Panel Data Research Center, Keio University

PDRC Discussion Paper Series

The Impact of Daycare Closures Owing to COVID-19 on Parental Stress: A Case of Japan

Michio Naoi, Tsubasa Ito, Kazuto Sumita, Qing Ye

22 September, 2021

DP2021-005
https://www.pdrc.keio.ac.jp/en/publications/dp/7388/

Panel Data Research Center, Keio University
2-15-45 Mita, Minato-ku, Tokyo 108-8345, Japan
info@pdrc.keio.ac.jp
22 September, 2021

The Impact of Daycare Closures Owing to COVID-19 on Parental Stress: A Case of Japan

Michio Naoi, Tsubasa Ito, Kazuto Sumita, Qing Ye

PDRC Keio DP2021-005

22 September, 2021

JEL Classification: I13; I18; C90

Keywords: Parents' mental health; State of emergency; COVID-19; Doubly robust difference-in-differences method

Abstract

This study aims to estimate the change in stress levels within households that had children under 6 years owing to the closure of daycare facilities in the wake of the state emergency declared in Japan in April 2020 to fight the COVID-19 crisis. Doubly robust difference-in-differences method is used based on household panel data. The results show that the stress of parents significantly increased when their children were forced to stay at home as daycare facilities closed. Another analysis also reveals that leaving children in daycare facilities during this period significantly reduced the stress of parents.

Michio Naoi
Faculty of Economics, Keio University
2-15-45 Mita, Minato-ku, Tokyo
naoi@econ.keio.ac.jp

Tsubasa Ito
Graduate School of Economics, Toyo University
5-28-20 Hakusan, Bunkyo-ku, Tokyo
tsubasa.itoh0917@gmail.com

Kazuto Sumita
Faculty of Economics, Toyo University
5-28-20 Hakusan, Bunkyo-ku, Tokyo
sumita@toyo.jp

Qing Ye
Graduate School of Economics, Toyo University
5-28-20 Hakusan, Bunkyo-ku, Tokyo
yeqing0127@gmail.com

Acknowledgement: The data for this analysis, the Japan Household Panel Survey (JHPS/KHPS) and the supplementary modules on COVID-19, were provided by the Keio University Panel Data Research Center, for which we gratefully acknowledge. This work was supported by JSPS KAKENHI Grant Numbers JP17H06086 and JP17H00988.

The Impact of Daycare Closures Owing to COVID-19 on Parental Stress: A Case of Japan *

Tsubasa Ito† Michio Naoi‡ Kazuto Sumita§ Qing Ye¶

September 12, 2021

Abstract

This study aims to estimate the change in stress levels within households that had children under 6 years owing to the closure of daycare facilities in the wake of the state emergency declared in Japan in April 2020 to fight the COVID-19 crisis. Doubly robust difference-in-differences method is used based on household panel data. The results show that the stress of parents significantly increased when their children were forced to stay at home as daycare facilities closed. Another analysis also reveals that leaving children in daycare facilities during this period significantly reduced the stress of parents.

Keywords: Parents' mental health, state of emergency, COVID-19, doubly robust difference-in-differences method

JEL Codes: I13, I18, C90

---

*This draft is preliminary and incomplete. Please do not cite without explicit permission. We welcome any comments and suggestions. The data for this analysis, the Japan Household Panel Survey (JHPS/KHPS), were provided by the Keio University Panel Data Research Center.

†Graduate School of Economics, Toyo University

‡Faculty of Economics, Keio University

§Faculty of Economics, Toyo University

¶Graduate School of Economics, Toyo University

1 Introduction

The COVID-19 pandemic led to governments enforcing the closure of daycare and elementary schools worldwide. Japan was no exception. Elementary and junior high schools were temporarily closed after "the first state of emergency" was declared from April 17 to May 31, 2020. However, daycare facilities were not asked to close their doors but were free to take their own initiative depending on the prefecture and the municipalities.[1] Against the backdrop of this situation, Takaku and Yokoyama (2021) examined the differences in parental stress between households with children under 6 years, who were able to attend daycare facilities continuously before and after the first state of emergency, and households with children over 7 years, who were not able to attend elementary school.

To begin with, however, we need to realize that there is a big difference in the need for parental support for children under 6 years and children over 7 years. For example, the former can attend daycare facilities almost all-year round without any long breaks,[2] whereas the latter have long breaks while attending elementary school and stay at home during these breaks. This implies that the burden on parents is different for households with children under 6 years and those with children over 7 years.

In households with children aged 6, it has been found that sending the children to daycare facilities reduces stress on parents, especially on mothers. Yamaguchi et al. (2018) noted that when children attend daycare facilities, the quality of childcare and subjective well-being of the parents improve. The study showed that not only children but also their parents benefit from daycare facilities. Therefore, when the first state of emergency during the COVID-19 crisis was declared, severe stress would have affected the population. In such a situation, if children aged below 6 continued to attend daycare facilities, parents would have found it more beneficial or less stressful than in normal times.

This paper examines the impact of the closure of daycare facilities after the first state of emergency, focusing only on parents with children aged under 6. First, considering all households that have at least one child aged under 6, we examine whether daycare closures during the first state of emergency had any impact on parental stress. In this exercise, we compare households in which childcare arrangement was affected by daycare closures with

[1] In Japan, preschoolers can attend nurseries or kindergartens. A nursery school takes care of children who need childcare as both parents are working or for some reason are not able to care for the child in the daytime. Children up to 6 years of age are eligible to join a nursery. Kindergartens are facilities designed to educate children between the ages of 3 and school-age, and their responsibility does not include caring for children of working parents. In this study, unless otherwise mentioned, nursery schools and kindergartens are collectively referred to as daycare facilities.

[2] Kindergartens are an exception. They have similar breaks as elementary schools. However, some kindergartens do accept children during long breaks.

those that were not. The former includes households experienced daycare closures. The latter includes two types of households: those with children attending daycare facilities but experienced no daycare closures, and those with children not attending daycare facilities even before the first state of emergency. This will allow us to examine the impact of temporary changes in childcare arrangement during the first state of emergency. Second, we further restrict our sample to households with children who attended daycare facilities before the first state of emergency and examine whether temporary daycare closures had any impact on parental stress. We examine the differences in parental stress among those whose children continued to attend daycare facilities during the first state of emergency and those whose children could not do so.

In this study, we use the "Japan Household Panel Survey" (JHPS) and it's supplementary module on COVID-19 (hereafter referred to as COVID-19 Supplement). The JHPS is conducted every February since 2004 by the Panel Data Research Center at Keio University under the name of Keio Households Panel Survey and continues even now. COVID-19 Supplement is a supplementary survey on COVID-19 conducted by the Center in May 2020. It surveyed the same subjects as the JHPS did, adding specific questions related to COVID-19 and the state of the emergency. Among these questions, it asked households with children under 6 years whether their children are attending daycare facilities and whether the centers were closed in April. We used these questions to examine whether there was any difference in the aggregate Kessler 6 (K6) scale of the households depending on the closure of the daycare facilities.

We face two problems when we estimate treatment effects caused by the state of emergency declared during the COVID-19 crisis. First, the closure of daycare facilities was not necessarily random because, in contrast to the temporary closure of elementary schools, this was voluntary. In other words, if there were households that send their children to daycare facilities that are likely to be closed during such a crisis and households that do not send their children, the estimation would be subject to pre-treatment bias. Therefore, we discuss the direction of the pre-treatment bias by estimating average treatment effect (ATE) by linear regression (LR) and inverse probability weighting (IPW). Second, K6 scale changes over time influenced by the COVID-19 situation, and hence, the change in K6 scale cannot be attributed purely to the daycare facilities' closure. Nevertheless, the common shocks assumption is satisfied because the shocks due to the state of emergency are likely to have affected all households. Therefore, we can estimate the effect of daycare facilities' closure on K6 scale using the doubly robust difference-in-differences (DR-DID) method, which is a combination of the IPW and DID methods. Furthermore, the DID allows us to identify the average treatment effect on the treated (ATT),[3] where daycare facilities' closure status after the first state of emergency is the average stress change in the treatment group. This finding could have policy implications.

This paper is structured as follows: the next section briefly reviews previous studies. Section 3 explains the data set used, section 4 explains the method of analysis, and section 5 presents the results. Section 6 concludes.

2 Review of Related Literature

2.1 Recent studies on childcare during COVID-19

In Japan, Takaku and Yokoyama (2021) and Yokoyama and Takaku (2020) analyzed data of 15,836 people sampled through a unique Internet survey, and used a natural experiment on closure of public elementary school closures and voluntary closure of daycare facilities during the COVID-19 crisis to estimate the impact of closure of public elementary schools nationwide. Takaku and Yokoyama (2021) conducted regression discontinuity design (RDD) analysis to examine the impact of school absence during the crisis on children and mothers whose oldest child was aged between 4 and 10. Their results showed that COVID-19-related withdrawal led to weight gain in children and increased parental anxiety. However, no significant effects of the crisis were seen on marital relationships or domestic violence. Yokoyama and Takaku (2020) focused on the stress (K6 scale) of parents with children aged 4 to 6 years to clarify how closure of daycare facilities affected them. They found that since daycare facilities were not completely closed, and parents could choose to send their children to these facilities, may be a factor affecting mothers' stress. To deal with this issue, they examined the relationship between attendance at daycare facilities and parental stress, using closure of daycare facilities as an instrumental variable (IV). They found that mothers' stress increased when their children did not attend daycare facilities.

As a study from other countries, Huebener et al. (2021) used DID to examine how the total closure of schools and daycare facilities services during the lockdown in Germany affected parents' well-being. Specifically, households with children attending school or daycare facilities services were the treatment group and the rest formed the control group. In their analysis, the well-being of these parents was identified before and after the COVID-19 crisis, showing that the well-being of parents, especially mothers, decreased significantly when schools and daycare facilities services were closed during the lockdown. In another research, Wu et al. (2020) conducted an Internet survey of households with children in elementary

[3] Athey and Imbens (2006) discusses how DID estimates will become the ATT.

school through college in China to investigate stress during the COVID-19 crisis and found that parents, especially mothers, of elementary school-aged children felt more stress. The COVID-19 crisis had increased parents' stress and child abuse at home (Brown et al., 2020, Griffith, 2020, Lawson et al., 2020). In addition to stress, Alon et al. (2020) points out that COVID-19 had a significant impact on women's participation in the workforce compared with mothers who did not use childcare services.

2.2 Studies on the impact of daycare facilities

Many studies have also shown that attending daycare facilities and receiving early childhood education improves the cognitive and non-cognitive abilities of children, especially disadvantaged children (e.g., Heckman et al., 2013). Yamaguchi et al. (2018) conducted an analysis on Japan's Longitudinal Survey of Newborns in the 21st Century (LSN21) to estimate how childcare enrollment affects children and their parents. They found that children of less-educated mothers developed better language skills and showed reduced inattention, hyperactivity, and aggression when they attended daycare facilities. Besides, daycare facilities also improved the quality of childcare and the subjective well-being of children's parents. In particular, they found that it reduced the stress of less-educated mothers. They estimated the impact of daycare facilities on children and their parents using a two-period DID method, with the enrollment quota of daycare facilities in a particular region as the instrumental variable. They estimated marginal treatment effects (MTE) of the impact of unobserved trends in daycare facilities use. The results showed that mothers with higher levels of education and work skills are more likely to use daycare facilities and they benefit the most from daycare facilities use.

3 Data

3.1 Survey

The data we analyzed are drawn from the "Japan Household Panel Survey" (JHPS) and its supplementary module on COVID-19 (COVID-19 Supplement), conducted in February and May 2020, respectively. The JHPS was originally conducted as two independent household surveys—the Keio Household Panel Survey (KHPS) and the former-Japan Household Panel Survey (former-JHPS). The KHPS began in 2004 surveying 4,005 households. The former-JHPS began in 2009 with an initial sample of 4,022 households. Since 2015, these two surveys have been merged and continue to survey covering general topics, including employment, education, lifestyle, time allocation, health, and living environment, as well as more detailed subjects, such as the composition of respondents' households, their income, expenditures, assets, and housing.

In the wake of the COVID-19 crisis, the COVID-19 Supplement was conducted as an additional survey to the JHPS sample in May 2020. A total of 5,470 JHPS respondents were asked to participate in the COVID-19 supplement, resulting in a response rate of 70.5% (N = 3,857). The COVID-19 Supplement aimed to understand the situation under the emergency, asking specific questions related to the COVID crisis. In the following analysis, we used the sample from the most recent JHPS conducted in February 2020 and its COVID-19 Supplement conducted in mid-May to analyze the impact of COVID-19 measures on households. These two surveys provides us data on the household situation before and after the first state of emergency and its resulting daycare closures.

3.2 Variables

3.2.1 Treatment variables

In the COVID-19 Supplement of the JHPS, households with children aged 6 or younger, as of April 2020, were asked about their childcare situation.[4] They were categorized as: 1) never attended daycare facilities, 2) attended daycare facilities without having any closures (including day care only), 3) refrained from going to daycare facilities even though the centers were not closed, 4) did not attend daycare facilities and are still not attending daycare facilities due to closure (including day care only), and 5) did not attend daycare facilities for a specific period, but are now attending the daycare facilities again (including day care only).[5]

We used the responses to these questions and defined the treatment and control groups. Our first set of treatment and control definition compares households whose children did not attend daycare facilities due to daycare closures (treatment group) and those whose children never attended daycare facilities or kept going to daycare facilities without any daycare closures (control group). Based on the responses of the above question, the treatment indicator is set to 1 for households that answered 4) or 5), and 0 for households that answered 1) or 2). In the following analysis, we do not use respondents who answered 3) in the above question due to endogeneity concerns, that is, the decision to not send their kids to daycare facilities might be endogenous.

A potential problem in our first definition is that the control group includes children

[4]However, if a household has more than one child aged 6 or younger, the response regarding the oldest child is the only one considered.

[5]These are households whose with children whose daycare facilities were closed as of April but reopened as of May.

who never attended daycare facilities and those who continued going to daycare facilities without any daycare closures. While these two groups of children did not experience any changes in their childcare situation during the first state of emergency, there might be some inherent differences between them. In the second set of treatment and control definition, we omit households whose children never attended daycare facilities and restrict our sample to households whose children had been in daycare facilities before the first state of emergency. For the sake of interpretation, we flip the treatment definition from the first set. As a result, our treatment indicator is set to 1 for households that answered 2) and 0 for households that answered 4) or 5) to the above question. Again, households that answered 3) is excluded because of endogeneity concerns.

3.2.2 Outcome variables

Our key outcome variable is parental stress measured by the Kessler Psychological Distress Scale (Kessler et al., 2002). K6 scale is a 6-item self-report measure of psychological distress to assess individual risk of serious mental disorders such as depression and anxiety. The K6 scale is constructed from the 6 different survey items about feelings or experiences during the past 30 days.[6] These 6 items ask respondents to rate how often they felt (1) nervous, (2) hopeless, (3) restless or fidgety, (4) so depressed that nothing could cheer them up, (5) that everything was an effort, and (6) that everything was worthless, in the past 30 days. The six survey items were all answered on a 5-point Likert scale: 0 (never), 1 (for a small period of time), 2 (for some time), 3 (for most of the time), and 4 (all the time). Responses to these 6-items were summed up to yield a K6 scale, with a total score of more than 10 indicating a high probability of worsening mental health. In this study, the total K6 scale is used as an outcome.

Here, we look at the distributions of K6 scale in January and April. Figure 1 shows K6 scale for males and females in January and April, which is the full sample in the JHPS and COVID-19 Supplement. Panel (a) is the distribution of K6 scale for males and panel (b) is the distribution of K6 scale for females. Looking at the K6 scale in January, we see that both K6 scales are higher and are skewed to the right. As far as the distribution of K6 scale in April is concerned, both K6 scales become lower than that of January and have a thicker right tail, especially in females. Here again, we see that K6 scale is worse for females than for males. From this, it may be inferred that although the COVID-19 crisis causes stress in both males and females, the stress was more on females than on males.

[6]We define the K6 scale of JHPS as that of January 2020, because the most recent JHPS was conducted in February 2020. The COVID-19 Supplement defines the K6 scale as of early April to early May (roughly April), because the survey was conducted in mid-May.

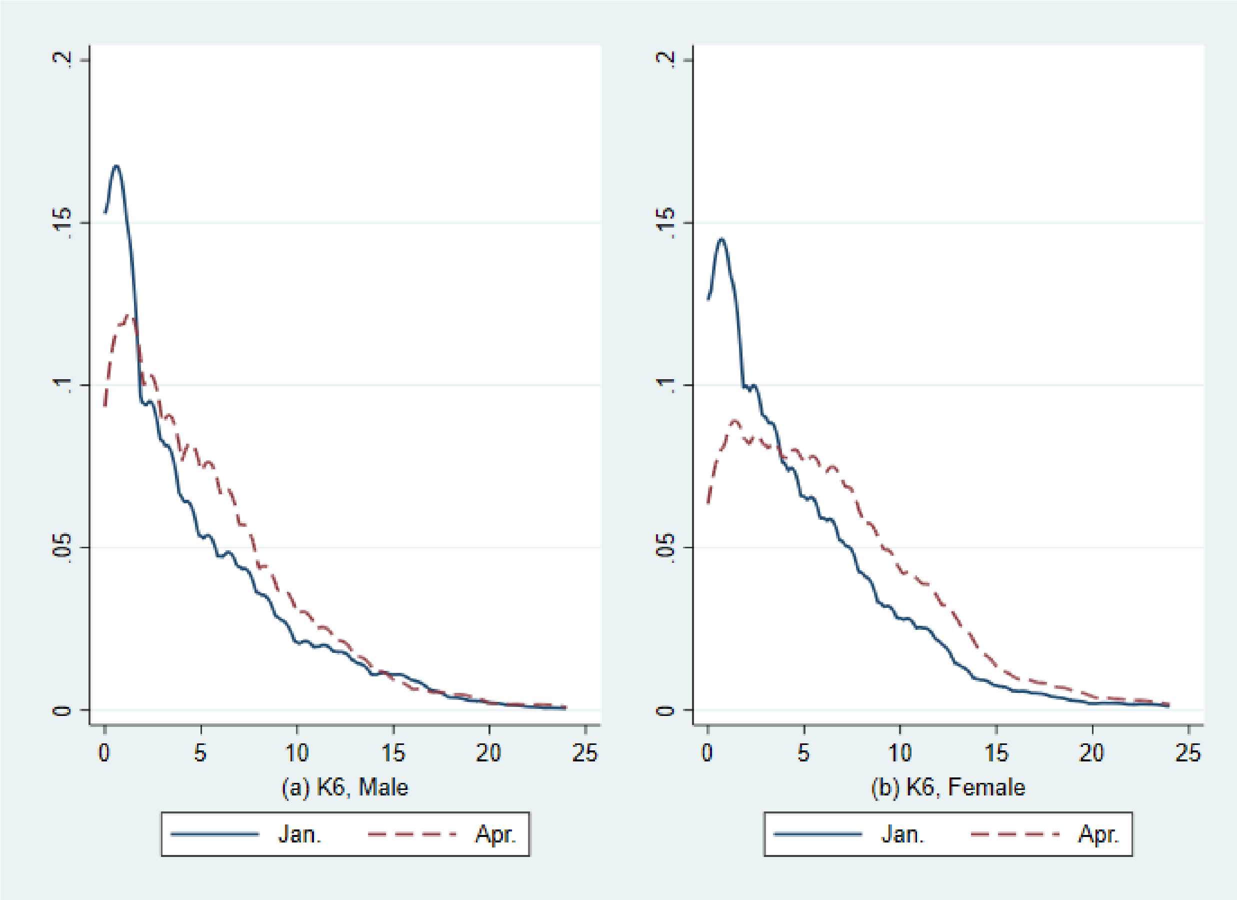
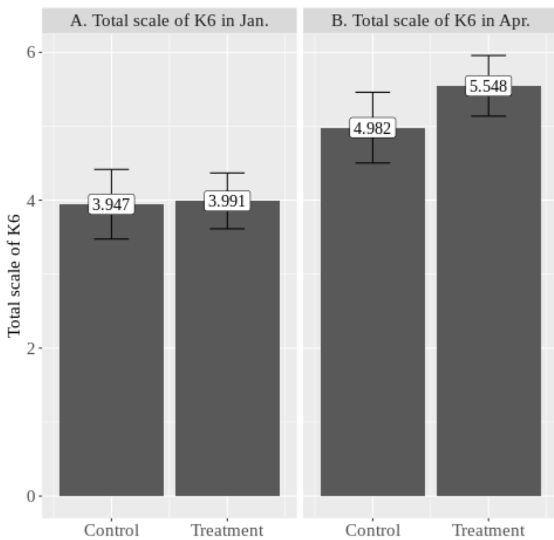
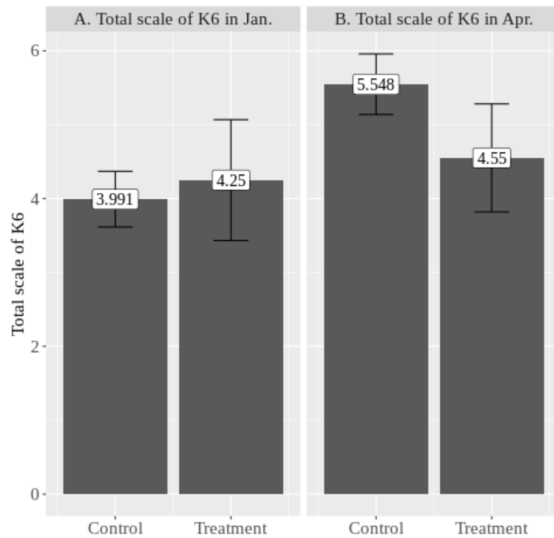


Figure 1: Distributions of K6 scale in January and April

Further, we examine the difference in K6 scale in the treatment groups above. Figure 2a depicts the difference in the treatment groups: that do not attend daycare facilities due to closures, and Figure 2b depicts the difference in the treatment group that: attends daycare facilities that did not have any closures. In both cases, we can see that COVID-19 crisis worsens the K6 scale in both the treatments and controls. However, no difference in K6 scale between treatment and control group is seen in January before the COVID-19 crisis. In Figure 2a, we see that the mean of K6 scale of the treatment groups is higher than that of the control groups after the COVID-19 crisis. Further, the K6 scale is worse for the group that does not attend daycare facilities due to closures. In Figure 2b, the difference between the treatment and control group is larger than that in Figure 2a. Furthermore, the group attending the daycare facilities showed less increase on the K6 scale after the COVID-19 crisis. The group that could not attend daycare facilities showed a larger and highly increased K6 scale.



(a) Not attend daycare due to its closures



(b) Attending daycare facilities without closures

Figure 2: Difference in K6 scale from treatment v.s. control

3.2.3 Other variables

In addition, the following variables are used as control variables: age group dummies of respondents, female dummy, college graduate dummy, dual-income dummy, children's age group dummies, household members dummies, variables related to nursing care and income, three major metropolitan areas (TMA), ordinance-designated cities (ODC), and core city dummies. The age group dummies are dummy variables set to 1 if the respondent is 30 to 34, 35 to 39, 40 to 44, or 45 years or older, and the reference is 30 to 34 years old. We use three dummies for dual-income households: (a) dual-income households where both individuals are employed full-time, (b) dual-income households where only one individual is employed full-time, and (c) dual-income households where both individuals are not employed full-time. The age group dummies for children are dummy variables that are set to 1 if the child is 0 to 6, 7 to 12, 13 to 15, or 16 years or older, and the reference is a household with a child between 0 and 6 years. For the household members, using 1 to 3 as a reference group, we use two dummy variables indicating 4 to 6, and 7 and above. For caregiving, we use a dummy variable that takes the value 1 when there is at least one hour of caregiving in a week. Income is the logarithmic value of household income. To identify the composition of the household members, we include a dummy for living with the respondent's parent and a dummy for living with the respondent's spouse's parent. In the doubly robust difference-in-differences used in this paper, the covariates for propensity score estimation and outcome regression need to be aligned (For see Sant'Anna and Zhao, 2020). Considering that the treatment group is one that was formed due to the COVID-19 crisis, the propensity score needs to be calculated using covariates before the COVID-19 crisis. Therefore, all covariates used in the analysis were obtained from the JHPS (as of January), which is before the COVID-19 crisis.[7] These descriptive statistics are summarized in the Table 1.

Table 1: Descriptive Statistics

	Treatment variation			
	Treatment1: Not attend daycare due to its closures		Treatment2: Attending daycare facilities without closures	
Variables	mean	std. dev	mean	std. dev
Full sample				
Dep. var				
Kessler 6 scale (January)	3.969	4.535	4.058	4.346
Kessler 6 scale (April)	5.254	4.756	5.290	4.462
Treatment	0.504	0.501	0.258	0.439
Observations	228		155	
Estimated sample				
Dep. var				
Kessler 6 scale (January)	3.732	4.237	4.023	4.320
Kessler 6 scale (April)	4.874	4.435	5.075	4.274
Treatment	0.521	0.501	0.256	0.438
Controls var.				
under 29 years (reference)	0.068	0.253	0.053	0.224
30 to 34 years	0.295	0.457	0.256	0.438
35 to 39 years	0.384	0.488	0.421	0.496
40 to 44 years	0.179	0.384	0.211	0.409
over 45 years	0.074	0.262	0.060	0.239
female dummy	0.516	0.501	0.504	0.502
college dummy	0.526	0.501	0.519	0.502
double-income of full-time employment	0.258	0.439	0.301	0.460
double-income with only one full-time employment	0.284	0.452	0.308	0.464
double-income outside of regular employment	0.258	0.439	0.301	0.460
children 0 to 6 years old (reference)	0.605	0.490	0.586	0.494
children 7 to 12 years old	0.332	0.472	0.361	0.482
children 13 to 15 years old	0.026	0.160	0.030	0.171
children over 16 years old	0.037	0.189	0.023	0.149
number of people in household (1 to 3) (reference)	0.316	0.466	0.203	0.404
number of people in household (4 to 6)	0.611	0.489	0.714	0.453
number of people in household (over 7)	0.074	0.262	0.083	0.276
number of parents living with respondents	0.042	0.269	0.045	0.271
number of parents of spouse living with respondents	0.079	0.355	0.075	0.340
income	46.221	48.864	49.180	57.328
care dummy	0.042	0.201	0.053	0.224
three metropolitan areas (TMA)	0.663	0.474	0.699	0.460
ordinance-designated city (ODC)	0.368	0.484	0.376	0.486
core city	0.568	0.497	0.564	0.498
Observations	190		133	

[7] However, the stress of parents who have children may be affected by the situation after the COVID-19 crisis. For example, if there is an infected person in their neighborhood, they have to be careful not to pass the disease to their children. Therefore, they are likely to be sensitive to the infection situation in their own area. Therefore, in Appendix A, we include the COVID-19 crisis variable in the linear regression and inverse probability weighting to perform a simple robustness check.

4 Method of Analysis

4.1 Identification strategy

The basic estimation model for the empirical analysis is eq.(1).

$$Y_i = \alpha + \tau D_i + \theta' X_i + \varepsilon_i. \tag{1}$$

Here, we will clarify how the closure of daycare facilities during the COVID-19 crisis affects parental K6 scale. Y is the total K6 scale for parents with children, D is the treatment variable, X is the control variable, and ε is the error term. Specifically, we will compare the difference in total K6 scale between treatments and controls as shown in eq.(2).

$$\tau = E[Y|D = 1] - E[Y|D = 0]. \tag{2}$$

However, it is conceivable that τ may have some biases due to the following issues. We sort these issues and consider how to deal with each of the biases.

4.1.1 Pre-treatment bias

If treatment assignment is not random and depends on individual attributes and other factors, the estimates of treatment effect will be biased. Such a bias is called pre-treatment bias. In other words, when a treatment group is a unit of children whose daycare facilities are closed due to the initial declaration of a state of emergency, if parents tend to enroll their children in daycare facilities that are likely to be closed in the first place, pre-treatment bias will occur. In Japan, enrollment in daycare facilities (excluding kindergartens) prioritizes households in which both parents are employed full-time. Therefore, households in which both parents are in full-time employment are more likely to choose daycare facilities. On the contrary, if one or both parents are not in regular employment (non-regular employment or self-employed), it becomes difficult for them to choose and enroll their child in daycare facilities that cater to their needs. These households may be able to cope with sudden closures of daycare facilities if both parents work more flexibly than those in full-time employment. Therefore, even if there is an absence of daycare facilities after the first state of emergency, there may be a downward bias in the estimates because the stress load is not high.

Parents may also enroll their children in daycare facilities that are more likely to be closed. Such a tendency is an unobservable confounding factor (U), and a pre-treatment bias is applied; because the usual linear regression cannot identify τ (Figure3a). Therefore, in addition to the usual LR, we use propensity score (PS) to address pre-treatment bias and to discuss the direction of bias. As an estimation method using propensity scores, inverse probability weighting (IPW) is used in this paper. Here, PS includes IPW as well as propensity score matching (PSM); the IPW is used in this study instead of PSM in view of the criticism of PSM by Smith and Todd (2005) and King and Nielsen (2019),[8] and others, claimed that the following problems exist in estimating using the PS method. IPW is a method of weighting an outcome variable by the inverse of the propensity score estimate. Propensity scores are estimated by logistic regression of the covariates on the treatment variables. All covariates are pre-treatment.

$$\pi(X_i; \gamma) = \Pr(D_{it} = 1 | X_{it-1}; \gamma). \tag{3}$$

The propensity score values in eq.(3) are used to weight the outcome variables. Since the IPW is used to weight the inverse of the propensity score and to adjust the balance between the treatment and control groups, the average treatment effect on the treated (ATT) is

$$\tau_i^{ipw} = \frac{1}{n}\sum_{i=1}^{n} \frac{D_i Y_i}{\hat{\pi}(X_i; \hat{\gamma})} - \frac{1}{n}\sum_{i=1}^{n} \frac{(1 - D_i) Y_i}{1 - \hat{\pi}(X_i; \hat{\gamma})}. \tag{4}$$

The treatment group is weighted $1/\hat{\pi}(X_i; \hat{\gamma})$, and the control group is weighted $1/(1 - \hat{\pi}(X_i; \hat{\gamma}))$. That is to say, the inverse of the probability that an individual is assigned to treatment is weighted to the treatment group, thus balancing the covariates in the treatment and control. The effect of the covariates X is controlled for by mediating the propensity score $\pi(X)$. This means that if controlling X makes the treatment unconfounded, then controlling $\pi(X)$ makes the treatment unconfounded as well (Rosenbaum and Rubin, 1983). Figure 3b is an intuitive illustration of how D reflects the causal effect on Y.

$$\begin{tikzcd}
X \arrow[r] & D \arrow[r, "\tau"] & Y \\
& U \arrow[ul] \arrow[ur] &
\end{tikzcd}$$

(a) Pre-treatment bias

$$\begin{tikzcd}
D \arrow[rr, "\tau^{ipw}"] & & Y \\
& X \arrow[ul, "\pi(X)"] \arrow[ur] &
\end{tikzcd}$$

(b) Propensity score

Figure 3: Directed acyclic graph of pre-treatment bias

[8]Smith and Todd (2005) and King and Nielsen (2019) point out that in PSM, the propensity score is highly sensitive to the balance of covariates in matching and the choice of data, and that bias remains/can remain in the estimates. Smith and Todd (2005) showed that combining PSM with DID can remove bias compared to the cross-section. However, PSM is based on a fairly strong assumption that the model for the propensity score was correctly specified (Wooldridge, 2007).

4.1.2 COVID-19 crisis shock

In estimating the impact of the closure of daycare facilities on parental stress in the post COVID-19 stage and the first state of emergency, we need to consider the possibility of exacerbated stress on many parents. Parental stress would include the effect of the crisis, not just the effect due to closure of daycare facilities. In other words, the COVID-19 crisis would affect treatment and control groups equally, conditional on observable characteristics (including geographic location and individual characteristics). This satisfies the common shock assumption, and hence, we use the difference in differences (DID) method to remove the COVID-19 shock from the estimates. In other words, in addition to weighting adjustment to deal with pre-treatment bias, we combine it with the DID method to remove the COVID-19 shock to estimate the impact of the closure of the daycare facilities on parental stress.

Here is a brief description of a simple DID. The first period refers to the period before the state of emergency was declared and the second period is the one after. We denote, Y_{it} as the outcome of individual i in period t. Suppose the first state of emergency is declared during these two periods, and some households experience the treatment of daycare facilities closure. If individual i is exposed to the treatment in period t, we denote it as $D_{it} = 1$, otherwise $D_{it} = 0$. Let $t = 0$ before the first state of emergency and $t = 1$ after the first state of emergency. Let individual i be $D_{i0} = 0$ because no one is exposed to the treatment at $t = 0$. Here, we discuss the impact of daycare facilities on parental stress, the ATT,

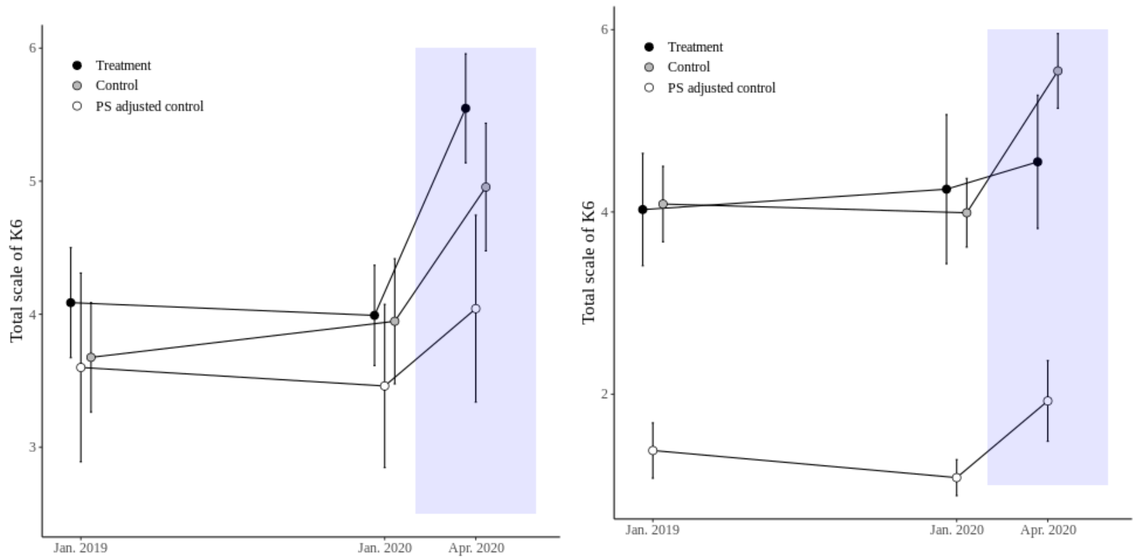
$$\tau^{did} = E[Y_{i1}(1) - Y_{i1}(0)|D_i = 1]. \quad (5)$$

Where $Y_{it}(0)$ is the outcome of the control and $Y_{it}(1)$ is the outcome of the treatment. The regression model assumed is a two-way fixed linear regression model.

$$Y_{it} = \alpha_1 + \alpha_2 T_i + \alpha D_i + \tau^{did}(T_i \cdot D_i) + \theta' X_i + \varepsilon_{it}. \quad (6)$$

T is a time variable. This eq.(6) is based on Angrist and Pischke (2008), and τ^{did} can be interpreted estimated as ATT.

An important assumption, the parallel trends assumption, exists in the use of DID. Therefore, we will look at the verification of parallel trends and their adjustment. Figure 4 shows the trend of the mean value of total scale of K6 divided by treatment and control in JHPS and COVID-19 Supplement for 2019 and 2020. Black dots are treatments, gray dots are controlled. The range shaded in blue represents the post COVID-19 crisis. First, figure4a shows the mean change in the total scale of K6 for treatment (a), not attending daycare due to its closure. This shows that the treatment has a higher K6 scale of only 0.411



(a) Not attend daycare due to its closures

(b) Attending daycare facilities without closures

Figure 4: Validation of the parallel trends assumption and adjustment by propensity score

Notes: In panel (a), the sample size for the treatment and control groups is 115 and 111, respectively. Further, the propensity score adjusted control group is 90. In panel (b), the sample size for the treatment and control groups are 38 and 115. Also, the propensity score adjusted control group is 99. The whiskers above and below each point indicate the standard error.

as of January 2019, but by January 2020, the difference has almost disappeared because the mean total scale of K6 in the treatment group has fallen while that in the control group has risen. After the COVID-19 crisis, the total scale of K6 increased significantly in both groups, indicating that the COVID-19 crisis was a common shock. However, the total scale of K6 for the group that experienced the closure of the daycare facilities that served as the treatment increased significantly compared to the control group that did not experience closure. In light of the fact that the trends of the total scale of K6 for treatment and control are not parallel, we see that we deviated from the parallel trends assumption. Therefore, we adjust the control group by calculating the propensity score. The adjustment method is the second term in equation 4; the controls adjusted for propensity score are illustrated by the white dots in Figure.4a[9] This shows that the trend of the control group is adjusted to be parallel to that of the treatment group so that, the use of the propensity score does not deviate from the parallel trends assumption.

Next, looking at the Figure 4b treatment (b), attending daycare facilities without closures, there is no significant difference in the mean of the total scale of K6 as of January

[9]The results of the propensity score estimation can be seen in the Appendix B

2019. The treatment group shows a slight increase over January 2020, while the control group shows a decrease. However, the trend is slight and there is no significant difference in the trend. After the COVID-19 crisis, the total scale of K6 increased in both treatments; but the increase varied greatly, but the increase was more gradual in the treatments. The difference in total scale of K6 between treatment and control as of April 2020 was about 0.998. However, the average of the total K6 scale was much smaller for the controls adjusted for the propensity score, indicating that the trend was not adjusted for. From this, it is possible that the propensity score was not calculated well for the second treatment. As a result, the second treatment (b) attending daycare facilities without deviates from the parallel assumption in the first place, or we do not know whether it is appropriate to use the propensity score. In view of these problems, we use doubly robust difference-in-differences (DR-DID) to obtain a consistency estimator. DR-DID is able to obtain consistent estimators even if there is a mis-specification in either outcome regression or IPW.

4.1.3 Infection status in the region

We, we use PS and DID to estimate the effect of daycare facilities closure on a parental K6 scale in the COVID-19 crisis. We use pre-treatment covariates to estimate the probability of treatment assignment in PS. Since PS is used to address pre-treatment bias, the inclusion of covariates in the COVID-19 crisis would be inconsistent in estimating the probability of treatment assignment. However, since the closure of daycare facilities is voluntary, it is likely to be highly dependent on the awareness and attitude toward COVID-19 crisis and the infection status in each region; in other word, daycare facilities in areas with high awareness of the COVID-19 crisis, where people refrain from going out, may have been actively closed. In addition, although the shock of the COVID-19 crisis can be considered a common shock, the impact of the shock is likely to be different by regions. There may be a difference in the impact on K6 scale between residents in areas where the infection rates were high compared with areas that were relatively free of COVID-19. In such a case, even if PS and DID are used, it is not possible to estimate τ accurately.

Therefore, we address the above problem by including regional dummies. The Figure 5 shows the average rate of refraining from going out in January, February, and April for each prefecture.[10] This rate is quantified by using the real-time population distribution estimated from the information of 78 million base stations of NTT DOCOMO, INC. cell phones.[11] The

[10] However, considering that most schools and companies remained closed until 5th January, this is the average from the 6th to the 31st.

[11] This rate of refraining from going out is produced by Mizuno Laboratory (National Institute of Informatics). Rate of refraining for prefectures are published from January 1, 2020 to March 31, 2021, and are available for download for each date (as of May 2021). As for municipalities, they are published

Figure 5: Rate of refraining from going out by prefecture

Notes: The following table shows the average rate at which people refrained from going out in January, February, and April for each prefecture. This rate is based on the one prepared by the Mizuno Laboratory (http://research.nii.ac.jp/ mizuno/).

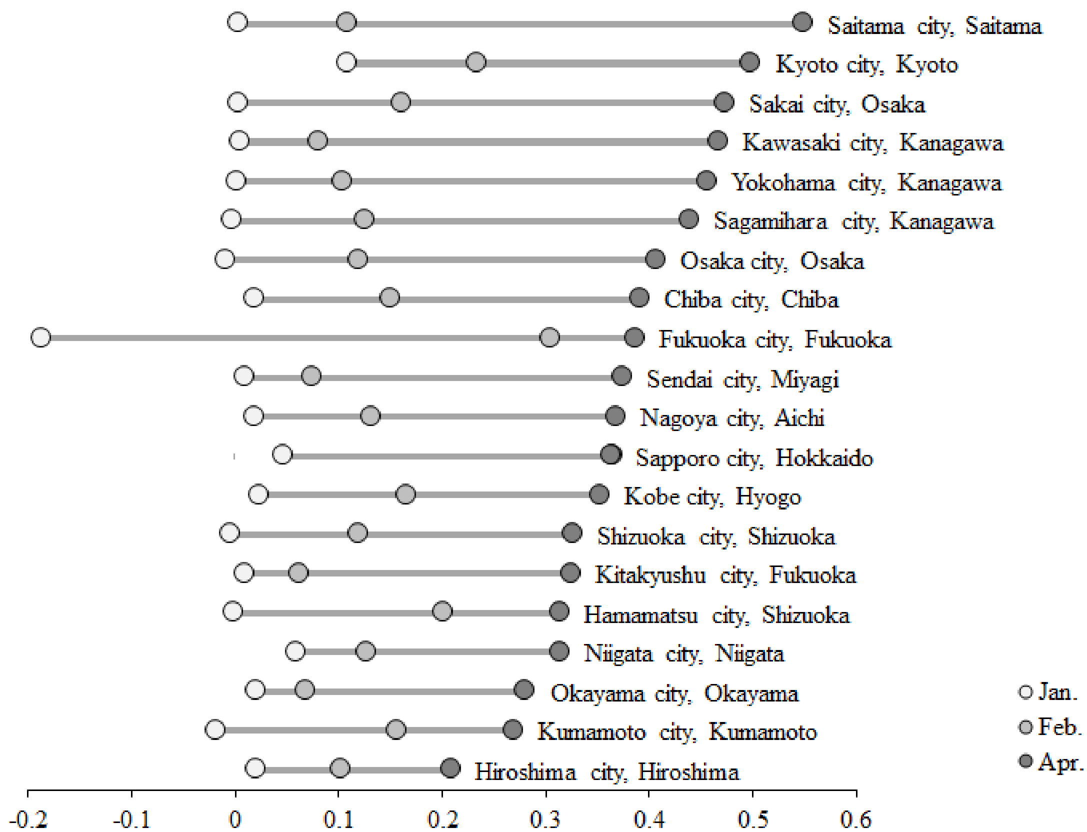


Figure 6: Rate of refraining from going out by ordinance-designated city

Notes: The following table shows the rate of refraining going out at the end of January, February, and April in ordinance-designated cities (ODC). This rate is based on the one prepared by the Mizuno Laboratory (http://research.nii.ac.jp/ mizuno/).

Figure 5 shows that the rate of refraining from going out is high in the Tokyo metropolitan area, including Tokyo, Kanagawa, Chiba, and Saitama prefectures, and in major cities such as Osaka and Hyogo prefectures. The Figure 6 also shows the refrain rate of going out at the end of January, February, and April in ordinance-designated cities (ODC). The ODC have a larger population than other cities, towns, and villages, and are given the same level of authority as prefectures by the government as compared to ordinary cities, towns, and villages. From this chart, we see that Saitama, Kyoto, Osaka, Kanagawa, Chiba, Fukuoka, and other ODC have the highest rate of refraining from going out. These cities have a high concentration of population and economic activity, to begin with. They also have a large number of children on waiting lists, which means that parents may not be able to easily enroll their children in the daycare facilities of their choice. Therefore, we believe that we can control the infection status of COVID-19 crisis to some extent by including a dummy variable that adds Fukuoka Prefecture to the three major metropolitan areas (Tokyo, Kanagawa, Chiba, Saitama, Aichi, Osaka, Hyogo, and Kyoto prefectures) and a cross term of dummies for ODC and core cities[12] in the estimation of the propensity score.

4.2 Doubly robust difference-in-differences

Based on the above, we correct pre-treatment bias by using the propensity score (PS) and COVID-19 shock by using difference in differences (DID) techniques. Combining PS and DID methods requires that (1) the model of treatment variables D and pre-treatment covariables X for calculating the PS is correctly specified, and that (2) the function of outcome regression is correctly specified. If either is mis-specified, no consistent estimator can be obtained. Therefore, implementing PS combined with the DID method does not always produce a robust result. To overcome such problems, Sant'Anna and Zhao (2020) proposed doubly robust difference-in-differences (DR-DID) estimators, where a consistent estimator is obtained if one of the above holds. DR-DID is an estimation equation using IPW combined with an outcome regression function, which can be written as follows. First, $\pi(X)$ is an arbitrary model for an unknown propensity score. In this study, we define $\Delta Y = Y_1 - Y_0$ as the difference from January($t = 0$) to April($t = 1$), and $\mu_{d,\Delta}(X) \equiv \mu_{d,1} - \mu_{d,0}$, where $\mu_{d,t}(x)$ as the model for the true unknown outcome regression $m_{d,t}(x) \equiv E[Y_t|D = d, X = x], d, t = 0, 1$.

from January 6, 2020 to March 24, 2021 (as of May 2021). For more information, see Mizuno Laboratory (http://research.nii.ac.jp/ mizuno/).

[12]A core city is a city that is not larger in population or economic scale than an ODC, but is larger than other municipalities.

Based on this, we calculate the estimand as follows:

$$\hat{\tau}^{dr\text{-}did} = E_n\left[\left(\hat{w}_1(D) - \hat{w}_0(D, X; \hat{\gamma})\right)\left(\Delta Y - \mu_{0,\Delta}(X; \hat{\beta}_{0,0}, \hat{\beta}_{0,1})\right)\right], \tag{7}$$

where

$$\hat{w}_1(D) = \frac{D}{E_n[D]}, \quad \text{and} \quad \hat{w}_0(D, X; \hat{\gamma}) = \frac{\hat{\pi}(X; \hat{\gamma})(1-D)}{1-\hat{\pi}(X; \hat{\gamma})} \bigg/ E_n\left[\frac{\hat{\pi}(X; \hat{\gamma})(1-D)}{1-\hat{\pi}(X; \hat{\gamma})}\right]. \tag{8}$$

The moment equation in eq.(7) illustrates a two-step strategy for estimating the ATT. The first step is to estimate the true unknown $p(\cdot)$ by estimating $\pi(\cdot)$ and the true unknown $m_{d,t}(\cdot)$ by $\mu_{d,t}(\cdot)$, $d, t = 0, 1$. In the second step, the estimand propensity scores and the fitting values of the regression model are plugged into the sample analogue of $\tau^{dr\text{-}did}$, $\pi(x; \hat{\gamma})$ in $\hat{w}_0$ is an estimated parametric model for $p(\cdot)$. Where $\hat{\gamma}$ is the estimator for the pseudo-true parameter γ^*. Additionally, $\mu_{d,t}(x; \beta^*_{d,t})$ is the parametric model of $m_{d,t}(x)$ for finite dimensional pseudo-true $\beta^*_{d,t}$, $t = 0, 1$, and for a generic β_0 and β_1, $\mu_{0,\Delta}(x; \beta_0, \beta_1) = \mu_{0,1}(x; \beta_1) - \mu_{0,0}(x; \beta_0)$. $\hat{\beta}$ is the estimator for $\beta^*_{d,t}$.

For $\hat{\tau}^{dr\text{-}did}$ to converge to $\tau^{dr\text{-}did}$, the following three conditions must be fulfilled: (1) data are independent and identically distributed, (2) conditional outcomes for treatment and control are parallel in the absence of treatment, and (3) the propensity score support for treatment is a subset of the propensity score support for control.[13] When these three assumptions hold, if either (but not necessarily both) $\pi(X) = p(X)$ almost surely or $\mu_\Delta(X) = m_{0,1}(X) - m_{0,0}(X)$ almost surely, then $\tau^{dr\text{-}did}$ is a consistent estimator of τ (For specific proofs, see Sant'Anna and Zhao, 2020.).

5 Result

We estimated the effect of the treatment on the parental stress in households whose children's school attendance changed due to the closure of schools after the first state of emergency. Columns (1) to (3) of Table 2 show the estimation results of treatment effect from linear regressions. Column (1) shows the estimate of treatment only, and column (2) shows the estimate of treatment effect including control variables. Compared to column (1), estimate of treatment effect in column (2) is larger and statistically significant. Next, column (3) shows the results of using IPW to address the pre-treatment bias: the estimated value of

[13] Assumptions (2) and (3) are standard and important assumptions in the conditional DID method (Heckman et al., 1997 and Abadie, 2005). Assumption (3), in particular, is to identify the average impact on treatments under selection by covariates (Heckman et al., 1997).

treatment is 1.817, which is larger than that in column (2) in the linear regression and is statistically significant at the 1% level of significance. From this, we may infer that the pre-treatment bias of the estimate hangs downward. Finally, in column (4), using DR-DID to remove the effect of common shocks, the estimated value is smaller than that in column (2) in the linear regression and that in column (3) in the IPW.[14] In other words, the upward bias in column (3) of the IPW was due to the increased stress during COVID-19. However, the estimate in column (4) of DR-DID is 1.127, which is significant at the 10% level.

Table 2: Comparison of estimated values: Treatment not attending daycare due to closures.

	LR		IPW	DR-DID
	(1)	(2)	(3)	(4)
Treatment:	0.592	1.541**	1.817***	1.127*
Not attending daycare due to closures	(0.631)	(0.674)	(0.698)	(0.641)
Period of variables				
Dependent var.	Post	Post	Post	Pre & Post
Control var.	-	Pre	-	Pre
Included variables				
Control var.	No	Yes	No	Yes
Num.Obs.	228	190	190	190
se type	HC1	HC1	HC1	Inf.Func.
Estimands	ATE	ATE	ATT	ATT

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Columns (1) and (2) are linear regression, column(3) is inverse probability weighting, and column (4) is a doubly robust difference in differences. In parentheses, column (1) to (3) are robust standard errors and column (4) is standard error obtained from the influence function. Outcomes are different only in column (4). In the IPW, only the treatment dummy was included in the regression to the total K6 scale because the control variables were adjusted for in the calculation of the propensity score.

Next, we considered households that were able to attend daycare facilities after the first state of emergency as the treatment group and households that were closed as the control group (Table 3). Columns (1) and (2) are estimates by linear regressions, column (3) and (4) show those of IPW and DR-DID, respectively. From columns (1) to (2) of the linear regression, we see that only column (2) is statistically significant. Compared to column (2),

[14]In DR-DID, the standard error is estimated using an influence function based on Sant'Anna and Zhao (2020). Roughly speaking, an influence function quantifies how a statistic changes when the data increase or decrease slightly (Hampel, 1974). The standard error obtained from the influence function is known to asymptotically become a consistent estimator with the standard error of the statistic (Deville, 1999, Jann, 2019). It is also a powerful and flexible approach to calculating and applying multiple estimates to various treatment effect estimators (Jann, 2019, Jann, 2020).

the coefficient in column (3) is smaller and not statistically significant. Furthermore, in column (4) of DR-DID, the coefficient is even larger at -3.341, which is statistically significant at the 5% level of significance.

Table 3: Comparison of estimated values: Treatment attending daycare facilities without closures.

	LR		IPW	DR-DID
	(1)	(2)	(3)	(4)
Treatment:	-0.998	-2.380**	-1.773	-3.406***
Attending daycare facilities without closures	(0.835)	(0.982)	(1.170)	(1.509)
Period of variables				
Dependent var.	Post	Post	Post	Pre & Post
Control var.	-	Pre	-	Pre
Included variables				
Control var.	No	Yes	No	Yes
Num.Obs.	155	133	133	133
se type	HC1	HC1	HC1	Inf.Func.
Estimands	ATE	ATE	ATT	ATT

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Columns (1) and (2) are linear regression, column(3) is inverse probability weighting, and column (4) is a doubly robust difference in differences. In parentheses, column (1) to (3) are robust standard errors and column (4) is standard error obtained from the influence function. Outcomes are different only in column (4). In the IPW, only the treatment dummy was included in the regression to the total K6 scale because the control variables were adjusted for in the calculation of the propensity score.

6 Conclusion

In this study, we estimated the change in stress of households with children under 6 years due to the closure of daycare facilities after the first state emergency was announced to deal with the COVID-19 crisis in Japan. Specifically, we estimated the difference in stress between households where children had to stay at home and households where children could attend daycare facilities. The results showed that parental stress increased when their children were forced to stay at home. We also estimated the direction of bias by linear regression, IPW, and DR-DID. When parents have a choice in sending their children to daycare facilities that are more likely to close during a crisis than to centers that are less likely to close, the estimates are subject to pre-treatment bias. Comparing the linear regression and the IPW estimation results, we see a downward pre-treatment bias. This suggests that households with children in daycare facilities that close do not experience a large change in stress and are able to cope with sudden closures; this could be the reason for the downward bias. However, when DR-DID is used, the estimates become smaller, suggesting that the shocks of COVID-19 crisis and the first state of emergency declared give an upward bias to the estimate when only post-treatment data such as after COVID-19 crisis and the first state of emergency were used. Thus, parental stress in closing the daycare facilities in COVID-19 crisis was found to be about 23.1% higher than the mean value of total K6 scale in April (4.874, see descriptive statistics in the Table 1). This was consistent with Yokoyama and Takaku (2020), which also used total K6 scale.

Furthermore, in contrast to the above, we estimated the change in stress among households that were able to continue using daycare facilities after the first state of emergency and those that took a break from daycare facilities. We found that the use of daycare facilities after the declaration of the first state of emergency reduced the stress on parents. On the other hand, the IPW estimates were smaller and less statistically significant than the linear regression in column (2), and the Figure 4b shows that the PS adjusted control group was not well adjusted, suggesting that the IPW may well be mis-specified. In the Table 3, the sample is limited to those who have been using daycare facilities before the COVID-19 crisis, so there may be no pre-treatment bias. However, DR-DID yields consistent estimates even if there is a mis-specification in either the linear regression or the IPW. The DR-DID estimates in column (4) are larger in magnitude than those in column (2) and are significant at the 5% level of significance. We found that there was an upward bias in the estimates from DR-DID with the emergence of a common shock such as the COVID-19 crisis. The results of DR-DID considering COVID-19 shock revealed that households that were able to continue to use daycare facilities after the declaration of the first state of emergency had significantly lower stress than households whose daycare facilities were closed after the declaration of the first state of emergency. This is consistent with Yamaguchi et al. (2018), who found that daycare facilities use was beneficial not only for children but also for their parents.

In summary, we found that changes in childcare conditions due to the closure of daycare facilities in the first state of emergency sufficiently exacerbated parental stress. Furthermore, it was again shown that the use of daycare facilities has a positive impact on parents. Therefore, households with a child under elementary school age (6 years old or under) should ensure that there is no change in the childcare situation even in emergency situations. In the future, measures that do not close nursery schools even in emergencies such as the COVID-19 crisis will be required. It will also be necessary to examine how telecommuting affects the frequency and environment in which parents raise their children to clarify the mechanisms by which parental stress worsens.

References

Abadie, Alberto (2005) "Semiparametric difference-in-differences estimators," *The Review of Economic Studies*, Vol. 72, No. 1, pp. 1–19.

Alon, Titan M, Matthias Doepke, Jane Olmstead-Rumsey, and Michele Tertilt (2020) "The impact of COVID-19 on gender equality,"Technical report, National Bureau of economic research.

Angrist, Joshua D and Jörn-Steffen Pischke (2008) *Mostly harmless econometrics: An empiricist's companion*: Princeton university press.

Athey, Susan and Guido W Imbens (2006) "Identification and inference in nonlinear difference-in-differences models," *Econometrica*, Vol. 74, No. 2, pp. 431–497.

Brown, Samantha M, Jenalee R Doom, Stephanie Lechuga-Peña, Sarah Enos Watamura, and Tiffany Koppels (2020) "Stress and parenting during the global COVID-19 pandemic," *Child abuse & neglect*, Vol. 110, p. 104699.

Deville, Jean-Claude (1999) "Variance estimation for complex statistics and estimators: linearization and residual techniques," *Survey methodology*, Vol. 25, No. 2, pp. 193–204.

Griffith, Annette K (2020) "Parental burnout and child maltreatment during the COVID-19 pandemic," *Journal of family violence*, pp. 1–7.

Hampel, Frank R (1974) "The influence curve and its role in robust estimation," *Journal of the american statistical association*, Vol. 69, No. 346, pp. 383–393.

Heckman, James, Rodrigo Pinto, and Peter Savelyev (2013) "Understanding the mechanisms through which an influential early childhood program boosted adult outcomes," *American Economic Review*, Vol. 103, No. 6, pp. 2052–86.

Heckman, James J, Hidehiko Ichimura, and Petra E Todd (1997) "Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme," *The review of economic studies*, Vol. 64, No. 4, pp. 605–654.

Huebener, Mathias, Sevrin Waights, C Katharina Spiess, Nico A Siegel, and Gert G Wagner (2021) "Parental well-being in times of Covid-19 in Germany," *Review of Economics of the Household*, pp. 1–32.

Jann, Ben (2019) "Influence functions for linear regression (with an application to regression adjustment)," University of Bern Social Sciences Working Papers 32, University of Bern, Department of Social Sciences, URL: https://ideas.repec.org/p/bss/wpaper/32.html.

——— (2020) "Influence functions continued. A framework for estimating standard errors in reweighting, matching, and regression adjustment," University of Bern Social Sciences Working Papers 35, University of Bern, Department of Social Sciences, URL: https://ideas.repec.org/p/bss/wpaper/35.html.

Kessler, R C, Gavin Andrews, L J Colpe, Eva Hiripi, Mroczek D K, S L T Normand, E E Walters, and A M Zaslavsky (2002) "Short screening scales to monitor population prevalences and trends in non-specific psychological distress," *Psychological medicine*, Vol. 32, No. 6, pp. 959–976.

King, Gary and Richard Alexander Nielsen (2019) "Why propensity scores should not be used for matching," *Political Analysis*, Vol. 27, No. 4, pp. 1–20.

Lawson, Monica, Megan H Piel, and Michaela Simon (2020) "Child maltreatment during the COVID-19 pandemic: consequences of parental job loss on psychological and physical abuse towards children," *Child abuse & neglect*, p. 104709.

Rosenbaum, Paul R and Donald B Rubin (1983) "The central role of the propensity score in observational studies for causal effects," *Biometrika*, Vol. 70, No. 1, pp. 41–55.

Sant'Anna, Pedro H.C. and Jun Zhao (2020) "Doubly robust difference-in-differences estimators," *Journal of Econometrics*, Vol. 219, No. 1, pp. 101–122.

Smith, Jeffrey A and Petra E Todd (2005) "Does matching overcome LaLonde's critique of nonexperimental estimators?" *Journal of econometrics*, Vol. 125, No. 1-2, pp. 305–353.

Takaku, Reo and Izumi Yokoyama (2021) "What the COVID-19 school closure left in its wake: Evidence from a regression discontinuity analysis in Japan," *Journal of public economics*, Vol. 195, p. 104364.

Wooldridge, Jeffrey M. (2007) "Inverse probability weighted estimation for general missing data problems," *Journal of Econometrics*, Vol. 141, No. 2, pp. 1281-1301.

Wu, Mengting, Wenyan Xu, Yuhong Yao, Lei Zhang, Lei Guo, Juan Fan, and Jue Chen (2020) Mental health status of students' parents during COVID-19 pandemic and its influence factorsGeneral Psychiatry334

Yamaguchi, Shintaro, Yukiko Asai, and Ryo Kambayashi (2018) "How does early childcare enrollment affect children, parents, and their interactions?" *Labour Economics*, Vol. 55, pp. 56–71.

Yokoyama, Izumi and Reo Takaku (2020) "How Serious Was It? The Impact of Preschool Closure on Mothers' Psychological Distress: Evidence from the First COVID-19 Outbreak," SSRN 3756638, URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3756638.

A Robustness check: covariates after COVID-19 crisis

In this paper, the analysis is conducted using doubly robust difference-in-differences (DR-DID), considering issues related to the identification of inverse probability weighting (IPW) using linear regression and propensity scores. However, in DR-DID, the covariates for the estimation of propensity score and outcome regression should be the same. Therefore, we were not able to include them in the estimation of variables after the COVID-19 crisis (COVID-19 Supplement), which may have an impact on the stress of parents with children. In this Appendix, we check the robustness of the inclusion of variables after the COVID-19 crisis in the treatments, (a) not attending daycare due to closures and (b) attending daycare facilities without closures, on the parents' total scale of K6. In the robustness check, variables after the COVID-19 crisis are included in addition to the same variables as in Tables 2 and 3. In the IPW, the propensity score estimation is the same as in Tables 2 and 3, but variables after the COVID-19 crisis are included in the regression to the total scale of K6. The variables after the COVID-19 crisis are in Table A1.

Table A1: Descriptive statistics after COVID-19 crisis (at April)

	Treatment variation			
	Treatment1: Not attending daycare due to closures		Treatment2: Attending daycare facilities without closures	
Variables	mean	std. dev	mean	std. dev
number of new positive cases in prefectures	1051.842	1281.124	1124.947	1298.953
rate of self-restraint in prefectures	0.406	0.085	0.411	0.085
moving dummy	0.021	0.144	0.023	0.149
PCR-tested or wanted dummy	0.089	0.286	0.09	0.288
closed elementary and junior high school dummy	0.195	0.397	0.203	0.404
Observations	190		133	

The number of new positive cases in prefectures was created from the "Map of New Coronavirus Infections by Prefecture - Dashboard Map of COVID-19 Japan Case" provided by J.A.G JAPAN Corp.[15] The variables are included as logarithmic values. The rate of self-restraint in prefectures is the same as in Figure 5 and is based on data from the Mizuno Laboratory. In addition, moving dummy is a dummy variable which takes 1 if the respondent moved between January and April, and PCR-tested or wanted dummy is a dummy variable where 1 indicates the respondent that received or wanted to receive a PCR test. The closed elementary and junior high school dummy is a dummy variable with 1 being the respondent with children between the ages of 7 and 15 and closed elementary and junior high schools.

The estimation results are shown in Table A2. It can be seen that all the treatments

[15] However, data on the current infection status were not collected.

are significant and that the magnitude of the estimates is also larger than that of Table 2 and 3. Further, for IPW in column (4), the treatment, (b) attending daycare facilities without closures, is negative and significant at the 10% level of significance. Nevertheless, the robustness of the effect of treatment on the total scale of K6 was generally confirmed, supporting the results estimated in Table 2 and 3. However, all variables after the COVID-19 crisis, except for the treatment, were not confirmed to be statistically significant.

Table A2: Robustness check: Include covariates after COVID-19 crisis.

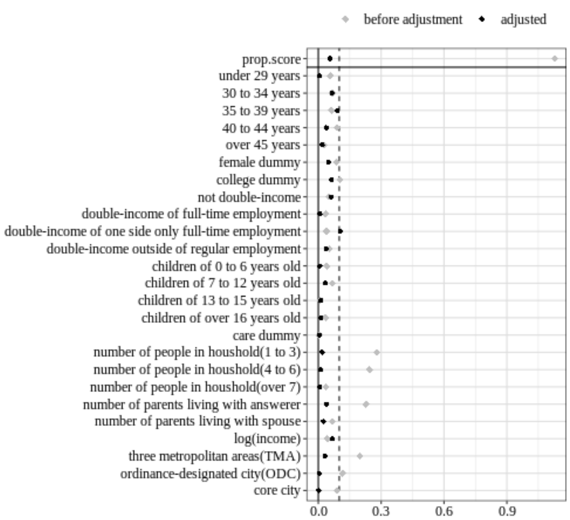
	Treatment variation			
	Treatment1: Not attend daycare due to closures		Treatment2: Attending daycare facilities without closures	
	LR	IPW	LR	IPW
	(1)	(2)	(3)	(4)
Treatment	1.664**	1.824***	-2.431**	-2.001*
	(0.666)	(0.669)	(0.976)	(1.155)
log(number of new positive cases in prefectures)	-0.379	0.025	-0.260	0.287
	(0.371)	(0.463)	(0.328)	(0.440)
rate of self-restraint in prefectures	5.264	-4.925	9.151	-6.361
	(7.325)	(8.511)	(7.951)	(12.948)
moving dummy	2.029	1.939	1.498	3.520
	(1.404)	(2.449)	(1.836)	(3.549)
PCR-tested or wanted dummy	2.236	1.281	2.713	3.323
	(1.363)	(1.467)	(1.844)	(2.097)
closed elementary and junior high school dummy	-1.412	-1.029	-1.415	1.420
	(0.865)	(0.777)	(1.094)	(1.639)
Period of variables				
Dependent var.	Post	Post	Post	Post
Control var.	Pre & Post	Post	Pre & Post	Post
Included variables				
Control var.	Yes	Yes	Yes	Yes
Num.Obs.	190	190	133	133
se type	HC1	HC1	HC1	HC1
Estimands	ATE	ATT	ATE	ATT

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Columns (1) and (3) denote linear regression, column(2) and (4) denote inverse probability weighting. In parentheses, column (1) to (4) are robust standard errors. In the IPW, only the treatment dummy was included in the regression to the total K6 scale because the control variables were adjusted for in the calculation of the propensity score. The covariates before the COVID-19 crisis for linear regression and the covariates for IPW propensity score estimation are the same in Tables 2 and 3.

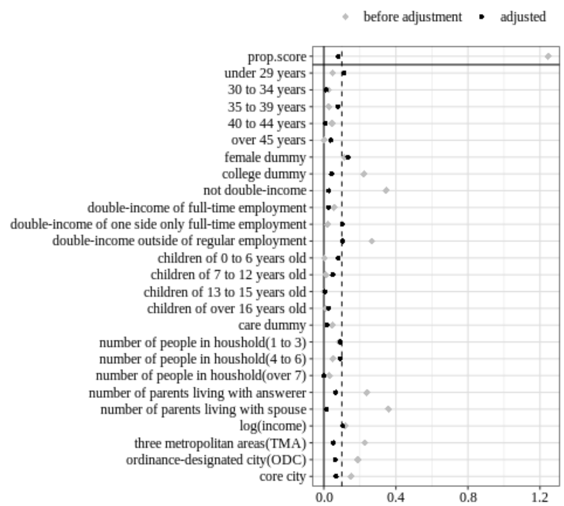
B Results of the propensity score estimation

For the balance of covariates based on the propensity score, see following Figure (Figure B1a, B1b).

Concerning the result of the propensity score estimation, see following Table B1.



(a) Not attending daycare due to closures



(b) Attending daycare facilities without closures

Figure B1: Balance of covariates

Table B1: Propensity score estimation

	Treatment variation	
	Treatment1: Not attending daycare due to closures	Treatment2: Attending daycare facilities without closures
	(1)	(2)
Constant	-3.596*	-0.832
	(2.011)	(3.068)
30 to 34 years	1.287	-1.692
	(1.420)	(2.203)
35 to 39 years	1.937	-2.423
	(1.399)	(2.182)
40 to 44 years	1.440	-1.422
	(1.454)	(2.225)
over 45 years	1.643	-1.104
	(1.627)	(2.437)
30 to 34 years×female dummy	-1.288	1.664
	(1.680)	(2.575)
35 to 39 years×female dummy	-2.444	2.913
	(1.672)	(2.507)
40 to 44 years×female dummy	-0.015	-0.941
	(1.800)	(2.822)
over 45 years×female dummy	-2.089	2.522
	(2.287)	(3.829)
female dummy	1.440	-1.388
	(1.582)	(2.382)
college dummy	0.461	-1.167**
	(0.377)	(0.592)
double-income of full-time employment	-1.065	2.493**
	(0.853)	(1.270)
double-income with only one full-time employment	-0.148	2.148**
	(0.446)	(0.921)
double-income outside of regular employment	-0.160	2.973***
	(0.462)	(0.936)
children 7 to 12 years old	-0.821*	0.177
	(0.473)	(0.679)
children 13 to 15 years old	-0.424	0.225
	(1.195)	(1.708)
children over 16 years old	-1.688	-0.448
	(1.334)	(2.944)
number of people in household(4 to 6)	1.810***	-0.145
	(0.454)	(0.714)
number of people in household(over 7)	2.533***	-2.747
	(0.970)	(2.019)
number of parents living with respondent	-1.533*	3.143*
	(0.830)	(1.622)
number of parents of spouse living with respondent	-0.111	-0.840
	(0.571)	(1.067)
log(income)	0.072	0.052
	(0.349)	(0.471)
care dummy	-0.345	0.610
	(0.865)	(1.159)
ordinance-designated city(ODC)	-0.117	2.132
	(1.249)	(2.260)
three metropolitan areas(TMA)	1.325	0.350
	(1.509)	(2.001)
core city	0.706	0.232
	(0.872)	(1.359)
TMA*ODC	0.105	-3.261
	(1.792)	(2.744)
TMA*core city	-1.053	-0.989
	(1.577)	(2.126)
Num.Obs.	190	133
AIC	273.8	163.1
BIC	364.8	244.1
Log.Lik.	-108.92	-53.564

Notes * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are reported in parentheses.