.4)	(184.6)	(297.6)	(291.5)	(287.3)	
Care level 1 - 2	900.6***	322.4	1,199.9**	546.1	875.4***	572.6	1,247.6**	465.5	588.5	5,515
	(264.5)	(402.7)	(552.2)	(688.2)	(263.3)	(400.8)	(549.7)	(681.6)	(672.8)	
Care level 2 - 3	1,541.8***	3,063.8***	4,180.8***	2,964.2**	1,482.5***	2,885.7***	4,089.6***	3,354.2**	3,359.9***	3,140
	(487.2)	(745.2)	(1,022.1)	(1,352.1)	(481.4)	(736.7)	(1,010.1)	(1,075.8)	(1,073.9)	
Care level 3 - 4	617.0	-626.3	-2,036.5	-2,162.2	717.0	-567.7	-1,730.8	-1,725.0	-2,167.7	2,257
	(669.4)	(985.7)	(1,338.8)	(1,395.3)	(638.8)	(940.7)	(1,278.3)	(1,316.1)	(1,462.2)	
Care level 4 - 5	256.2	1108.3	-1,500.7	-2,743.2	428.1	1,450.5	-846.0	-1,384.5	-1,364.0	1,532
	(987.1)	(1,413.8)	(1,813.0)	(2,802.8)	(900.5)	(1,289.7)	(1,654.6)	(2,146.7)	(2,204.3)	

Notes: This table presents the first-stage estimates β_0^c in equation 3.6. The first to fourth column represent the estimates for the case of using standardized care time as a health outcome. The fifth to ninth columns represent the case of utilization or whether to enter a nursing home. Estimates are separately presented for different specifications of $f(Caretime_{it})$: linear, quadratic, cubic and LPR representing nonparametric local polynomial regression estimates using the robust confidence intervals proposed by Calonico, Cattaneo, and Titiunik (2014). The eighth and ninth columns represent the estimates of LPR for the case of utilization and nursing home respectively. Standard errors in parentheses under each estimate. *** p<0.01, ** p<0.05, * p<0.10.

Table 3A.7: First-Stage Estimates for Coverage Reduction

	Linear		Quadratic		Cubic		LPR			
	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE	Obs.	Cluster
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Support level 1 - 2										
All recipients	-2,575.9***	(136.6)	-2,141.7***	(264.4)	-1,689.6***	(455.8.7)	-2,361.8***	(614.9)	109,026	5,576
Low-demand	-3,092.4***	(186.0)	-2,781.9***	(378.2)	-1,997.1***	(643.7)	-2,512.1***	(438.5)	50,115	3,063
High-demand	-2,019.0***	(194.7)	-1,507.0***	(359.6)	-1,597.4***	(621.5)	-1,593.5***	(606.7)	58,911	3,181
Care level 1 - 2										
All recipients	-2,911.0***	(28.9)	-3,093.0***	(36.7)	-2,928.2***	(43.7)	-2,983.7***	(47.1)	101,012	5,816
Low-demand	-2,938.0***	(33.2)	-3,064.8***	(41.4)	-2,915.5***	(47.8)	-2,961.1***	(50.8)	71,773	4,194
High-demand	-2,848.6***	(59.4)	-3,173.7***	(82.5)	-2,997.3***	(103.4)	-3,119.0***	(106.5)	29,248	1,779
Care level 2 - 3										
All recipients	-3,007.7***	(219.0)	-2,252.0***	(316.5)	-1,802.5***	(410.1)	-1,720.4***	(503.3)	66,433	3,759
Low-demand	-3,944.8***	(290.9)	-3,282.1***	(447.3)	-3,407.1***	(610.7)	-2,897.6***	(716.1)	30,475	1,766
High-demand	-2,002.9***	(308.9)	-1,152.1***	(402.8)	-231.7	(470.5)	-626.3	(501.3)	35,958	2,114
Care level 3 - 4										
All recipients	-1,569.1***	(130.2)	-1,359.3***	(190.0)	-1,104.4***	(245.3)	-1,134.7***	(291.1)	44,911	2,609
Low-demand	-2,029.1***	(187.9)	-1,522.3***	(275.8)	-1,423.0***	(367.1)	-1,512.1***	(474.7)	19,555	1,149
High-demand	-1,222.8***	(177.5)	-1,240.7***	(257.0)	-863.0***	(326.2)	-817.3	(376.0)	25,356	1,528
Care level 4 - 5										
All recipients	-1,232.8***	(195.2)	-1,290.4***	(277.6)	-1,134.0***	(362.5)	-390.2	(462.2)	36,574	2,013
Low-demand	-1,696.8***	(334.7)	-1,685.4***	(538.8)	-1,104.3**	(698.2)	-1,336.3	(883.5)	12,582	706
High-demand	-988.6***	(257.5)	-1,142.4***	(323.6)	-1,274.4***	(421.7)	-1,209.3***	(391.2)	23,992	1,370

Notes: This table presents the first-stage estimates of β_0^c in equation 3B.2. The first to sixth columns represent the estimates for different specifications of $f_0(Caretime_{it})$: linear, quadratic and cubic respectively. The seventh and eighth columns show the nonparametric local polynomial regression (LPR) estimates using the robust confidence intervals proposed by Calonico, Cattaneo, and Titiunik (2014). *** p<0.01, ** p<0.05, * p<0.10.

Table 3A.8: The Effects of Coverage Reduction on Long-Term Care Utilization

	Linear		Quadratic		Cubic		LPR			
	Coef. (1)	SE (2)	Coef. (3)	SE (4)	Coef. (5)	SE (6)	Coef. (7)	SE (8)	Obs. (9)	Cluster (10)
<u>Support level 1 - 2</u>										
All recipients	0.228***	(0.033)	0.175***	(0.086)	0.510***	(0.199)	0.270	(0.213)	109,026	5,574
Low-demand	0.066***	(0.026)	0.018	(0.059)	-0.138	(0.161)	0.007	(0.080)	50,115	3,063
High-demand	0.528***	(0.055)	0.520***	(0.154)	0.997***	(0.302)	0.861	(0.241)	58,911	3,181
<u>Care level 1 - 2</u>										
All recipients	0.192***	(0.063)	0.129	(0.094)	0.119	(0.134)	0.105	(0.157)	101,012	5,816
Low-demand	0.208***	(0.060)	0.127	(0.088)	-0.006	(0.119)	0.020	(0.133)	71,773	4,194
High-demand	0.284	(0.205)	0.394	(0.322)	0.950***	(0.471)	0.806	(0.539)	29,248	1,779
<u>Care level 2 - 3</u>										
All recipients	0.142	(0.109)	0.230	(0.219)	0.402	(0.360)	0.732*	(0.506)	66,433	3,759
Low-demand	0.057	(0.093)	0.157	(0.161)	0.119	(0.204)	0.241	(0.240)	30,475	1,766
High-demand	0.404	(0.273)	0.692	(0.713)	5.204	(9.947)	3.875	(3.294)	35,958	2,114
<u>Care level 3 - 4</u>										
All recipients	0.381	(0.240)	0.240	(0.483)	0.188	(0.818)	-0.089	(0.945)	44,911	2,609
Low-demand	0.174	(0.256)	0.368	(0.482)	0.121	(0.671)	0.186	(0.695)	19,555	1,149
High-demand	0.733	(0.502)	0.326	(0.813)	0.309	(1.632)	-0.747	(2.048)	25,356	1,528
<u>Care level 4 - 5</u>										
All recipients	-0.779*	(0.469)	-0.584	(0.647)	-1.450	(1.010)	-4.903	(3.182)	36,574	2,013
Low-demand	-0.941	(0.502)	-1.396*	(0.825)	-2.447	(1.957)	-2.735	(2.052)	12,582	706
High-demand	-0.648	(0.741)	-0.135	(0.848)	-1.068	(1.013)	-0.329	(1.030)	23,992	1,370

Notes: This table presents the second-stage estimates of β^c in equation 3B.1. The first to sixth columns represent the estimates for different specifications of $f(Caretime_{it})$: linear, quadratic and cubic respectively. The seventh and eighth columns show the nonparametric local polynomial regression (LPR) estimates using the robust confidence intervals proposed by Calonico, Cattaneo, and Titiunik (2014). Low-demand recipients are those whose average monthly long-term care utilization during the previous certification term is less than 50% of a given coverage limit. High-demand recipients are those whose average monthly long-term care utilization during the previous certification term is more than 80% of a given coverage limit. *** p<0.01, ** p<0.05, * p<0.10.

Table 3A.9: The Impact of Counterfactual Policy on Long-Term Care Costs

High-Demand		Low-Demand	
Lower (1)	Upper (2)	Lower (3)	Upper (4)
\$0.34 million (2.06%)	\$0.72 million (4.37%)	\$1.13 million (6.80%)	\$1.80 million (10.86%)

Notes: This table presents the expected increase in long-term care costs due to the counterfactual policy which expands all recipients' insurance coverage by one-stage (except Support level 2 and Care level 5). The values in parenthesis represent the percentage rises of the total cost of home-based long-term care through price or behavioral effects. The upper and lower bound are calculated by using the largest and smallest estimates in Table 3.6 respectively, excluding those with a negative value. The approximation of 1 unit = 0.1 USD is assumed for the calculation.

Appendix 3B: Estimation for a Coverage Reduction

When recipients face a coverage reduction, they are influenced by price effects if their service utilization is higher than the *one-stage lower* coverage limit. Therefore, the first step focuses on recipients whose average monthly utilization during a given term s is lower than 80% of the one-stage lower coverage limit. As explained in the summary statistics section, 80% is set here as the criterion rather than 50% in order to ensure sufficient statistical power due to the relatively low number of LTCI recipients whose coverage falls over time. This set of modified low-demand recipients is further narrowed in the second step to those whose standardized care time in the following term $s+1$ is in the same or one stage *lower* care-needs level compared with that of term s. The causal relationship of interest is the same as equation 3.3:

$$\begin{aligned}\Delta Utilization_{it}^{s,s+1} = \alpha^c + \beta^c \Delta Coverage_i^{s,s+1} + f^c(Caretime_{is+1}) \\ + \gamma^c Caretime_{is} + X_{it}^{s,s+1}\eta^c + \varepsilon_{it},\end{aligned} \tag{3B.1}$$

The first-stage regression is:

$$\begin{aligned}\Delta Coverage_i^{s,s+1} = \alpha_0^c + \beta_0^c \mathbb{1}\{Caretime_{is+1} < Cutoff\} + f_0^c(Caretime_{is+1}) \\ + \gamma_0^c Caretime_{is} + X_{it}^{s,s+1}\eta_0^c + \varepsilon_{it}.\end{aligned} \tag{3B.2}$$

where $\mathbb{1}\{Caretime_{is+1} < Cutoff\}$ is a dummy variable that takes value 1 if standardized care time is less than a given cutoff value. Other notations and functional specifications are the same as the estimation for coverage expansion. Thus β^c in equation 3B.1 indicates behavioral effects of coverage reduction on long-term care utilization. As in the case of coverage expansion, I estimate the regression with high-demand recipients and with entire recipients to make comparisons between responses due to behavioral biases and economic incentives.

Chapter 4

The Impact of Manipulation of Social Insurance Benefits

4.1 Introduction

Many social welfare programs rely on the means test to target individuals who need government assistance. For example, Medicaid and other welfare programs in the United States use income, asset, and non-financial criteria to determine whether a particular individual is eligible for the program. These means test results are high-stakes for disadvantaged individuals because their lives could strongly be affected by welfare programs. The reliance on high-stakes means test leads to concerns about the manipulation of test results. If there is room for discretion in the means test, the government might manipulate test results and discretionary make certain individuals eligible for the program for political gain or other reasons. The discretionary manipulation of the means test is not only unfair but also affect policy effectiveness of welfare programs; the individuals judged eligible for welfare programs under manipulation should be different from the ones expected by the policy rules. To properly measure the impact of welfare programs, it is important to understand the extent to which a given policy scheme induces manipulation of the means test and how it affects policy effectiveness.

In this paper, we examine the impact of manipulation of social insurance benefits in the context of the public long-term care insurance system (LTCI) in Japan. LTCI is a means-tested social insurance program that is designed to provide disabled elderly with subsidies for long-term care (LTC). A means test in LTCI is based on an individual's health status, the degree to which the person needs for LTC, rather than economic status. Applicants for LTCI subsidy take health checkups and have health index called standardized care time (SCT) calculated. In determining eligibility and generosity of LTCI benefits, SCT plays a key role. Applicants whose SCT is higher than a certain threshold are eligible for LTCI, and eligible applicants are classified into one of seven categories (care-needs levels) divided by other thresholds that determine LTCI benefits. The generosity of LTCI benefits is mainly characterized by a monthly coverage limit which sets the monthly maximum spending on home-based LTC services that can be covered by insurance. Due to the coverage limit, recipients' LTC utilization is strongly affected by their care-needs level. Importantly, recipients have to regularly retake the health checkups and they are classified into particular care-needs level each time based on newly calculated SCT.

We first document large scale manipulation of SCT by showing clear discontinuity and bunching in the distribution of SCT conditional on the previous care-needs level. Specifically, the conditional distributions suggest that LTC investigators manipulate SCT to avoid assigning recipients into a lower care-needs level than the previous term. Figure 4.1 presents one of the examples of conditional distributions which indicate the existence of manipulation. It is a distribution of SCT conditional on recipients whose previous SCT was between the indicated area. If SCT is lower than the threshold indicated by the red line, the care-needs level (coverage limit) of these recipients will be lowered from the previous level. The distribution has a discontinuity just at the threshold, and the number of recipients who below the threshold is much smaller than those above it. The distributions conditional on other care-needs levels exhibit the same property. This suggests that recipients who would otherwise have fallen into lower care-needs levels are classified into higher levels by manipulation of SCT.

We then quantify the extent to which the manipulation of SCT increases aggregate LTC expenditure. To do this, we need to recover the counterfactual distribution of SCT without manipulation and compare observed and counterfactual aggregate expenditure. In the previous bunching literature, a common method of constructing counterfactual distribution is to fit a particular polynomial to the observed distribution, excluding manipulable range around the threshold, and extrapolate the fitted distribution to the manipulable range. The main problem of this approach is that when the manipulable range is wide, we have to extensively extrapolate the fitted distribution based on a narrow range of observed distribution. Both extensive extrapolation and narrow range of parametric fitting undermine the validity of counterfactual distribution. This problem is especially common when the threshold provides discrete changes in the level of choice set ("notches") and induce strong behavioral responses. As shown in Figure 4.1, manipulation of SCT covers a wide range and the same problem applies.

To address this issue, we develop a novel methodology to partially identify and estimate the counterfactual distribution of SCT without manipulation. Our method imposes general shape restrictions such as log-concavity and stochastic dominance on counterfactual distribution without assuming any particular functional forms. Under the set of distribution satisfying the shape restrictions, we can identify the upper and lower bound of expected LTC expenditure based on the counterfactual distribution of SCT. We also propose a consistent estimator for these bound. By comparing the observed expenditure and counterfactual ones, we can consistently estimate the bound of the impact of manipulation on LTC expenditure.

The estimation for the manipulation effects shows that the manipulation of SCT significantly increases LTC expenditure even with the most conservative estimates. The lower bound of the estimate indicates that the manipulation increases annual LTC expenditure by 629 USD per recipient, which is equivalent to a 3 percent increase. The bound estimates show that the upper bound is much higher than the lower bound: the upper bound indicates that the increase in LTC expenditure due to the manipulation amounts to 2,393 USD. Therefore, the manipulation effects could be four times larger than the lower bound. The difference between upper and lower bound suggests that it is important to partially identify counterfactual distribution for SCT rather than point identify with restrict assumptions. Using the narrower manipulable range make the bound estimate narrow, but the lower bound is robust to the change in manipulable range.

Empirically, this paper contributes to the literature that analyzes the discretionary decision-making in public policies. Camacho and Conover (2011) and Brollo, Kaufmann, and La Ferrara (2019) examine the causes and consequences of discretionary enforcement of welfare program in Colombia and Brazil, respectively. They demonstrate that, in developing countries, politicians strategically manipulate the enforcement of welfare programs for political interests. Outside welfare programs, the manipulation of test scores by school teachers is also a well-studied topic. Diamond and Persson (2017) and Dee, Dobbie, Jacob and Rockoff (2019) use bunching method to detect score manipulation and quantify the impact of manipulation on children's future outcomes. In contrast, few studies analyze the manipulation of means-tested welfare programs in developed countries despite its importance. We show that the means-test of LTCI in Japan is extensively manipulated not to lower recipients' LTCI benefits, despite little political or financial benefits to those who conduct the means test. This result suggests that the manipulation of means test could be prevalent even in developed countries.

Our study also related to the growing literature on the bunching approach. In the past decade, the bunching approach developed by Saez (2010), Chetty, Friedman, Olsen, and Pistaferri (2011), and Kleven and Waseem (2013) has been applied to various topics. The key idea of this approach is to exploit bunching around points at which incentives change discontinuously, such as "kinks" and "notches", to observe behavioral responses and estimate the structural parameter. For example, Saez (2010) and Chetty et al. (2011) use bunching at kink points of taxable income to estimate taxable income elasticity. While the bunching approach has gained much popularity in many fields, some recent studies pointed out its un-identifiability of the parameter of interest. Blomquist, Newey, Kumar, and Liang (2019) and Bertanha, McCallum, and Seegert (2019) independently demonstrate that the elasticity is not point-identified using kinks when individual heterogeneity is unrestricted. The un-identifiability is attributed to the fact that the counterfactual distribution without bunching is not observable. When individual heterogeneity is flexible and counterfactual distribution is not point identified, the amount of bunching and parameter of interest are not point identified, either.

The advantage of our method is that we can partially identify counterfactual distribution with flexible shape restrictions tailored to various applications. The most relevant study in this respect is Bertanha et al. (2019), which proposed a partial identification strategy that bound the slope of counterfactual probability density function (PDF) by imposing Lipchitz continuity. Although we use log-concavity and stochastic dominance as shape restrictions for counterfactual PDF, we can impose Lipchitz continuity or other conditions instead if needed. This flexibility indicates that our method can apply to a variety of bunching analyses. In practice, the reasonable restrictions on counterfactual distribution varies among situations. Our method allows researchers to impose shape restrictions flexibly according to their bunching analyses.

This paper proceeds as follows. Section 4.2 introduces institutional background of LTCI in Japan. Section 4.3 documents a bunching in the distribution of SCT which indicates the manipulation of LTCI benefits. Section 4.4 describes the administrative data and its summary statistics. Section 4.5 proposes partial identification strategy using shape restrictions and nonparametric estimation method. Section 4.6 shows estimation results. Section 4.7 concludes.

4.2 Institutional Background

The long-term care insurance system in Japan (LTCI) was launched in April 2000 to address growing needs for public long-term care services due to the rapid aging of the population. The government subsidizes various long-term care (LTC) services through the insurance so that eligible recipients can choose and use necessary services with a moderate out-of-pocket payment. The LTCI is a mandatory social insurance system and a person need to be aged 65 or older to receive the LTCI benefits.

The generosity of the LTCI benefits depend on how much each recipient needs for LTC. That is, the more LTC a recipient needs, the more generous benefits she can receive. The purpose of such a system is to prevent relatively healthy people from overusing services while ensuring that sufficient services are available to those who need them. In the case of home-based care, the generosity of the LTCI benefits is mostly characterized by that of monthly coverage limits. Recipients can use LTC services with 10 or 20 percent of prices up until the coverage limit, after which they have to pay full price of the services. Due to the salient price difference between within and outside of coverage limit, LTC utilization is strongly influenced by the generosity of coverage limits (Takahashi, 2019).

In determining the generosity of the LTCI benefits, the health index called *standardized care time* (SCT) plays key role. SCT summarizes how much a recipient needs for LTC. It is calculated for each recipient through face-to-face health checkups, and it is used to classify recipients into specific category which has corresponding benefits. Therefore, SCT determines the generosity of the LTCI benefits, and consequently affects recipients' LTC utilization. The following describes the generosity of the LTCI benefits, how SCT is calculated, and how people use LTC services under the LTCI.

Generosity of the LTCI Benefits

In the LTCI, the generosity of insurance benefits depend on the degree of necessity of LTC of each recipient. Specifically, there are seven categories called *care-needs levels* which determine insurance coverage and available services. Care-needs levels consist of Support level 1 and 2, and Care level 1 - 5 in ascending order of needs for LTC. The broad categories of "Support level" and "Care level" determine LTC services available to recipients. Recipients who are classified as Care level are allowed to use a wide range of LTC services, while those classified as Support level can only use services mainly focusing on preventive care.

Each care-needs level within the broad categories determines the generosity of insurance coverage, and recipients who have a relatively severe condition are supposed to be entitled to more generous coverage. The insurance coverage of LTCI for home-based care is characterized by monthly coverage limits.[1] Recipients pay only 10 or 20 percent of full price of LTC up until the coverage limit, after which they have to pay full price of services.[2] Table 4.1 presents the coverage limit for each care-needs level expressed as a total unit value for LTC services. One unit is approximately 10 JPY, or about 0.1 USD. As Table 4.1 shows, the more care recipients need, the more generous coverage limit they can get.

Care-Needs Certification

As described above, SCT plays deterministic role in classifying recipients into specific care-

[1]In this paper, the terms "insurance coverage" and "coverage limit" are used interchangeably.

[2]The 20 percent coinsurance, introduced in 2015, applies to those with total annual income of more than 1.6 million JPY (16K USD), and that of first-insured family members is more than 3.46 million JPY (34.6K USD), or 2.8 million JPY (28K USD) for a single-person household).

needs level. It is calculated through detailed health checkups called *care-needs certification*. People who want to use LTC services under the LTCI must apply to the local government and take the health checkups.

Care-needs certification is based on a nationally-standardized face-to-face survey which is conducted by a trained investigator (the LTC investigator). The LTC investigator first checks 74 items about the applicant's physical and mental conditions related to LTC. Based on the results of the health checkup, a special formula is used to generate "hypothetical care times" for eight categories of assistance. Table 4.2 lists the possible time ranges for each category. The sum of these care times is the *SCT*, which indicates how much LTC is needed for the applicant. The longer the SCT is, the more an applicant is considered to need LTC.

Then applicants are tentatively assigned to a corresponding care-needs level based on the calculated SCT. Table 4.1 presents a range of SCT and corresponding care-needs level. After the tentative assignment, a committee consisting of physicians, nurses, and other health and social service experts (the Certification Committee) assesses whether the SCT appropriately reflects the applicant's needs for LTC. The Certification Committee reassigns the applicant to the proper care-needs level if needed. Although a reassignment is possible, care-needs level (and corresponding generosity of LTCI benefits) is mostly determined by SCT.

To accommodate changes in LTC needs, recipients are required to take care-needs certification regularly and are reclassified into a different care-needs level if necessary. In principle, the first care-needs certification is valid for a half year and the following certification is valid for one year. Recipients need to retake the care-needs certification before the term expires to continue using LTCI services. Hereafter, I call "certification term" or "term" for representing each valid term of the care-needs certification.

Long-Term Care Utilization

Recipients who use home-based care prepare a monthly usage plan ("care plan") given a coverage limit and available services determined by the SCT (care-needs level). They usually create care plan with the help of qualified specialists of LTC ("care manager"), which allows them to choose their services from a wide variety of options. Thus LTC utilization is a joint decision of a recipient, family members, and care manager.

As discussed in Takahashi (2019), the care-needs level to which recipients are assigned have a significant impact on the LTC utilization through multiple factors. First, recipients have a strong incentives to keep their utilization below the coverage limit due to the higher prices outside the coverage. If recipients are assigned to a lower care-needs level, their utilization are constrained by a lower level of coverage limit. Second, even if recipients do not face an effective price change due to coverage limit, they could respond to a change in their care-needs level. Takahashi (2019) shows that recipients who use only less than half the coverage limit significantly increase LTC utilization when their care-needs level (and coverage limit) rise. This could be explained by a behavioral hazard of decision-makers such as anchoring effects and/or heuristics. Overall, recipients' LTC utilization is strongly affected by the care-needs level regardless of their utilization level. In the later section, we will show to what extent the assignment of care-needs level affects LTC utilization.

4.3 Motivating Facts: Manipulation of the LTCI Benefits

One of our main dataset is LTCI administrative data on care-needs certification. The most important information in this dataset is each recipient's SCT, which is used for assigning recipients to a specific care-needs level. This data is available for each certification term, and contains a breakdown of how the final SCT was calculated (that is, it provides a hypothetical care time for each category of assistance) as well as other information related to care-needs certification such as the start and end dates of each certification term. The available sample period of this data is from June 2012 to March 2018.

As described in the previous section, recipients of the LTCI are required to take care-needs certification at the beginning of each certification term. We present notable reclassification pattern under repeated care-needs certification. Figure 4.2 shows distribution of recipients' SCT conditional on previous care-needs level. For instance, Panel (a) is the distribution of SCT conditional on those whose previous care-needs level is Care level 2. All of these distribution have a sharp discontinuity at a certain level of SCT. These discontinuities are unique characteristics of distribution conditional on previous care-needs level. Figure 4.5 presents the distribution of SCT calculated in the first certification. In contrast to the conditional distributions, the distribution of SCT calculated in the first certification has little discontinuity except threshold at 50. Our analysis thus focuses on the conditional distribution and reclassification of care-needs levels.

It is important to note that the position of discontinuities in the conditional distribution is exactly consistent with a threshold determining whether recipients can maintain the same LTCI benefits (including monthly coverage limit) as previous term. In the case of those whose previous care-needs level is Care level 2, recipients' care-needs level (and LTCI benefits) is lowered if their SCT is below 50 minutes. Panel (a) in Figure 4.2 shows that the position of the discontinuity coincides with the threshold and the number of recipients below the threshold is much smaller than those above it. In contrast, there is no clear discontinuity at a threshold determining whether recipients can receive more generous benefits as previous term (70 minutes in the case of Panel (b)). There exists the same tendency in other distribution.

The discontinuity in the conditional distribution suggests that the SCT is somehow manipulated to avoid classifying recipients into lower care-needs level than previous term. In the care-needs certification, recipients are informed of their care-needs level, not their SCT. Thus if the SCT is manipulated, it is considered to be manipulated by LTCI investigators who calculate it based on health checkups. The SCT tend not to decreases due to the deterioration of physical and mental ability of the elderly, but the discontinuity of distribution cannot be explained by itself. Figure 4.3 presents the conditional distribution of SCT where the previous SCT was normalized to zero. These distributions have no discontinuity so that the discontinuity in Figure 4.2 should be due to the thresholds.

Although we do not analyze the specific motivation of LTCI investigators to manipulate the SCT, there could be several possible reasons. First, LTCI investigators might want to guarantee that recipients can continue to use the services they have been using. If recipients' care-needs level and generosity of LTCI benefits are lowered, some recipients have to give up some LTC services to meet new benefit level. LTCI investigators may manipulate SCT and give recipients same LTCI benefits as before to prevent recipients from reducing services.

Second, LTCI investigators could have some psychological resistance to reducing recipients' care-needs level *per se*. Figure 4.4 shows that the distribution of the SCT conditional on recipients whose LTC utilization during the previous term is less than coverage limit of a lower care-needs level. These recipients do not have to give up their services even if they are assigned to a lower care-needs level. Nevertheless, there are clear discontinuities at the thresholds of the distribution, which clearly shows the manipulation of the SCT. This suggests that LTCI investigator try to avoid a lower care-needs level for reasons other than the reduction of recipients' LTC utilization. As shown in Figure 4.5, the SCT distributions conditional on the first certification (therefore no previous SCT) have little discontinuity. Therefore we focus on the distribution conditional previous care-needs level in the following analysis.

The object of this paper is to quantify the impact of manipulation of SCT on long-term are costs. If the manipulation does not exist, the SCT should be distributed smoothly conditional of previous care-needs levels. In this counterfactual case, the number of recipients whose care-needs level is lowered would be higher while the number of those whose care-needs level is same as before would be smaller. As is clear from the previous section, aggregate expenditure to public LTC services should be increased due to the manipulation because more recipients could receive generous LTCI benefits and fewer of them receive less generous ones. In the long run, the effect of manipulation will accumulate as more and more recipients are assigned into higher care-needs levels through repeated care-needs certification. In this paper, we propose a new empirical method that generalizes conventional ones used in the literature on "bunching", and estimate the upper and lower bound of the impact of manipulation on aggregate LTC costs.

4.4 Data

LTCI Administrative Data

In this study, we use administrative data on the LTCI from an anonymous local government near Tokyo. The first dataset is the data on care-needs certification, which contains the SCT of each recipient. This data is available for each certification term and the available sample period is from June 2012 to March 2018. It also contains a breakdown of how the final SCT was calculated (that is, it provides a hypothetical care time for each category of assistance) as well as other information related to care-needs certification such as the start and end dates of each certification term.

We also use LTCI claims data to observe recipients' LTC utilization. This dataset contains monthly information on LTC utilization for all LTCI recipients in the city as well as demographic characteristics and eligibility status. The available sample period of this data is the same as the data on care-needs certification. The claims data provides monthly information on how much each recipient used and paid for each type of service. It contains relatively limited information on the demographic characteristics of recipients, providing age and gender but no information on income and family structure. Its eligibility information contains the care-needs level, start and end dates of each certification term, coinsurance rate, and public subsidy eligibility.

The dataset for analysis is constructed by linking the data on care-needs certification and the claims data, which allows me to associate recipients' LTC utilization and their SCT. From this preliminary dataset, several types of recipients were excluded from the baseline analysis sample. First, I omit recipients whose care-needs level was altered by the Certification Committee to maintain the relationship between SCT and care-needs level.[3] Second, I omit recipients who live in a nursing home, because the coverage limit is applied only to those who use home-based care. Finally, I omit recipients who receive public subsidies for LTCI, because these individuals face a different incentive scheme than usual recipients. The remaining recipients constitute a baseline sample.

Summary Statistics

We show the summary statistics of the baseline sample in Table 4.3. Panel A to D are separated based on recipients' previous care-needs level, and each column represents recipients' current care-needs level. The first column shows that recipients' demographic characteristics are almost identical among care-needs levels in the previous term. Because LTCI is a social insurance for the elderly population, recipients are relatively old; the mean age of recipients is around 82 years old. More than half of recipients are women, and the fraction of those who pay a higher coinsurance rate (20%) is about 10 percent. The first column also shows that the higher the previous care-needs level, the higher the average current expenditure on LTC.

All panels show that older recipients are more likely to be assigned to higher care-needs

[3] About 11% of recipients are omitted by this procedure.

levels probably because their health conditions are more vulnerable. Other characteristics such as gender and coinsurance do not vary among current levels. In contrast, LTC expenditure monotonically increases as the care-needs level becomes higher. Figure 4.6 presents the relationship between SCT and monthly LTC expenditure conditional on the previous care-needs level. The grey area indicates that the previous SCT was within this range. This figure shows that the expenditure significantly varies at thresholds separating care-needs levels. Therefore the manipulation of SCT which move recipients to higher care-needs level should increase LTC expenditure.

4.5 Empirical Strategy

Our goal is to quantify to what extent the manipulation of SCT increases LTC expenditure. A central issue is how to construct the counterfactual distribution of SCT without manipulation. Previous literature on bunching assumes a specific functional form such as polynomials for a counterfactual distribution and constructs it by extrapolating the un-manipulable range of observed distribution.

Counterfactual distributions constructed by the traditional approach could be unreliable when the manipulable range of distribution is sufficiently large. In this case, as Kleven (2016) pointed out, researchers are required to extrapolate over a large range relying on a narrow un-manipulable range of distribution. The unreasonable extrapolation raises doubts about the credibility of counterfactual distributions. Figure 4.2 indicates that conditional distributions of the SCT have exactly the same problem; the SCT could be manipulated even if it is far below the threshold so that we need a large extrapolation for counterfactual distribution.

To address this issue, we propose a new method to construct counterfactual distribution which leads to partial identification of the impact of manipulation. The key to our approach is to impose shape restrictions on counterfactual distribution without particular functional forms. This partial identification approach can take into account the statistical error of extrapolation by a bound of estimates. Equally important, comparing the estimates under a different set of assumptions can provide a deeper understanding of the association between the estimates and assumptions we impose.

4.5.1 Model and Identification

We formulate our empirical strategy based on a conventional program evaluation framework. Let X be an observed SCT and X may be manipulated. Let X^* be a counterfactual SCT in the absence of manipulation. We define the support of X and X^* as $\mathcal{X}$. Following previous bunching literature, we assume that SCT could be manipulated only in a certain range around the threshold. In other words, SCT is not manipulated if it is far from the threshold and outside the manipulatable range. That is, we assume

$$X = X^* \quad \text{if } X \in K, \tag{4.1}$$

where $K \subset \mathcal{X}$ is an un-manipulatable range far from the threshold while $K^C \equiv \mathcal{X} \setminus K$ is manipulable range. When $X \in K^C$, X may not be consistent with X^*. In the following discussion, we assume that K is known. How to set the un-manipulatable range will be discussed at the end of this section.

We consider the following potential outcome framework. Potential outcomes are indexed against potential values x of X and denoted by $Y(x)$. For example, $Y(x)$ is monthly expenditure to LTC services per recipient given SCT x. Using the above notation, the observed outcome Y can be expressed as follows:

$$Y = Y(X). \tag{4.2}$$

Then, the counterfactual outcome Y^* can be written as

$$Y^* = Y(X^*). \tag{4.3}$$

Hence, the average treatment effect of manipulation is defined as $E[Y^*] - E[Y]$. Because $E[Y]$ is simply an average of observed outcomes, we focus on identification of $\theta_0 \equiv E[Y^*]$.

We define the conditional expectation of counterfactual outcomes $g(x) \equiv E[Y(x)|X^* = x]$ for all x. For example, $g(x)$ is LTC expenditure conditional on SCT x. Hereafter, we assume that $g(x)$ is identified for all x. Let f_X and f_{X^*} denote the PDF of X and X^*. Using the above notation, θ_0 can be expressed as follows:

$$\theta_0 = \int_{\mathcal{X}} g(x) f_{X^*}(x) dx. \tag{4.4}$$

This implies that if $f_{X^*}(x)$ is identified for all x we can identify the target parameter θ_0.

Shape Restrictions on Counterfactual Distribution

We now specify shape restrictions on the counterfactual distribution of SCT. Without restrictions, the shape of the counterfactual distribution can be anything and we cannot obtain any information from observed distribution. We impose the following three restrictions on a counterfactual PDF $f_{X^*}(x)$. The first restriction is that $f_{X^*}(x)$ coincides with observed PDF $f_X(x)$ in un-manipulatable range:

$$f_{X^*}(x) = f_X(x) \quad \text{for } x \in K. \tag{4.5}$$

This follows from (4.1) assuming that SCT is not manipulated outside manipulatable range, and most bunching literature relies on the same type of assumption. Because K^C is an interval, the above condition can be rewritten using cumulative distribution functions:

$$F_{X^*}(x) = F_X(x) \quad \text{for } x \in K, \tag{4.6}$$

where $F_{X^*}(x)$ and $F_X(x)$ are cumulative distribution functions corresponding to counterfactual PDF $f_{X^*}(x)$ and observed PDF $f_X(x)$.

The second restriction is that $f_{X^*}(x)$ is log-concave. A PDF f is log-concave if $f(x) = exp(h(x))$ with concave function h. Any log-concave PDF is continuous and single-peaked, and it includes well-known univariate parametric families such as normal, gamma, and beta distribution[4]. This assumption specifies a global shape of the counterfactual PDF which has no discontinuity and bunching around thresholds. Diamond and Persson (2017) also impose log-concavity on counterfactual PDF of test score. Because of the wide manipulable range, it is important to restrict the global shape of distribution to obtain informative bound of

[4] $Gamma(r, \lambda)$ for $\lambda \geq 1$ and $Beta(a, b)$ for $a, b \geq 1$ are log-concave.

counterfactual PDF and target parameters.

The third restriction is that the counterfactual distribution is first-order stochastically dominated by the observed distribution :

$$F_{X^*}(x) \geq F_X(x) \quad \text{for all } x. \tag{4.7}$$

This restriction implies that SCT could be increased, but not decreased by manipulation. In other words, the true SCT without manipulation could be lower than the observed value so that the counterfactual PDF of SCT should be shifted to the left of the observed one. We only impose above shape restrictions on counterfactual PDF and do not assume any particular functional forms.

The advantage of our method is that various restrictions can be incorporated flexibly. We use restrictions that are appropriate for our study, but what is appropriate restriction depends on each application. Our method allows researchers to flexibly impose restrictions based on institutional background and available information regarding counterfactual distribution. For example, researchers might want to restrict the slope of counterfactual PDF rather than restrict its global shape. In this case, researchers can impose Lipschitz continuity as in Bertanha et al. (2019) instead of log-concavity. The following discussion can be easily modified based on different restrictions.

Identified Set

Under the shape restrictions specified above, the identified set of un-manipulated distribution $f_{X^*}(x)$ is

$$\begin{aligned} \mathcal{F}_{IS} \quad \equiv \quad & \left\{ f \in \mathcal{F} : \int_{-\infty}^{x} f(u)du = F_X(x) \text{ for } x \in K, \right. \\ & \qquad \left. \text{and} \quad \int_{-\infty}^{x} f(u)du \geq F_X(x) \text{ for all } x. \right\}, \end{aligned} \tag{4.8}$$

where $\mathcal{F}$ is a parameter space of f and satisfies the following condition:

$$\mathcal{F} \subset \{f : f \text{ is a log-concave PDF and } supp(f) = \mathcal{X}.\}. \tag{4.9}$$

Note that we implicitly assume that $\int f(u)du = 1$ from the definition of PDF. Without parametric specification, there are infinite number of log-concave PDF in $\mathcal{F}_{IS}$.

For estimation procedure described later, it is useful to summarize the conditions in (4.8) as a single equation. To do this, we first define the following functional:

$$Q(f) \equiv w \int_K |F_X(x) - F_f(x)| \, dx + (1-w) \int_{K^C} |F_X(x) - F_f(x)|_+ \, dx, \tag{4.10}$$

where $F_f(x) \equiv \int_{-\infty}^{x} f(u)du$, $|a|_+ \equiv \max\{0, a\}$, and $w \in (0,1)$. The first term of $Q(f)$ can be interpreted as a penalty for the discrepancy between the counterfactual and empirical distributions in the un-manipulatable range K. Similarly, the second term represents a penalty for violating the assumption that $F_f(x)$ is first-order stochastically dominated by $F_X(x)$ ($F_f(x)$ should be larger than $F_X(x)$) in the manipulatable range K^C. If the un-manipulability in K and stochastic dominance condition are satisfied, both terms in $Q(f)$ is zero so that $Q(f)$ is also zero. Therefore, we have another expression for the identified set using $Q(f)$:

$$\mathcal{F}_{IS} = \{ f \in \mathcal{F} : Q(f) = 0 \}, \tag{4.11}$$

The parameter of interest is θ_0 in (4.4), the expected expenditure to LTC services under un-manipulated SCT distribution. θ_0 can only be partially identified under the shape restrictions because the un-manipulated distribution $f_{X^*}(x)$ is not point identified. The identified set $\mathcal{F}_{IS}$ provides lower and upper bounds of θ_0:

$$\begin{aligned} \overline{\theta^*} &= \sup_{f \in \mathcal{F}_{IS}} \int g(x) f(x) dx, \\ \underline{\theta^*} &= \inf_{f \in \mathcal{F}_{IS}} \int g(x) f(x) dx. \end{aligned} \tag{4.12}$$

Under given restrictions, $\overline{\theta^*}$ is the largest value of the expected average expenditure to LTC services based on the un-manipulated distribution of SCT. $\underline{\theta^*}$ is the analogous smallest value.

4.5.2 Estimation

We use sieve estimation methods to estimate $\overline{\theta^*}$ and $\underline{\theta^*}$ following Chen (2007) and Mogstad, Santos, and Torgovitsky (2018). We first specify a finite dimensional subset of PDF, $\mathcal{F}_n \subseteq \mathcal{F}$, satisfying log-concavity:

$$\mathcal{F}_n \equiv \Big\{ f = exp(h) : h(x) = \sum_{s=0}^{S_n} a_s b_s^{S_n}(x), \\ \text{and } \ a_s - 2a_{s+1} + a_{s+2} \leq 0 \text{ for } s = 0, 1, \cdots, S_n - 2. \Big\}, \quad (4.13)$$

where n is sample size, $b_s^{S_n}$ is the s-th Bernstein basis polynomial of degree S_n and $S_n \to \infty$, and $\{a_s\}_{s=0}^{S_n}$ parameterizes functions in $\mathcal{F}_n$. From the property of the Bernstein polynomials, any functions in $\mathcal{F}_n$ are log-concave (Wang and Ghosh, 2012; Diamond and Persson, 2017).

The identified set defined as (4.11) should satisfy the un-manipulability in K and stochastic dominance as well as log-concavity. In estimation, we have to take into account statistical error related to these restrictions due to finite sample. We can show that the estimator of $\mathcal{F}_{IS}$ can be written as

$$\hat{\mathcal{F}}_{IS} = \left\{ f \in \mathcal{F}_n : \sqrt{n}\hat{Q}_n(f) \leq \inf_{\tilde{f} \in \mathcal{F}_n} \sqrt{n}\hat{Q}_n(\tilde{f}) + \kappa_n \right\}, \quad (4.14)$$

where

$$\hat{Q}_n(f) = w \int_K \left| \hat{F}_n(x) - F_f(x) \right| dx + (1-w) \int_{K^C} \left| \hat{F}_n(x) - F_f(x) \right|_+ dx, \quad (4.15)$$

κ_n is a tolerance parameter that diverges to $+\infty$ more slowly than $\sqrt{n}$, $F_f(x) \equiv \int_{-\infty}^x f(u)du$, $\hat{F}_n$ is the empirical distribution function of X. In our estimation, we set $w = 0.5$. The constraints in (4.15) allow for f that come closest to satisfying the restriction in (4.11) up until a tolerance κ_n, instead of making f satisfy the restriction without any error. The first step of this estimation is to minimize $\hat{Q}_n(f)$ using log-concave PDF to derive $\inf_{\tilde{f} \in \mathcal{F}_n} \sqrt{n}\hat{Q}_n(\tilde{f})$ in (4.14).

The next step is to calculate a tolerance parameter κ_n. In our estimation, we calculate a tolerance parameter using bootstrap. Let s be an index for simulation of bootstrap and S be the number of simulation. We calculate empirical distribution $\hat{F}_n^s$ using different support $\mathcal{X}^s \subset \mathcal{X}$ for each s. Given $f^* \equiv \text{argmin}_{\tilde{f} \in \mathcal{F}_n} \sqrt{n}\hat{Q}_n(\tilde{f})$ and its CDF F_{f^*}, we calculate $\hat{Q}_n^s(f^*)$

by replacing $\hat{F}_n(x)$ and $F_f(x)$ in (4.15) with $\hat{F}_n^s(x)$ and $F_{f^*}(x)$ respectively:

$$\hat{Q}_n^s(f^*) = w \int_K \left| \hat{F}_n^s(x) - F_{f^*}(x) \right| dx + (1-w) \int_{K^C} \left| \hat{F}_n^s(x) - F_{f^*}(x) \right|_+ dx. \tag{4.16}$$

Then, we specify κ_n as follows:

$$\kappa_n = \left(C_{0.9}^S - \frac{1}{S} \sum_{s=1}^{S} \sqrt{n} \hat{Q}_n^s(f^*) \right) \times \log n, \tag{4.17}$$

where $C_{0.9}^S$ and $1/S \sum_{s=1}^{S} \sqrt{n}\hat{Q}_n^s(f^*)$ are the 0.9-th quantile and mean of $\sqrt{n}\hat{Q}_n(f^*)$ calculated by the bootstrap.

It is also necessary to estimate LTC expenditure conditional on SCT, $g(x)$. Let $\hat{g}(x)$ be an estimator of $g(x)$ and $\hat{g}(x)$ is calculated as a predicted value of the following regression model:

$$\begin{aligned} g(x) \quad &= \quad \beta_0 + \beta_1 x + \gamma_1 \mathbf{1}\{x \geq 32\} + \gamma_2 \mathbf{1}\{x \geq 50\} \\ &\quad\; + \gamma_3 \mathbf{1}\{x \geq 70\} + \gamma_4 \mathbf{1}\{x \geq 90\} + \gamma_5 \mathbf{1}\{x \geq 110\} + \varepsilon, \end{aligned} \tag{4.18}$$

where $\mathbf{1}\{\cdot\}$ is an indicator function which is equal to one if the statement in parenthesis is true, and zero otherwise. The specification of $g(x)$ takes into account that LTC expenditure vary depending not only on SCT x but also on the care-needs level. The indicator function allows the expenditure to vary by care-needs level. If SCT is between 32 and 50, recipients are assigned to either Support level 2 or Care level 1, and LTC expenditure is affected by which category they are assigned. We allocate recipients into one of these two categories based on the observed assignment probability if $x \in [32, 50]$.

Then, we estimate $\overline{\theta^*}$ and $\underline{\theta^*}$ with the following optimization problems:

$$\begin{aligned} \hat{\overline{\theta^*}} &= \sup_{f \in \hat{\mathcal{F}}_{IS}} \int \hat{g}(x) f(x) dx, \\ \underline{\hat{\theta^*}} &= \inf_{f \in \hat{\mathcal{F}}_{IS}} \int \hat{g}(x) f(x) dx. \end{aligned} \tag{4.19}$$

The proof of consistency under general condition is provided in Appendix A.

(Un-)Manipulable Range of Distribution

To implement above estimation, we have to determine the manipulable range of distribution, K^C, around thresholds determining whether recipients can maintain the same LTCI benefits. Figure 4.2 shows that when SCT is smaller than the threshold value, it is manipulated to exceed the threshold value even if it is far from the threshold. This suggests that the start of the manipulable range should be as far away from the threshold as possible. Therefore, we assume that the starting point of K is 32, which is the lowest SCT value separating care-needs levels (Support level 1 and 2). We also assume that SCT is manipulated to maintain the previous care-needs level, which is equivalent to assuming that the "excess mass" of conditional distribution is concentrated in the range of previous care-needs level. Therefore, we specify K^C as follows:

$$K^C = \begin{cases} [32, 70] & \text{if threshold} = 50 \\ [32, 90] & \text{if threshold} = 70 \\ [32, 110] & \text{if threshold} = 90 \\ [32, 130] & \text{if threshold} = 110 \end{cases} \tag{4.20}$$

In the case of Care level 5, we set the end point of K^C as 130 so that the range of excess mass is same as other care-needs levels. Un-manipulable range, K, is automatically determined by K^C.

How to set the manipulable range K^C could significantly affect the estimates for the impact of manipulation. We thus use another manipulable range and check how the setting of K^C affects the estimates. Because we specify above K^C as wide as possible, another manipulable range is set to be relatively narrow. We assume that SCT can be manipulated to raise care-needs level by one category:

$$K^C = \begin{cases} [32, 70] & \text{if threshold} = 50 \\ [50, 90] & \text{if threshold} = 70 \\ [70, 110] & \text{if threshold} = 90 \\ [90, 130] & \text{if threshold} = 110 \end{cases} \tag{4.21}$$

In this specification, the manipulable range covers only two care-needs levels separated by the threshold.

4.6 Results

Using the above estimation method, we estimate the upper and lower bound of the impact of SCT manipulation on LTC expenditure. Specifically, we compare the mean LTC expenditure based on the observed and counterfactual distribution of SCT. Figure 4.7 presents the observed and counterfactual distribution of SCT based on the wide manipulable ranges specified as (4.20). The solid vertical line indicates the threshold, and two dashed vertical lines indicate the start and endpoint of the manipulable range. The red and green lines represent the counterfactual PDF which produces the upper and lower bound of the mean LTC expenditure. Based on the above estimation method, these log-concave PDF are constructed so as to be close to the observed PDF in the un-manipulable range, while being first-order stochastically dominated by the observed PDF in the manipulable range. Because higher SCT leads to higher LTC expenditure, the upper bound distributions are to the right of the lower bound ones. Figure 4.7 shows that the upper and lower bound of counterfactual PDF are relatively close in Panel (a) and (b) while they are widely separated in Panel (c) and (d). This suggests that the extensive extrapolation and narrow reliable ranges cause wide bounds of counterfactual PDF, which is consistent with the discussion in Section 4.5.

Table 4.4 summarizes the impact of manipulation on LTC expenditure based on the wide manipulable range. Panel A shows the mean monthly LTC expenditure per recipient conditional on the previous care-needs level. The mean LTC expenditure is separately presented based on the observed PDF and counterfactual ones. The first to fourth columns represent each previous care-needs level, and the fifth column represents the mean expenditure un-conditional on the previous level. Regardless of observed and counterfactual PDF, the LTC expenditure increases as the previous care-needs level get higher. This indicates that the LTC expenditure is affected by the previous expenditure and health conditions. Our partial identification and nonparametric estimation approach produce reasonable bounds for counterfactual LTC expenditure; the upper bound of mean LTC expenditure is higher than the lower bound one by roughly 10%. It is important to note that even the upper bound of counterfactual expenditure is lower than the observed expenditure, which suggests the manipulation of SCT increases LTC expenditure.

Based on the observed and counterfactual expenditure, we are able to calculate the impact of manipulation on LTC expenditure. Panel B in Table 4.4 presents the annual expenditure calculated from the monthly expenditure in Panel A. Based on the annual expenditure, we then calculate the extent to which the manipulation of SCT increases LTC expenditure by subtracting counterfactual expenditure from the observed one. Thus the upper bound of the impact of manipulation is the difference between observed expenditure and lower bound of expenditure in Panel B, and the lower bound of the impact of manipulation is the difference between observed expenditure and upper bound of expenditure. The lower bound of expenditure increase is 629 USD per recipient so that even the most conservative estimates show that the manipulation significantly increases LTC expenditure. The upper bound is 2,393 USD, which indicates that the manipulation effects could be four times the lower bound.

To examine how the setting of (un-)manipulable range affects the estimates, we also estimate the manipulation effect using the narrow manipulable range specified as (4.21). Figure 4.8 presents the observed and counterfactual PDF based on the narrow manipulable range. Due to the relatively less extrapolation and wide un-manipulable range, the upper and lower bound PDF are closer than the case of using the wide manipulable range shown in Figure 4.7. The bound of manipulation effect is narrower than in the case of the wide manipulable range. The lower bound is 658 USD and slightly higher than the estimate from wide manipulable range, while the upper bound is 1,702 USD and much lower than the estimate from wide manipulable range. Therefore the lower bound of manipulation effect is robust to the change in manipulable ranges, and it appears that the manipulation of SCT increases LTC expenditure by at least around 650 USD annually per recipient.

4.7 Conclusion

We examine the impact of manipulation of social insurance benefits in the context of LTCI in Japan. The generosity of LTCI benefits is determined by care-needs levels and the LTC investigator categorizes recipients into a specific level based on health score (SCT). We first find a novel discontinuity and bunching in the distribution of SCT conditional of the previous care-needs level. The conditional distribution indicates that the LTC investigator, who calculates SCT, manipulate SCT and avoid assigning recipients into the lower care-needs level. Because the higher care-needs level provides more generous LTCI benefits, the manipulation is likely to increase LTC expenditure. We quantify the manipulation effect by comparing LTC expenditure based on observed SCT distribution and counterfactual one.

To construct counterfactual distribution without manipulation, we propose a partial identification strategy assuming no functional specification. The advantage of our method is to allow for incorporating flexible shape restrictions tailored to specific applications. In our application, we assume that the counterfactual PDF is log-concave, and the observed distribution has first-order stochastic dominance over the counterfactual one. We also propose the estimation method that provides a consistent estimator for the upper and lower bound of manipulation effects.

We find that the manipulation of SCT increases annual LTC expenditure by at least 629 USD per recipient from the lower bound estimates. The upper bound indicates that the increase in LTC expenditure due to the manipulation could amount to 2,393 USD, four times larger than the lower bound. The difference between lower and upper bound estimates suggests that it is important to partially identify counterfactual distribution when we need to extensively extrapolate to construct it. In this study, we do not analyze specific causes of manipulation and this issue is left for future research.

Bibliography

[1] **Aliprantis, Charalambos D., and Kim C. Border.** 2006. "Infinite Dimensional Analysis: A Hitchhiker's Guide." *Springer.*

[2] **Bertanha, Marinho, Andrew H. McCallum, and Mathan Seegert.** 2019. "Better Bunching, Nicer Notching." *Working Paper.*

[3] **Best, Michael Carlos, Janes Cloyne, Ethan Ilzetzky, and Henrik Kleven.** 2018. "Estimating the Elasticity of Intertemporal Substitution Using Mortgage Notches." *Working Paper.*

[4] **Blomquist, Soren, Whiteney K. Newey, Anil Kumar, and Che-Yuan Liang.** 2019. "On Bunching and Identification of the Taxable Income Elasticity." *Working Paper.*

[5] **Blundell, Richard, Amanda Gosling, Hidehiko Ichimura, and Costas Meghir.** 2007. "Changes in the Distribution of Male and Female Wages Accounting for Employment Composition Using Bounds." *Econometrica*, 75(2), pp. 323-363.

[6] **Brollo, Fernanda, Katja Kaufmann, and Eliana La Ferrara.** 2019. "The Political Economy of Program Enforcement: Evidence from Brazil." *Journal of European Economic Association*, forthcoming.

[7] **Camacho, Adriana and Emily Conover.** 2011. "Manipulation of Social Program Eligibility." *American Economic Journal: Economic Policy*, 3(2), pp. 41-63.

[8] **Chen, Xiaohong.** 2007. "Chapter 76 Large Sample Sieve Estimation of Semi-Nonparametric Models." *Handbook of Econometrics*, Vol. 6, Part 2, ed. by J. J. Heckman and E. E. Leamer. Elsevier, pp. 5549-5632.

[9] **Chen, Zhao, Zhikuo Liu, Juan Carlos Suárez Serrato, and Daniel Yi Xu.** 2019. "Notching R&D Investment with Corporate Income Tax Cuts in China." *Working Paper.*

[10] **Chetty, Raj, John N. Friedman, and Emmanuel Saez.** 2013. "Using Differences in Knowledge Across Neighborhoods to Uncover the Impacts of the EITC on Earnings." *American Economic Review*, 103(7), pp. 2683-2721.

[11] **Chetty, Raj, John N. Friedman, Tore Olsen and Luigi Pistaferri.** 2011. "Adjustment Costs, Firm Responses, and Macro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records." *Quarterly Journal of Economics*, 126(2), pp. 749-804.

[12] **Chetverikov, Denis, Andres Santos, and Azeem M. Shaikh.** 2018. "The Econometrics of Shape Restrictions." *Annual Review of Economics*, 10, pp. 31-63.

[13] **Dee, Thomas S., Will Dobbie, Brian A. Jacob, and Jonah Rockoff.** 2018. "The Causes and Consequences of Test Score Manipulation: Evidence from the New York Regents Examinations." *Working Paper.*

[14] **DeFusco, Anthony A. and Andrew Paciorek.** 2017. "The Interest Rate Elasticity of Mortgage Demand: Evidence from Bunching at the Conforming Loan Limit." *American Economic Journal: Economic Policy*, 9(1), pp. 210-240.

[15] **Diamond, Rebecca and Petra Persson.** 2017. "The Long-term Consequences of Teacher Discretion in Grading of High-stakes Tests." *Working Paper.*

[16] **Duflo, Esther, Michael Greenstone, Rohini Pande, and Nicholas Ryan.** 2018. "The Value of Regulatory Discretion: Estimates from Environmental Inspection in India." *Econometrica*, 86(6), pp. 2123-2160.

[17] **Garicano, Luis, Claire Lelagre, and John Van Reenen.** 2016. "Firm Size Distortions and the Productivity Distribution: Evidence from France." *American Economic Review*, 106(11), pp. 3439-3479.

[18] **Gerard, Francois, Christoph Rothe, and Miikka Rokkanen.** 2019. "Bounds on Treatment Effects in Regression Discontinuity Designs with a Manipulated Running Variable." *Working Paper.*

[19] **Kleven, Henrik J.** 2016. "Bunching." *Annual Review of Economics*, 8, pp. 435-464.

[20] **Kleven, Henrik J. and Mazhar Waseem.** 2013. "Using Notches to Uncover Optimazation Frictions and Structural Elasticities: Theory and Evidence from Pakistan." *Quarterly Journal of Economics*, 128(2), pp. 669-723.

[21] **Manoli, Day and Andrea Weber.** 2016. "Nonparametric Evidence on the Effects of Financial Incentives on Retirement Decisions." *American Economic Journal: Economic Policy*, 8(4), pp. 160-182.

[22] **Marx, Benjamin M.** 2019. "Dynamic Bunching Estimation with Panel Data." *Working Paper.*

[23] **Mogstad, Magne, Andres Santos, and Alexander Torgovitsky.** 2018. "Using Instrumental Variables for Inference about Policy Relevant Treatment Parameters." *Econometrica*, 86(5), pp. 1589-1619.

[24] **Persson, Petra.** 2018. "Social Insurance and the Marriage Market." *Journal of Political Economy*, forthcoming.

[25] **Saez, Emmanuel.** 2010. "Does Taxpayer Bunch at Kink Points?" *American Economic Journal, Economic Policy*, 2(3), pp. 180-212.

[26] **Takahashi, Masaki.** 2019. "The Behavioral Effect of Insurance Coverage and Its Health Consequences: Evidence from Long-Term Care." *Working Paper*.

[27] **Wang, Jiangdian and S.K. Ghosh.** 2012. "Shape Restricted Nonparametric Regression with Bernstein Polynomials." *Computational Statistics & Data Analysis*, 56(9), pp. 2729-2741.

A. Figures

Figure 4.1: Bunching in the Conditional Distribution of SCT

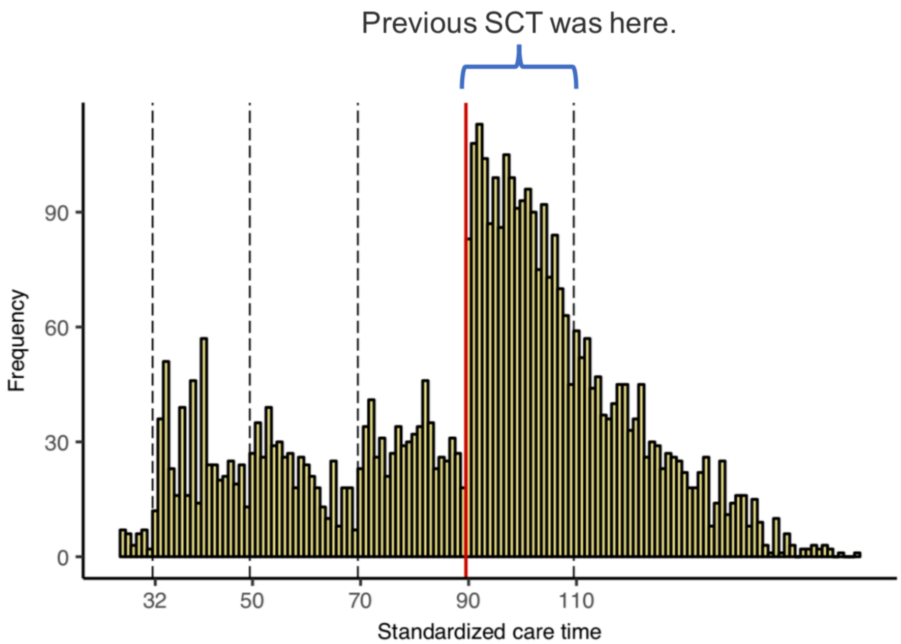


Figure 4.2: Distribution of the Standardized Care Time

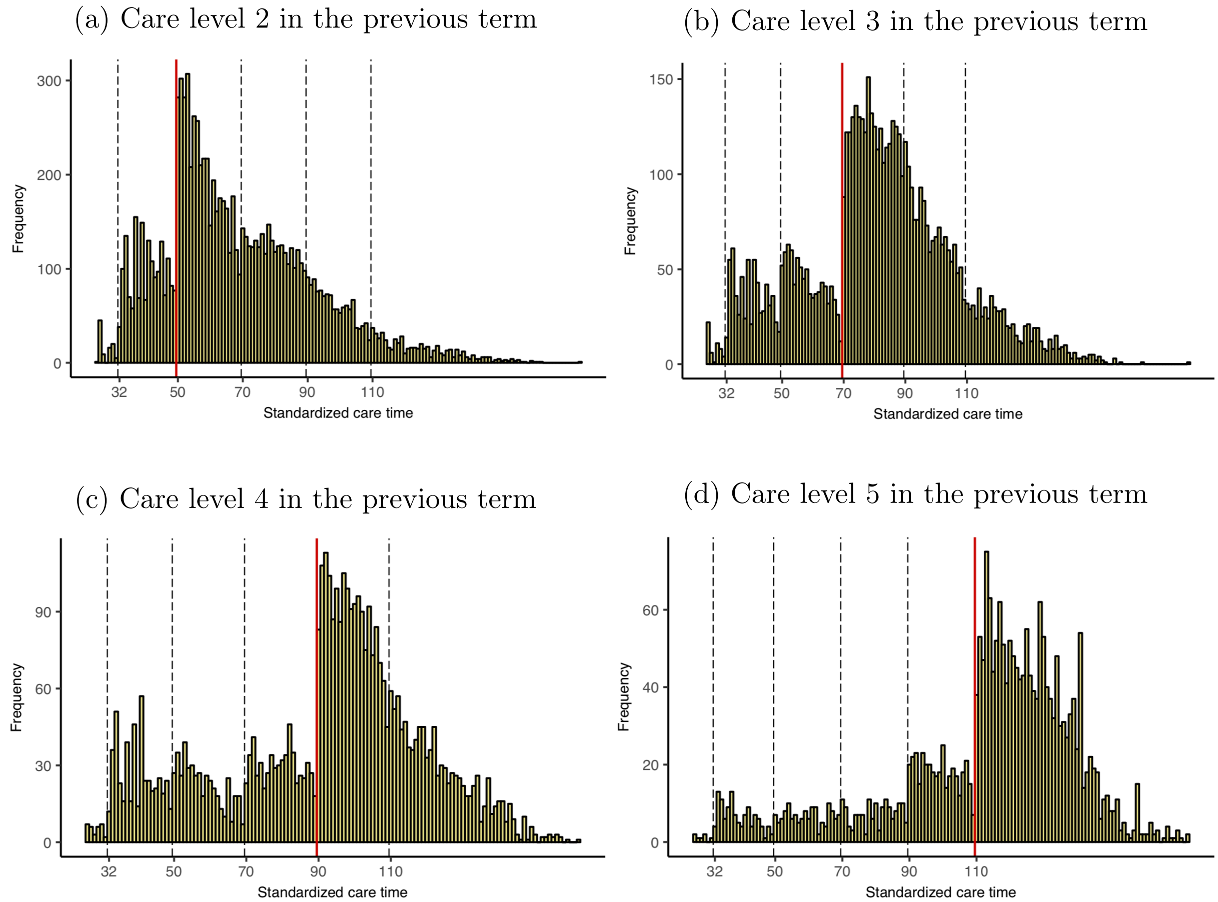
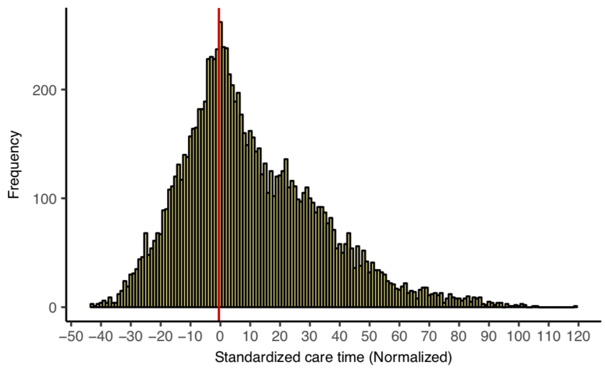
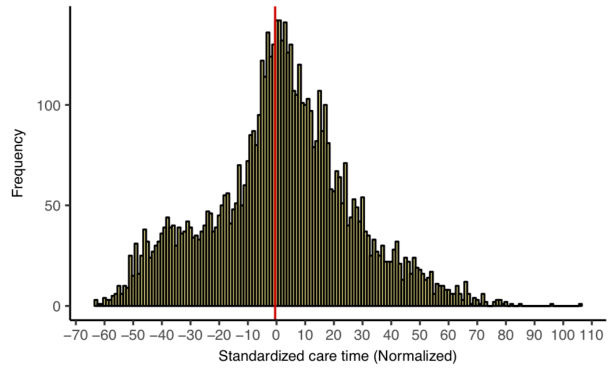


Figure 4.3: Distribution of the Standardized Care Time (Normalized)

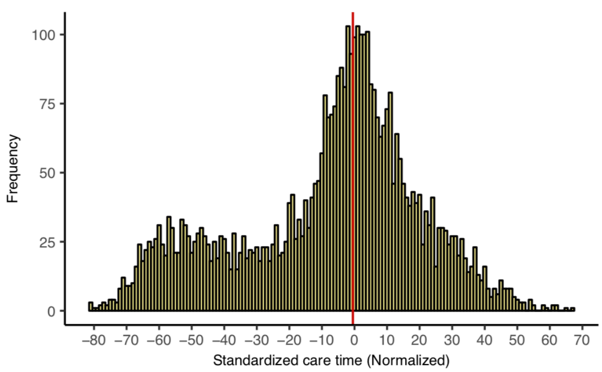
(a) Care level 2 in the previous term



(b) Care level 3 in the previous term



(c) Care level 4 in the previous term



(d) Care level 5 in the previous term

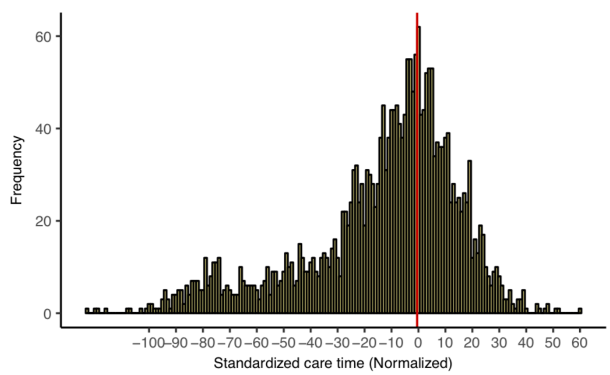
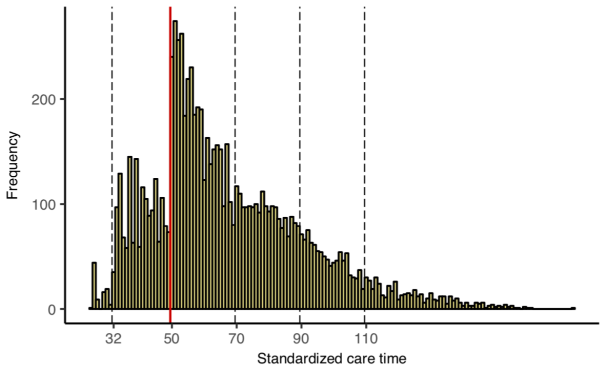
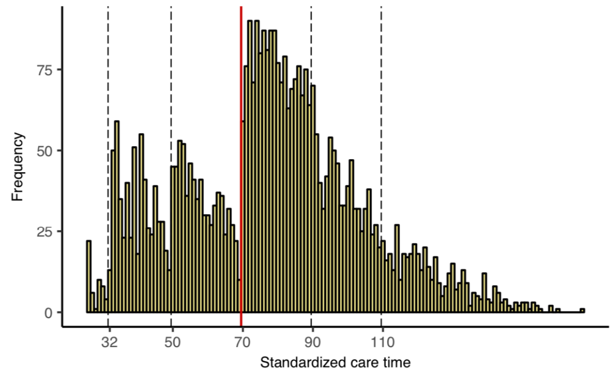


Figure 4.4: Distribution of the Standardized Care Time (Low-Demand Recipients)

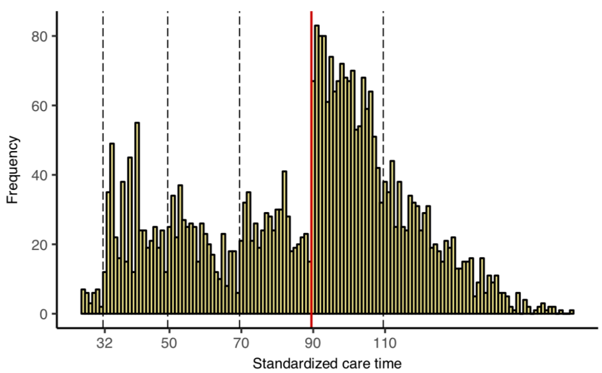
(a) Care level 2 in the previous term



(b) Care level 3 in the previous term



(c) Care level 4 in the previous term



(d) Care level 5 in the previous term

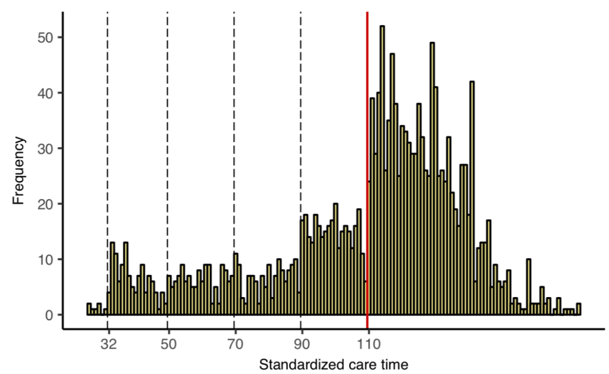


Figure 4.5: Distribution of the Standardized Care Time (First Certification)

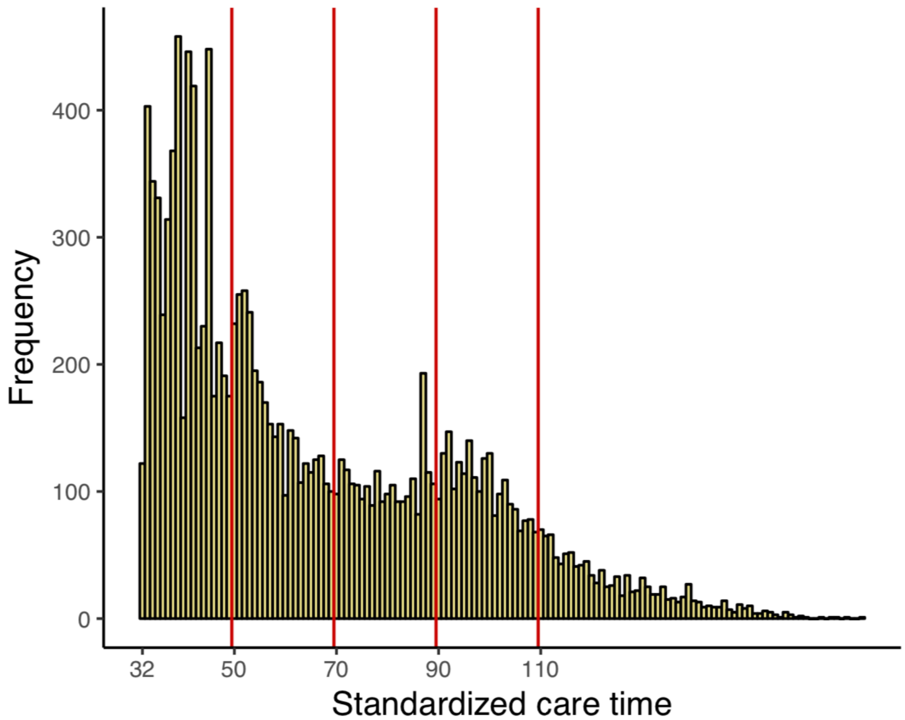
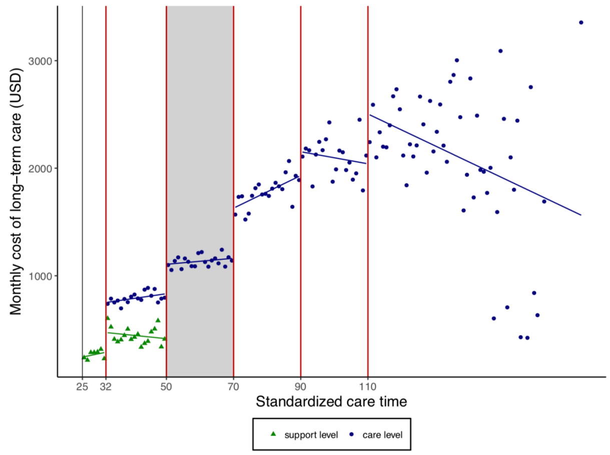
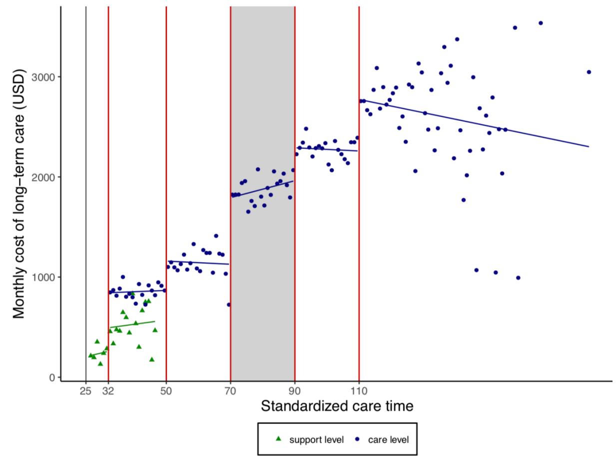


Figure 4.6: Monthly Expenditure to Long-Term Care Services

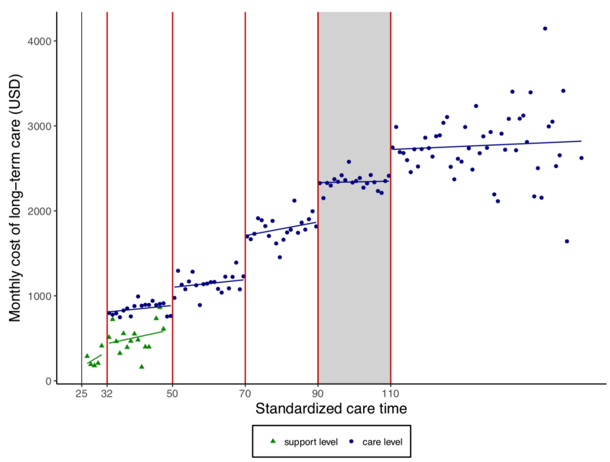
(a) Care level 2 in the previous term



(b) Care level 3 in the previous term



(c) Care level 4 in the previous term



(d) Care level 5 in the previous term

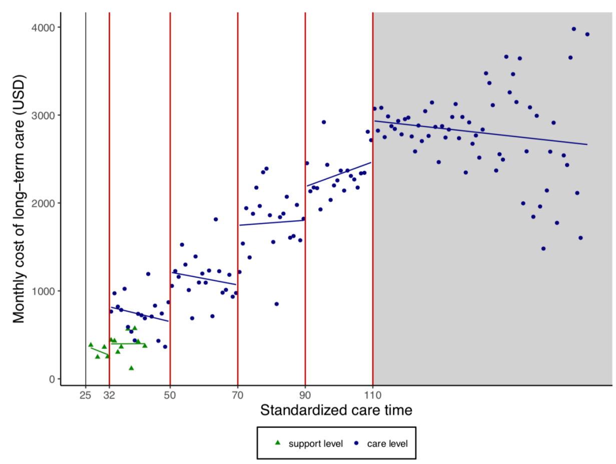
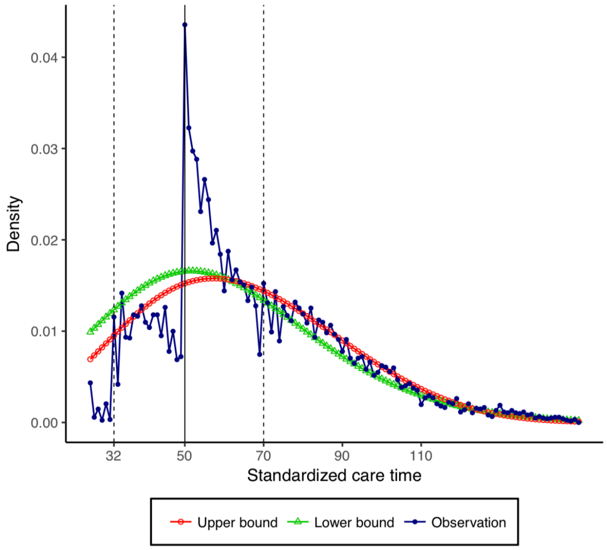
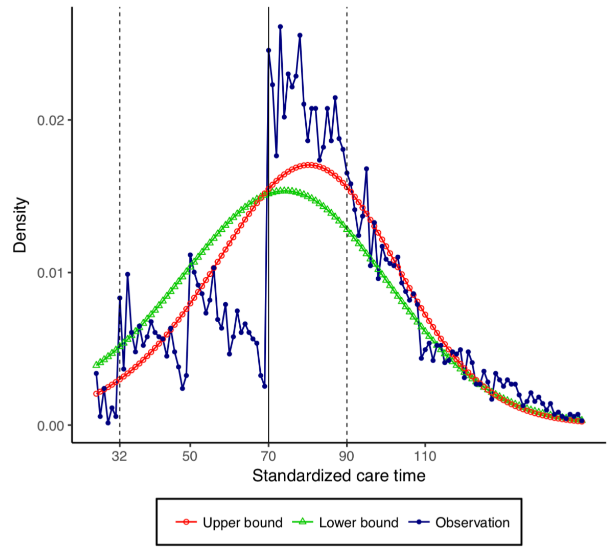


Figure 4.7: Observed and Counterfactual Distribution of the SCT

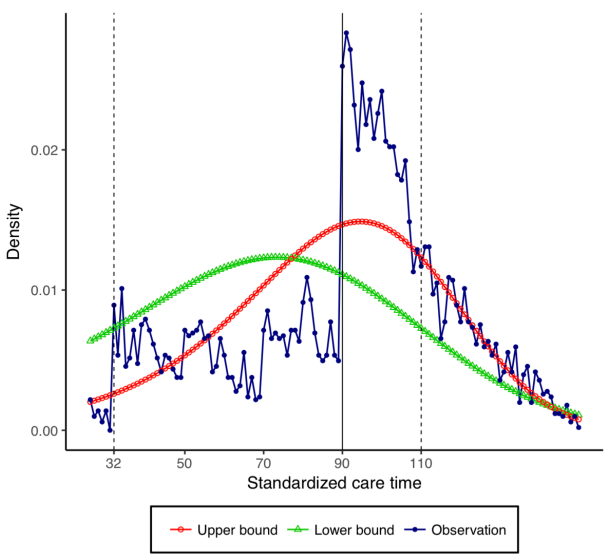
(a) Care level 2 in the previous term



(b) Care level 3 in the previous term



(c) Care level 4 in the previous term



(d) Care level 5 in the previous term

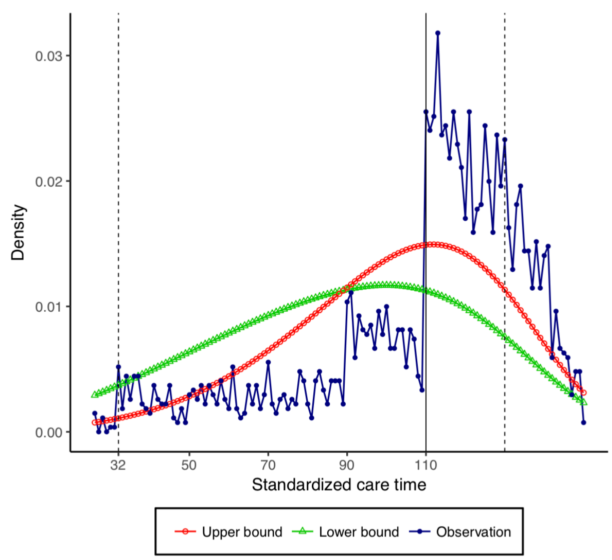
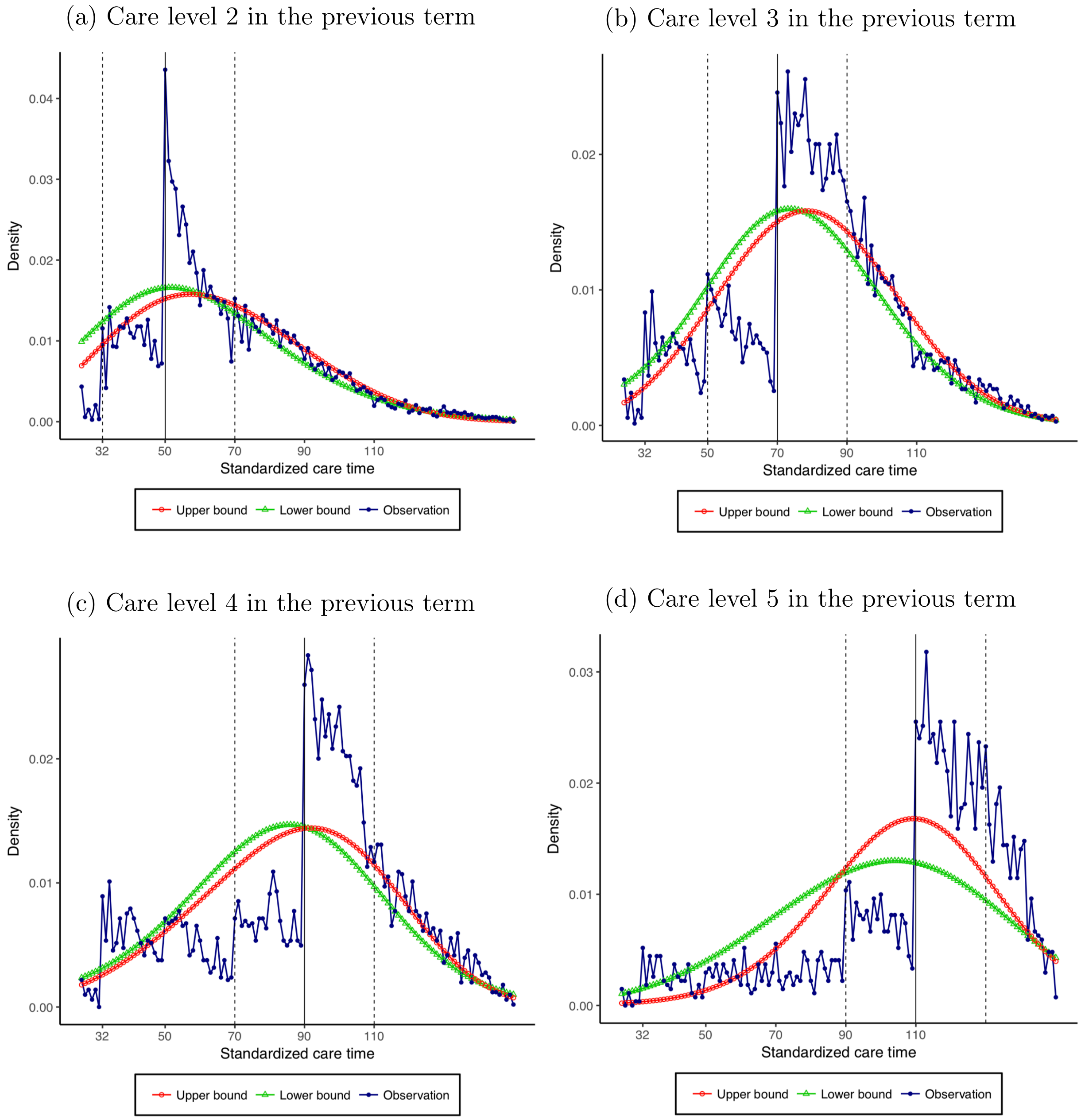


Figure 4.8: Observed and Counterfactual Distribution of the SCT



(a) Care level 2 in the previous term

(b) Care level 3 in the previous term

(c) Care level 4 in the previous term

(d) Care level 5 in the previous term

Table 4.1: Monthly Coverage Limits for Each Care-needs Level

Care-needs level	Standardized care time	Coverage cap (unit)
Support level 1	25.0 – 31.9	5,003
Support level 2	32.0 – 49.9	10,473
Care level 1	32.0 – 49.9	16,692
Care level 2	50.0 – 69.9	19,616
Care level 3	70.0 – 89.9	26,931
Care level 4	90.0 – 109.9	30,806
Care level 5	≥110.0	36,065

Table 4.2: Category of Assistance and Range of Time Length

Category of assistance	Range of time length (minutes)
Eating	1.1 – 71.4
Transferring	0.4 – 21.4
Toileting	0.2 – 28.0
Hygiene	1.2 – 24.3
Housework	0.4 – 11.3
Dementia	5.8 – 21.2
Exercise	0.5 – 15.4
Medical care	1.0 – 37.2
Standardized care time	10.6 – 230.6

Table 4.3: Summary Statistics

		Support level		Care level				
	All levels (1)	level 1 (2)	level 2 (3)	level 1 (4)	level 2 (5)	level 3 (6)	level 4 (7)	level 5 (8)
		A. Care level 2 in the previous term						
Age	82.6	80.3	80.3	81.6	82.0	83.9	84.1	83.7
Woman	0.62	0.72	0.70	0.64	0.61	0.62	0.63	0.60
20% coinsurance	0.11	0.10	0.05	0.11	0.11	0.11	0.10	0.08
Monthly expenditure (USD)	1,341	254	439	777	1,138	1,794	1,993	1,877
Obs. (Recipient × Term)	12,237	112	365	1,877	5,031	2,794	1,412	646
		B. Care level 3 in the previous term						
Age	82.6	79.9	79.9	81.3	81.0	82.6	83.9	84.1
Woman	0.61	0.66	0.74	0.65	0.56	0.60	0.64	0.62
20% coinsurance	0.11	0.09	0.08	0.12	0.13	0.11	0.09	0.10
Monthly expenditure (USD)	1,801	259	477	827	1,154	1,875	2,263	2,436
Obs. (Recipient × Term)	7,111	58	119	587	988	2,967	1,590	802
		C. Care level 4 in the previous term						
Age	82.3	80.7	81.3	80.4	81.6	81.8	82.6	83.2
Woman	0.63	0.48	0.64	0.62	0.61	0.59	0.64	0.65
20% coinsurance	0.10	0.18	0.13	0.11	0.13	0.10	0.09	0.09
Monthly expenditure (USD)	2,024	247	472	825	1.159	1,775	2,124	2,620
Obs. (Recipient × Term)	5,064	33	94	443	501	698	2,108	1,187
		D. Care level 5 in the previous term						
Age	81.4	75.1	80.6	80.7	80.7	81.2	81.4	81.4
Woman	0.63	0.44	0.50	0.49	0.51	0.56	0.61	0.67
20% coinsurance	0.09	0.33	0.12	0.12	0.13	0.11	0.11	0.07
Monthly expenditure (USD)	2,460	325	365	741	1,140	1,720	2,323	2,787
Obs. (Recipient × Term)	2,804	9	26	102	152	169	414	1,932

Table 4.4: The Impact of Manipulation on Long-Term Care Expenditure (Wide Manipulable Range)

	Care-needs level in the previous term				
	Care level 2 (1)	Care level 3 (2)	Care level 4 (3)	Care level 5 (4)	Total (5)
	A. Mean monthly expenditure per recipient (USD)				
Lower bound	1,225	1,625	1,651	1,965	1,485
Upper bound	1,302	1,738	1,944	2,246	1,632
Observed	1,333	1,790	2,010	2,361	1,684
	B. Mean annual expenditure per recipient (USD)				
Lower bound	14,700	19,500	19,812	23,580	17,820
Upper bound	15,624	20,856	23,328	26,952	19,584
Observed	15,998	21,481	24,122	28,328	20,213
	C. Annual increase in expenditure per recipient (USD)				
Lower bound	374	625	794	1376	629
Upper bound	1,298	1,981	4,310	4748	2,393

Table 4.5: The Impact of Manipulation on Long-Term Care Expenditure (Narrow Manipulable Range)

	Care-needs level in the previous term				
	Care level 2 (1)	Care level 3 (2)	Care level 4 (3)	Care level 5 (4)	Total (5)
	A. Mean monthly expenditure per recipient (USD)				
Lower bound	1,225	1,640	1,847	2,144	1,544
Upper bound	1,302	1,731	1,907	2,323	1,631
Observed	1,333	1,790	2,010	2,361	1,684
	B. Mean annual expenditure per recipient (USD)				
Lower bound	14,700	19,680	22,164	25,728	18,528
Upper bound	15,624	20,772	22,884	27,876	19,572
Observed	15,998	21,481	24,122	28,328	20,213
	C. Annual increase in expenditure per recipient (USD)				
Lower bound	374	709	1,238	452	658
Upper bound	1,298	1,801	1,958	2,600	1,702

Appendix for Chapter 4: Proofs

Assumption 1. *(i) $\mathcal{F}_{IS}$ is nonempty and $\mathcal{F}$ is compact in the $L^1(\mathcal{X})$ space. (ii) $\mathcal{F}_n \subset \mathcal{F}$, and there exists a map $\Pi_n : \mathcal{F} \mapsto \mathcal{F}_n$ such that $\sup_{f\in\mathcal{F}} \|f - \Pi_n(f)\|_1 = O(n^{-1/2})$. (iii) We have $\sup_{x\in\mathcal{X}} |g(x)| \leq C$ for some $C > 0$ and $\mathcal{X}$ is compact. (iv) We have $\sup_{x\in\mathcal{X}} |\hat{g}(x) - g(x)| = o_p(1)$. (v) $\kappa_n \to \infty$ and $\kappa_n/\sqrt{n} \to 0$.*

Theorem 1. *Under Assumption 1, we have $\hat{\overline{\theta^*}} \to_p \overline{\theta^*}$ and $\underline{\hat{\theta^*}} \to_p \underline{\theta^*}$.*

Proof. We define $L(f) \equiv \int_{\mathcal{X}} g(x)f(x)dx$ and $\Gamma(F) \equiv w\int_K |F(x)|dx + (1-w)\int_{K^C} |F(x)|_+ dx$. Then, we have

$$\begin{aligned}\hat{Q}_n(f) &= \Gamma(\hat{F}_n - F_f),\\ Q(f) &= \Gamma(F_X - F_f).\end{aligned}$$

By the definition of $\Gamma(\cdot)$, we have $\Gamma(F + F') \leq \Gamma(F) + \Gamma(F')$.

First, we observe that

$$\begin{aligned}&P\left(\Pi_n(\mathcal{F}_{IS}) \subset \hat{\mathcal{F}}_{IS}\right)\\ =\;& P\left(\sup_{f\in\mathcal{F}_{IS}} \sqrt{n}\hat{Q}_n(\Pi_n(f)) \leq \inf_{f\in\mathcal{F}_n} \sqrt{n}\hat{Q}_n(f) + \kappa_n\right)\\ \geq\;& P\left(\sup_{f\in\mathcal{F}_{IS}} \sqrt{n}\Gamma\left((\hat{F}_n - F_X) + (F_X - F_f) + (F_f - F_{\Pi_n(f)})\right) \leq \kappa_n\right)\\ \geq\;& P\left(\Gamma(\sqrt{n}(\hat{F}_n - F_X)) + \sqrt{n}\sup_{f\in\mathcal{F}_{IS}} \Gamma(F_f - F_{\Pi_n(f)}) \leq \kappa_n\right),\end{aligned}$$

where the last inequality holds because we have $\Gamma(F_X - F_f) = 0$ for $f \in \mathcal{F}_{IS}$. Because $\hat{F}_n(x)$ is an empirical distribution and $\mathcal{X}$ is compact, we have $\Gamma(\sqrt{n}(\hat{F}_n - F_X)) = O_p(1)$. Hence, it follows from Lemma 1 and Assumption 1 (v) that we have

$$\liminf_{n\to\infty} P\left(\Pi_n(\mathcal{F}_{IS}) \subset \hat{\mathcal{F}}_{IS}\right) = 1.$$

This implies that

$$\begin{aligned}
\overline{\theta^*} &= \sup_{f\in\mathcal{F}_{IS}} L(f) \;=\; \sup_{f\in\mathcal{F}_{IS}} L(\Pi_n(f)) + o(1) \\
&\leq \sup_{f\in\hat{\mathcal{F}}_{IS}} L(f) + o_p(1), \qquad (4A.1)
\end{aligned}$$

where the second equality follows from Lemma 1. Similarly, we have

$$\underline{\theta^*} \geq \inf_{f\in\hat{\mathcal{F}}_{IS}} L(f) + o_p(1). \qquad (4A.2)$$

By the property of $\Gamma(\cdot)$, we have

$$\begin{aligned}
\sup_{f\in\hat{\mathcal{F}}_{IS}} Q(f) &= n^{-1/2} \sup_{f\in\hat{\mathcal{F}}_{IS}} \sqrt{n}\Gamma\left((F_X - \hat{F}_n) + (\hat{F}_n - F_f)\right) \\
&\leq n^{-1/2}\Gamma\left(-\sqrt{n}(\hat{F}_n - F_X)\right) + n^{-1/2} \sup_{f\in\hat{\mathcal{F}}_{IS}} \sqrt{n}\hat{Q}_n(f) \\
&\leq n^{-1/2}\Gamma\left(-\sqrt{n}(\hat{F}_n - F_X)\right) + n^{-1/2} \inf_{f\in\mathcal{F}_n} \Gamma\left(\sqrt{n}(\hat{F}_n - F_f)\right) + \kappa_n/\sqrt{n} \\
&= n^{-1/2} \inf_{f\in\mathcal{F}_n} \Gamma\left(\sqrt{n}(\hat{F}_n - F_f)\right) + o_p(1).
\end{aligned}$$

Because $\Pi_n(\mathcal{F}_{IS}) \subset \mathcal{F}_n$, we obtain

$$\begin{aligned}
& n^{-1/2} \inf_{f\in\mathcal{F}_n} \Gamma\left(\sqrt{n}(\hat{F}_n - F_f)\right) \\
\leq\; & n^{-1/2} \inf_{f\in\mathcal{F}_{IS}} \Gamma\left(\sqrt{n}(\hat{F}_n - F_{\Pi_n(f)})\right) \\
\leq\; & n^{-1/2}\Gamma\left(\sqrt{n}(\hat{F}_n - F_X)\right) + \inf_{f\in\mathcal{F}_{IS}} \Gamma\left(F_f - F_{\Pi_n(f)}\right) \;=\; o_p(1),
\end{aligned}$$

where the last inequality holds because we have $\Gamma(F_X - F_f) = 0$ for $f \in \mathcal{F}_{IS}$. Therefore, we obtain

$$\sup_{f\in\hat{\mathcal{F}}_{IS}} Q(f) = o_p(1). \qquad (4A.3)$$

For any $\delta > 0$, we define $\mathcal{F}_{IS}(\delta) \equiv \{f \in \mathcal{F} : Q(f) \leq \delta\}$. It follows from (4A.3) that there exists a positive sequence $\delta_n \downarrow 0$ such that

$$\liminf_{n\to\infty} P\left(\hat{F}_{IS} \subset \mathcal{F}_{IS}(\delta_n)\right) = 1. \qquad (4A.4)$$

We view $\delta \mapsto \mathcal{F}_{IS}(\delta)$ as a correspondence, and we denote its graph by

$$\mathrm{Gr}(\mathcal{F}_{IS}(\cdot)) \equiv \{(f, \delta) \in \mathcal{F} \times \mathbb{R}_+ : Q(f) \leq \delta\}.$$

By Lemma 1, $\mathrm{Gr}(\mathcal{F}_{IS}(\cdot))$ is closed in $L^1(\mathcal{X}) \times \mathbb{R}_+$. Because $\mathcal{F}$ is compact, it follows from Theorem 17.11 in Aliprantis and Border (2006) that $\delta \mapsto \mathcal{F}_{IS}(\delta)$ is upper hemicontinuous. Let $\tilde{L} : L^1(\mathcal{X}) \mapsto \mathbb{R}$ be a continuous functional. Then, by $\mathcal{F}_{IS}(0) = \mathcal{F}_{IS}$ and Lemma 17.30 in Aliprantis and Border (2006), we have

$$\limsup_{n\to\infty} \sup_{f\in\mathcal{F}_{IS}(\delta_n)} \tilde{L}(f) \leq \sup_{f\in\mathcal{F}_{IS}} \tilde{L}(f).$$

Because $\mathcal{F}_{IS} \subset \mathcal{F}_{IS}(\delta_n)$ for all n, we obtain

$$\sup_{f\in\mathcal{F}_{IS}} \tilde{L}(f) \leq \liminf_{n\to\infty} \sup_{f\in\mathcal{F}_{IS}(\delta_n)} \tilde{L}(f).$$

Hence, we obtain

$$\lim_{n\to\infty} \sup_{f\in\mathcal{F}_{IS}(\delta_n)} \tilde{L}(f) = \sup_{f\in\mathcal{F}_{IS}} \tilde{L}(f). \tag{4A.5}$$

Because equation (4A.5) holds with $\tilde{L} = L$ and $\tilde{L} = -L$, it follows from (4A.4) that

$$\begin{aligned}
\overline{\theta^*} &= \sup_{f\in\mathcal{F}_{IS}} L(f) = \lim_{n\to\infty} \sup_{f\in\mathcal{F}_{IS}(\delta_n)} L(f) \geq \sup_{f\in\hat{\mathcal{F}}_{IS}} L(f) + o_p(1), \\
\underline{\theta^*} &= \inf_{f\in\mathcal{F}_{IS}} L(f) = -\lim_{n\to\infty} \sup_{f\in\mathcal{F}_{IS}(\delta_n)} -L(f) \leq \inf_{f\in\hat{\mathcal{F}}_{IS}} L(f) + o_p(1).
\end{aligned}$$

Combined with (4A.1) and (4A.2), we obtain

$$\begin{aligned}
\overline{\theta^*} &= \sup_{f\in\hat{\mathcal{F}}_{IS}} L(f) + o_p(1), \\
\underline{\theta^*} &= \inf_{f\in\hat{\mathcal{F}}_{IS}} L(f) + o_p(1).
\end{aligned} \tag{4A.6}$$

By Assumption 1 (iv), we have

$$\begin{aligned}\sup_{f\in\mathcal{F}}\left|\int_{\mathcal{X}}\hat{g}(x)f(x)dx - L(f)\right| &\leq \int_{\mathcal{X}}|\hat{g}(x)-g(x)|f(x)dx \\ &\leq \sup_{x\in\mathcal{X}}|\hat{g}(x)-g(x)| = o_p(1).\end{aligned}$$

This implies that

$$\begin{aligned}\hat{\overline{\theta^*}} &= \sup_{f\in\hat{\mathcal{F}}_{IS}}\int_{\mathcal{X}}\hat{g}(x)f(x)dx = \sup_{f\in\hat{\mathcal{F}}_{IS}} L(f)+o_p(1), \\ \hat{\underline{\theta^*}} &= \inf_{f\in\hat{\mathcal{F}}_{IS}}\int_{\mathcal{X}}\hat{g}(x)f(x)dx = \inf_{f\in\hat{\mathcal{F}}_{IS}} L(f)+o_p(1).\end{aligned}$$

Hence, it follows from (4A.6) that $\hat{\overline{\theta^*}} = \overline{\theta^*}+o_p(1)$ and $\hat{\underline{\theta^*}} = \underline{\theta^*}+o_p(1)$. □

Lemma 1. *Under Assumption 1, we have* $\sup_{f\in\mathcal{F}}|L(f)-L(\Pi_n(f))| = O(n^{-1/2})$ *and* $\sup_{f\in\mathcal{F}}\Gamma\left(F_f - F_{\Pi_n(f)}\right) = O(n^{-1/2})$.

Proof. Because $g(x)$ is bounded, we have

$$\begin{aligned}|L(f)-L(\Pi_n(f))| &\leq \int_{\mathcal{X}}|g(x)|\cdot|f(u)-\Pi_n(f)(u)|\,du \\ &\leq C\|f-\Pi_n(f)\|_1.\end{aligned}$$

Hence, we obtain $\sup_{f\in\mathcal{F}}|L(f)-L(\Pi_n(f))| = O(n^{-1/2})$ from Assumption 1 (ii).

Next, we show that $\sup_{f\in\mathcal{F}}\Gamma\left(F_f - F_{\Pi_n(f)}\right) = O(n^{-1/2})$. We observe that

$$\begin{aligned}\sup_x|F_f(x)-F_{\Pi_n(f)}(x)| &= \sup_{f\in\mathcal{F}}\sup_x\left|\int_{-\infty}^{x}(f(u)-\Pi_n(f)(u))du\right| \\ &\leq \int|f(u)-\Pi_n(f)(u)|\,du \\ &= \|f-\Pi_n(f)\|_1.\end{aligned}$$

By the definition of Γ, for any $f \in \mathcal{F}$, we have

$$\begin{aligned} & \Gamma\left(F_f - F_{\Pi_n(f)}\right) \\ = \ & w \int_{K \cap \mathcal{X}} \left|F_f(x) - F_{\Pi_n(f)}(x)\right| dx + (1-w) \int_{K^C \cap \mathcal{X}} \left|F_f(x) - F_{\Pi_n(f)}(x)\right|_+ dx \\ \leq \ & \int_{\mathcal{X}} \left|F_f(x) - F_{\Pi_n(f)}(x)\right| dx. \end{aligned}$$

Because $\mathcal{X}$ is compact, there exists $C_\Gamma > 0$ such that

$$\Gamma\left(F_f - F_{\Pi_n(f)}\right) \leq C_\Gamma \times \sup_x |F_{\Pi_n(f)}(x) - F_f(x)|.$$

Hence, we obtain $\sup_{f \in \mathcal{F}} \Gamma\left(F_f - F_{\Pi_n(f)}\right) \leq C_\Gamma \times \sup_{f \in \mathcal{F}} \|f - \Pi_n(f)\|_1 = O(n^{-1/2})$. $\square$

Lemma 2. *Under Assumption 1, the set*

$$\mathcal{A} \equiv \{(f, \delta) \in \mathcal{F} \times \mathbb{R}_+ : Q(f) \leq \delta\}$$

is closed in $L^1(\mathcal{X}) \times \mathbb{R}_+$.

Proof. We show that $L^1(\mathcal{X}) \times \mathbb{R}_+ \setminus \mathcal{A}$ is open with respect to the product topology. To this end take any $(\tilde{f}, \tilde{\delta}) \in L^1(\mathcal{X}) \times \mathbb{R}_+ \setminus \mathcal{A}$.

Case 1: Suppose that $\tilde{f} \notin \mathcal{F}$. Because $\mathcal{F}$ is compact, $L^1(\mathcal{X}) \setminus \mathcal{F}$ is open. Hence, there exists an open neighborhood $\mathcal{N}(\tilde{f})$ of $\tilde{f}$ such that $f \notin \mathcal{F}$ for any $f \in \mathcal{N}(\tilde{f})$. Let $\mathcal{N}(\tilde{\delta})$ be an open neighborhood of $\tilde{\delta}$. Then, by the definition of $\mathcal{A}$, we have $\mathcal{N}(\tilde{f}) \times \mathcal{N}(\tilde{\delta}) \subset L^1(\mathcal{X}) \times \mathbb{R}_+ \setminus \mathcal{A}$.

Case 2: Suppose that $\tilde{f} \in \mathcal{F}$. Then, we have $Q(\tilde{f}) > \tilde{\delta}$. Hence, there exists $\eta > 0$ such that $Q(\tilde{f}) > \tilde{\delta} + \eta$. We define the set

$$\mathcal{O}(\tilde{f}) \equiv \left\{f \in L^1(\mathcal{X}) : Q(f) > \tilde{\delta} + \eta\right\}.$$

We show that $\mathcal{O}(\tilde{f})$ is open. For any $f, f' \in L^1(\mathcal{X})$, we have

$$\begin{aligned} \|F_f - F_{f'}\|_1 &\leq \int_{\mathcal{X}} \int_{-\infty}^{x} |f(u) - f'(u)| du dx \\ &\leq C_{\mathcal{X}} \|f - f'\|_1, \end{aligned}$$

where the last inequality follows from the compactness of $\mathcal{X}$. Hence, $f \mapsto F_f$ is a continuous map from $L^1(\mathcal{X})$ to $L^1(\mathcal{X})$. This implies that $Q : L^1(\mathcal{X}) \mapsto \mathbb{R}$ is continuous. Hence, the set $\mathcal{O}(\tilde{f})$ is open. We define

$$\mathcal{O} \equiv \mathcal{O}(\tilde{f}) \times ((\tilde{\delta} - \eta, \tilde{\delta} + \eta) \cap \mathbb{R}_+),$$

which is open with respect to the product topology. If $(f, \delta) \in \mathcal{O}$, then we have $Q(f) > \tilde{\delta} + \eta > \delta$. Therefore, we obtain $\mathcal{O} \subset L^1(\mathcal{X}) \times \mathbb{R}_+ \setminus \mathcal{A}$. $\square$