Journal of Public Economics 195 (2021) 104364

Contents lists available at ScienceDirect

Journal of Public Economics

journal homepage: www.elsevier.com/locate/jpube

What the COVID-19 school closure left in its wake: Evidence from a regression discontinuity analysis in Japan ☆

Reo Takaku, Izumi Yokoyama *

Graduate School of Economics, Hitotsubashi University, 2-1, Naka, Kunitachi, Tokyo 186-8601, Japan

ARTICLE INFO

Article history:
Received 15 September 2020
Revised 17 December 2020
Accepted 28 December 2020
Available online 8 January 2021

JEL classification:
I10
I20;D10

Keywords:
School closures
Fuzzy RDD
Marital relationship
Children's weight
Mothers' anxiety

ABSTRACT

To control the spread of COVID-19, the national government of Japan abruptly started the closure of elementary schools on March 2, 2020, but preschools were exempted from this nationwide school closure. Taking advantage of this natural experiment, we examined how the proactive closure of elementary schools affected various outcomes related to children and family well-being. To identify the causal effects of the school closure, we exploited the discontinuity in the probability of going to school at a certain threshold of age in months and conducted fuzzy regression discontinuity analyses. The data are from a large-scale online survey of mothers whose firstborn children were aged 4 to 10 years. The results revealed a large increase in children's weight and in mothers' anxiety over how to raise their children. On the outcomes related to marital relationships, such as the incidence of domestic violence and the quality of marriage, we did not find statistically significant changes. These findings together suggest that school closures could have large unintended detrimental effects on non-academic outcomes among children.

1. Introduction

During the COVID-19 pandemic in 2020, many countries closed schools in order to control infection. According to UNESCO (2020), more than 80% of students worldwide experienced school closures at the end of March 2020. Of course, there has been heated debate on the pros and cons of this policy. On the one hand, school closure has been seen as a natural policy response to the sudden outbreak of new respiratory diseases, since young children are extremely efficient at catching and passing them on, as has been found in the case of influenza (Cauchemez et al., 2008). However, opponents emphasize that children's education is severely disrupted and their mental health may suffer in countries with national school closures. Several medical studies also show that school closure is not an efficient way to control COVID-19 infection, because unlike influenza, children are not among the populations that suffer greatly from COVID-19 (Armbruster and Klotzbücher, 2020; Iwata et al., 2020).

Despite the importance of understanding the various consequences of past school closures in the current policy debate, there has not been a sufficient number of studies that directly reveal the effects of school closure on children and their families. So far, some studies have explored the effects of anti-COVID-19 policies on children's health outcomes and daily life, such as effects on obesity (Pietrobelli et al., 2020) and child maltreatment (Baron et al., 2020). Many studies also explore how lockdown affects the incidence of domestic violence (DV), as a prominent outcome that affects families (Piquero et al., 2020; Sanga and McCrary, 2020; Leslie and Wilson, 2020; Mohler et al., 2020; Campedelli et al., 2020; Payne and Morgan, 2020; Baron et al., 2020)[1].

Even with so much effort, it may be potentially impossible to identify the effects of school closures specifically in the countries

☆ This research is supported by HIAS (Hitotsubashi Institute for Advanced Study) and Grants-in-Aid for Scientific Research (18K12793, PI: Izumi Yokoyama). We thank Hideki Hashimoto and Naoki Kondo for granting permission to use J-SHINE data. We thank Arisa Shichijo for her outstanding research assistance. Also, we appreciate the comments from Michihito Ando. All errors are our own.

* Corresponding author.
E-mail addresses: reo.takaku@r.hit-u.ac.jp (R. Takaku), izumi.yokoyama@r.hit-u.ac.jp (I. Yokoyama).

[1] On the effects of the lockdown on mental health, some studies found negative effects soon after the implementation of lockdown. (Armbruster and Klotzbücher, 2020; Sibley et al., 2020). Scale (Kessler et al., 2002) before and after the lockdown in New Zealand and found a statistically significant increase in psychological distress.

https://doi.org/10.1016/j.jpubeco.2020.104364
0047-2727/

that enforced a lockdown, because, in these countries, schools were closed jointly with the implementation of numerous other kinds of anti-COVID-19 policies, including stay-at-home orders and business suspension orders. Therefore, it still remains a challenge for researchers to isolate the effects of "school closure" from those of other anti-COVID-19 policies. For example, many papers on the effects of COVID-19 policies compare the trend of outcomes between 2019 and 2020 (Brodeur et al., 2021; Leslie and Wilson, 2020), but this strategy may not be adequate for the evaluation of school closures, as other concurrent policies may contribute to changes in the behavior of children and their families entirely.

In contrast, utilizing the experience of school closure in Japan, this study successfully estimates the *pure* impacts of school closure on the well-being of families comprehensively, which is our study's largest contribution. This was made possible for the following two reasons: First, as we will see in Section 2, Japan is the rare country that experienced school closure without any heavy restrictions on daily life activities, which makes it possible to separate the effects of school closure from the effects of other anti-COVID-19 policies such as lockdown. Second, we utilize the prominent features of the Japanese school closure: all elementary schools were closed in March 2020, while preschools were exempted from this nationwide school closure. This enables us to compare the two groups of children and parents who faced totally different school closure situations even with only a very small difference in the timing of the children's birth. Due to this small difference in birth timing, these children and their families are likely to have similar characteristics and experiences of other anti-COVID-19 policies. Thus, by comparing these two groups, we can identify the *pure* impact of school closure.

Another contribution of our study is the data collection approach. We implemented a large-scale online survey in a timely manner. By doing so, our study can explore the impacts of school closures on comprehensive outcomes before the memories of potential respondents fade, which prevents measurement error in their answers. Furthermore, by creating an original questionnaire covering almost all the potential impacts of school closure on families, including novel and unique questions, we could obtain valuable findings and implications that would not have been obtainable from readily available public data (Leslie and Wilson, 2020; Baron et al., 2020).

Further, as we will see in the conclusion section, the results we have obtained yielded very important policy implications, which are also among our contributions.

In this study, we explore how a marginal difference in the timing of children's birth changed their experiences of school closure in March and eventually changed children's and families' outcomes, through a fuzzy regression discontinuity design (RDD) with an age-based threshold (Lee and Lemieux, 2010; Canaan, 2020). As results of our regressions, we gain several valuable findings, as follows: According to our fuzzy estimates of the impact of "non-schooling" due to school closure, the following two can be said to be the most conspicuous results: The fraction of mothers whose child(ren) gained weight rose by 14.4 to 15.4 percentage points, and mothers who worry over how to raise their children rose by 17.8 to 20.2 percentage points. Note that the magnitude of these numbers is non-negligible: these impacts are statistically significant even at a 1 % significance level. In contrast, we do not see any significant effect in other family outcomes, such as incidence of DV or quality of marriage index (Norton, 1983).

The remainder of this paper is constructed as follows: Section 2 offers an explanation of the natural experiment in Japan. Section 3 provides a description of the data and the main outcome variables. Section 4 explains the identification strategy and empirical methods. Section 5 reports the main results, and the results of the subsample analyses are presented in Section 6. Last, Section 7 concludes.

2. Background

As mentioned in Section 1, unlike in many other countries, in Japan, there have been no strict restrictions on daily life activities except for official requests to stay at home and not travel to other regions. This was due to the fact that COVID-19 did not spread rapidly at the time and the national government had no legal basis to implement a city-wide lockdown. Therefore, throughout February and early March in 2020, most economic activities went on as usual. However, following the rapid spread in nearby countries (China and South Korea), Prime Minister Shinzo Abe abruptly requested that all schools nationwide close as of March 2 (Cabinet Office, 2020), most likely in the hope that Japan would be able to host the Tokyo Olympics as planned (New York Times, 2020).

This sudden and unpredictable request for school closure is one of the most prominent features of Japan's anti-COVID-19 measures. In fact, Japan's nationwide closure was suddenly implemented despite the fact that the cumulative number of COVID-19 deaths was only three as of the day of the announcement. Because the school closure on March 2 was unexpected and implemented so abruptly, it caused substantial confusion to families. As a suggestive piece of evidence, we found a sharp increase in the number of Google searches for the word "divorce" on March 2—the first day of school closure—which is explained in Online Appendix A. This made us realize that it is necessary to more comprehensively and thoroughly investigate the impact of school closure on family well-being compared to existing studies because marital relationships can affect parents and children in many ways, which may include unexpected side effects.[2]

On the other hand, another surprising and sudden event occurred right after the announcement of the requested nationwide school closure by the prime minister: The Ministry of Health, Labour and Welfare announced that preschools were exempted from the nationwide school closure because of the potential impacts of the closure on working parents. Therefore, whether children were affected by school closure in March depended on children's school grades. Specifically, given the school grade system in Japan, the first graders born in March, who are the youngest within the same school grade, experienced school closure in March 2020. In contrast, children born in April, who are the eldest preschoolers, did not experience school closure because preschools were generally open at that time, which implies that children very close in age-in-months were exposed to different schooling policies. These two groups seemed to experience a similar threat caused by the spread of COVID-19, and they were also exposed to other policies such as requests for physical distancing similarly[3], but whether they experienced school closure in March was totally different between them.[4] Utilizing this natural experiment, we implemented a large-scale online survey to uncover the *pure* impact of school closure on the well-being of families comprehensively.

3. Data

3.1. Survey

For this study, we hired an Internet-survey company called Cross Marketing, Inc., and employed random sampling from about

[2] For more details on school closure in Japan, see Online Appendix A.

[3] For international readers, Ando et al. (2020) provide a comprehensive overview of the Japanese government's response to the COVID-19 crisis in terms of the fiscal measures taken between January and June 2020.

[4] Since school closure lasted until June, the group who was eldest in preschools as of March also experienced school closure in April when they entered elementary school.

4,790,000 people across the nation who had pre-registered as potential survey participants. The survey was implemented during the period from July 22, 2020, to August 19, 2020.

In the survey, we targeted married women whose co-resident firstborn child was born between April 2, 2010 and April 2, 2016, which roughly corresponds to 4–10 years old.[5]

In the actual implementation, we sent out invitations to our survey to 44,218 women, and among them, 22,553 mothers responded to our invitations and satisfied the requirements of the sample. We also included a question asking about their willingness to participate in the main survey after having explained that the main survey includes some sensitive questions such as inquiring about their mental health and the marital relationship. Through this question, 17,860 mothers eventually agreed to move on to our main survey. Ultimately, 15,836 mothers answered all the necessary questions, and thus, the number of the sample in the main analyses is 15,836.[6]

3.2. Descriptive statistics and representativeness

Next, we report descriptive statistics in our survey in comparison with other representative surveys. From the planning of the questionnaire, we made several questions the same as those in an already-existing survey in order to check the representativeness of our online internet survey afterward. To check the representativeness, we utilized two waves of the Japanese Study of Stratification, Health, Income and Neighborhood (J-SHINE), which were conducted in 2010 and 2012. The reason we use J-SHINE here is that it asks about the incidence of DVs in a solid manner proposed by Straus and Douglas (2004) and also includes other basic variables common to our covariates.[7]

The comparison with J-SHINE in Table 1 provides useful information about the representativeness of the respondents in our survey. First, we do not see any large difference between our data and J-SHINE in most of the mean values of basic covariates such as the age of respondents and the number of children.[8]

Next, the incidence of physical DVs, which is a useful indicator of marital quality, seems to be similar between our survey and J-SHINE: The total physical DV score was 0.3 in our survey and 0.26 in J-SHINE. This supports that our survey did not pick up a very specific population in terms of marital quality and family environment related to children.

[5] We did not impose any restriction on siblings of the firstborn child in the sampling process because restricting the sample to mothers who have only one child would likely be biased toward families with special features such as low socioeconomic status or strongly career-oriented double-income couples.

[6] We have also checked that there is no significant discontinuity at the threshold for both the fraction of the sample drop at the sensitivity question and the fraction of those who moved on to the main survey but did not complete the survey. For more details, see Online Appendix B.

[7] The main purpose of J-SHINE is to provide an interdisciplinary longitudinal survey database with comprehensive measures of living conditions, social environments, health, and biomarkers among Japanese residents aged less than 50. J-SHINE respondents were chosen from four metropolitan areas of Japan (Takada et al., 2014). Specifically, adults aged 25–50 years were randomly selected from the residential registry data. The first wave of data was collected in 2010, and the second was collected in 2012. The first wave includes 4,357 respondents, out of which 2,961 persons also participated in the second wave. From the entire sample, we chose female respondents whose firstborn child was 4–10 years old to compare with our survey.

[8] Concerning the educational level of mothers, we find a non-negligible difference in educational levels between the participants in our survey and J-SHINE. According to the School Basic Survey (MEXT, 2020), the ratio of female students who go on to four-year university studies has dramatically increased since the late 2000s; the ratio increased from 36.8% in 2005 to 45.2% in 2010, and most of the targets of this survey were from this cohort. Considering this fact, the mean value of the college dummy from our data is considered to be reasonable.

3.3. Dependent variables

- **Changes Related to Children Caused by the COVID-19 Outbreak**
  Our survey contains many yes-or-no questions about the changes in the respondent and her family members due to the COVID-19 outbreak. Among these questions, we report the results on the items regarding changes in the respondent's child and the mother–child relationship caused by the COVID-19 outbreak.

- **Domestic Violence**
  Following several studies (Straus and Douglas, 2004; Hidrobo and Fernald, 2013), we adopted multidimensional concepts of incidents of DVs that cover the emotional and physical aspects of DVs. For emotional DVs, we asked about incidents of "neglect or ignoring," "insult," and "behavior control." For much more physical DVs, we asked whether a respondent has broken their spouse's possessions or threatened their spouse by attempting to strike or actually striking them. For each positive response on DV, whether the violence was initiated by the father or the mother is noted. Thus, we asked 10 questions on DVs (i.e., 5 types and who did it). We also asked the frequency of the type of DV using the following three categories (i.e., 1. Never, 2. Sometimes, and 3. Frequently). Then, we calculated the total score of DVs by summing up the frequency measure over each of the 10 questions, which results in a score range from 10 to 30.

- **Satisfaction with Marriage**
  As a convenient way to evaluate the quality of a marriage, we asked: "Are you satisfied with marital life?" with a 5-point Likert-type scale ranging from 1 for "not at all" to 5 for "very satisfied.".

- **Risk of Divorce**
  We measured the risk of divorce from four aspects: the incidence of 1. quarrel, 2. discussion of divorce with the spouse, 3. self-thinking of divorce, and 4. proposal of divorce from husband. The frequency of each item was evaluated on a 5-point Likert-type scale, ranging from "usually" for 5 to "none" for 1. We report the total score, which ranges from 4 to 20.

- **Quality of Marriage Index (QMI)**
  QMI, which is one of the most popular indexes for evaluating marital quality in the academic literature, was developed by Norton (1983). In this study, we used Moroi (1996)'s marital quality scale, which incorporated and translated the concepts contained in Norton (1983)'s QMI into Japanese. Moroi (1996)'s marital quality scale consists of six items regarding marital life,[9] and each question was answered on a 4-point scale. We report the total score, which ranges from 6 to 24. Note that a higher score indicates a higher quality of marriage.

4. Empirical strategy

We will explore the effects of school closure on family well-being during the outbreak of COVID-19. Here, note that we use the word "schooling" to refer to school-aged children (generally over seven years old) attending elementary school as well as preschool children (younger than seven years) going to nursery school or kindergarten as usual, despite the COVID-19 outbreak. Similarly,

[9] The six responses are the following: (1) We have a good marriage, (2) my relationship with my partner is very stable, (3) Our marriage is strong, (4) my relationship with my partner makes me happy, (5) I really feel like part of a team with my partner, (6) I am really happy with my marriage.

Table 1
Descriptive statistics: comparison with J-SHINE.

	Our Survey		J-SHINE	
	Mean (1)	S.D. (2)	Mean (3)	S.D. (4)
Basic Characteristics				
Age	37.31	5.25	36.97	4.72
Number of Children	1.89	0.71	1.88	0.73
Firstborn Child Is a Girl	0.49	0.50	0.50	0.50
4-Year College Graduates: Mother	0.40	0.49	0.28	0.45
4-Year College Graduates: Father	0.47	0.50	0.54	0.50
Working (Mother)	0.55	0.50	0.48	0.50
Regular Worker (Mother)	0.20	0.40	0.14	0.35
Domestic Violence				
More than once = 1				
Physical DV	0.30	0.96	0.26	0.75
Physical DV by Wife	0.15	0.49	0.14	0.45
Physical DV by Husband	0.15	0.50	0.12	0.42
Frequently = 1				
Physical DV	0.11	0.59	0.14	0.59
Physical DV by Wife	0.05	0.30	0.06	0.32
Physical DV by Husband	0.05	0.31	0.08	0.35
N	15,836		746	

Notes: The total score of DV measures is 10 at maximum because we asked 10 questions on DVs, and here we used dummies for each item of DV. The 10 DV items consist of five types of DVs (e.g., "ignoring" and "hitting") and who did it (i.e., wife or husband). In addition to this, we measured the frequency of DVs in three categories (i.e., Never, Sometimes, and Frequently). Thus, For the results of "More Than Once = 1," we count the number of DVs for which the respondent chose "Sometimes" or "Frequently." For the results of "Frequently = 1" in our survey, we counted the number of DVs for which the respondent chose "Frequently" only, while the results of "Frequently = 1" in J-SHINE, we count the number of DVs in which the respondent chose "More than Twice," since the J-SHINE survey counted the number of DVs in three categories (i.e., None, Once, and More than twice).

"non-schooling" refers to children not going to school, regardless of whether they belong to an elementary school or preschool.

4.1. Identification

As explained earlier, for school-aged children, elementary schools were completely closed as of March 2.[10] In contrast, nursery schools and kindergartens were generally open at that time, following the announcement of the Ministry of Health, Labour and Welfare.

Interestingly, these governmental decisions on the status of school closures created two groups of children that faced totally different statuses of school closures even with a very small time difference in the matter of their births. More concretely, as of March 2020, children at the age of 89 months (at the time of the survey) belonged to the first grade at elementary school, while children at the age of 88 months were still preschool children in the highest grade in preschool facilities. Note that in spite of the fact that the age-in-months between the two groups differs by only one month–88 and 89–whether or not they could go to school differed between the two groups.

Based on these facts, we uncover the effects of proactive school closures by comparing several outcomes between mothers who had children barely below and barely above the threshold of the age-in-months of 89. This comparison is based on the idea that both the observable and unobservable factors that could potentially affect outcomes of interest (i.e., y) are continuous at the age-in-months of 89. Thus, if we find any discontinuity in y at the threshold, it can be interpreted as the "pure" impact of the school closures.

4.2. Fuzzy regression discontinuity design

Although it was announced that preschools were exempted from this nationwide school closure, in truth, not all preschools were open: The decision to close or open nursery schools and kindergartens was left to each facility and to the municipality where they were located.

In other words, exceeding the threshold age-in-months, that is, 89, did not mean that the probability of "non-schooling" changed from 0 to 1, since there were some preschools that also made a decision to close the facilities. Furthermore, even though preschools were not closed, children or their parents could choose whether to attend. Thus, being a preschool child did not necessarily mean "full schooling," while school closure in elementary schools was enforced fully. Due to this imperfect compliance among preschools and the available option for preschool children and their parents whether to go to preschool, we apply a fuzzy RDD to estimate the causal effects of not going to school in March 2020.

Since our running variable is age-in-months, which is uncontrollable, there should in principle be no general manipulation problem around the threshold. Regarding another potential threat to identification, the timing of school entry cannot be manipulated, since school admission dates are strictly enforced (Kawaguchi, 2011). Also, grade retention is extremely rare in the Japanese school system.[11]

Finally, note that age-in-months of the *firstborn* child is used as the running variable. It is technically possible to use the age of the *youngest* child in the household as the running variable, but if that was used, it could lead to mistakenly treating some households with more than one child, for example households that consist of a preschool child whose preschool was open plus a school-aged child facing school closure, as those that did not face school closure at all. Note that this mistake cannot happen if we use the age-in-

[10] According to the MEXT (2020b), 99.9% of elementary schools were closed as of March 10.

[11] School grades change on April 2, not April 1, in Japan. Note that school closure was implemented by grade level, not by birth month, and that the information we have is children's age-in-months and their school grade. Thus, to construct a valid running variable, we made an adjustment to include those who were born on April 1 in the group of those who were born in March. Those born on April 1 are in a lower grade than those born on April 2, and thus it is necessary to separate the two. Only those who were born on April 1 can be identified even without information about the children's exact date of birth, if we use both the information of their grade and their birth month. By doing so, we can know whether they really experienced school closure at the level of age-in-months.

months of the firstborn child in the household as the running variable.

4.3. Local-linear regression

Since our identification framework is the fuzzy RDD, the treatment effect is recovered by dividing the marginal change of outcome variables around the threshold by the fraction of children who did not go to school due to the nationwide request of school closure in March 2020. Specifically, for a respondent (mother) i in our survey, we estimate a system of local-linear regressions of the following form. The first-stage specification is:

$$Non-Schooling_i = \alpha_0 + \alpha I(m_i \geqslant 89) + \alpha_L(m_i - 89) + (\alpha_R - \alpha_L)I(m_i \geqslant 89)(m_i - 89) + \varepsilon_i, \tag{1}$$

and the reduced-form specification is:

$$Y_i = \beta_0 + \beta I(m_i \geqslant 89) + \beta_L(m_i - 89) + (\beta_R - \beta_L)I(m_i \geqslant 89)(m_i - 89) + \epsilon_i, \tag{2}$$

where $89 - b \leqslant m_i \leqslant 89 + b$ and b is the optimal bandwidth around the cutoff point. Next, $Non-Schooling_i$ is a binary variable which takes a value of one if mother i's firstborn child did not go to school in March 2020, and y_i is the outcome variable of mother i's firstborn child or herself depending on the questions. m_i represents age-in-months of the firstborn child. $I(m_i \geqslant 89)$ is an indicator function that takes 1 if the firstborn child's age-in-months is 89 or older and otherwise takes 0. Utilizing the estimates on coefficients of $I(m_i \geqslant 89)$ separately obtained from these two equations, the fuzzy regression discontinuity (RD) estimate can be written as $\hat{\beta}/\hat{\alpha}$.

Note that the parameter $(\hat{\beta})$ obtained from Eq. (2) corresponds to the estimate that will be obtained from a sharp RD regression, which is also equivalent to the magnitude of the discontinuity at the threshold in each figure of outcomes.

Concerning the actual implementation and presentation of this framework, in addition to reporting the results of the conventional local-linear regression, we also report results from the robust bias-corrected inference method (Calonico et al., 2014, 2020). In the implementation, we use the triangular kernel function that weighs points near the threshold more heavily than those distant from the threshold. Regarding the choice of bandwidth, we use the mean square error optimal bandwidths proposed by the Calonico et al. (2014) (henceforth referred to as the CCT bandwidth).

In the estimations, we use heteroskedasticity-robust standard errors as suggested by Kolesár and Rothe (2018). They showed that the practice of clustering by the running variable does not resolve specification bias problems in discrete RDD settings and can even lead to CIs with substantially worse coverage properties than those based on the conventional heteroskedasticity-robust standard error. Recent papers such as Canaan (2020) have also tended to use the conventional robust standard error in response to the results of Kolesár and Rothe (2018), and so do we.

5. Empirical results

5.1. Checks for continuity assumption

Before presenting the empirical results, we check the validity of the continuity assumption in Online Appendix B. In the check for continuity assumption, we examined the continuity of observed covariates, such as the education level of respondents and age, and found that the basic characteristics of the respondents and their children are sufficiently continuous around the cutoff month. In addition to this, we checked how unobservable characteristics of respondents were distributed around the cutoff month by using the participation rate in the main survey after it was explained that the main survey included some sensitive questions, such as inquiring about negative impacts on family well-being. In short, we found the participation rate in the main survey was also continuous around the cutoff, suggesting that sample selection due to some sensitive questions was not so serious in our survey.

5.2. Impact of school closure on "Non-Schooling"

Fig. 1(a) shows the fraction of preschools and elementary schools in our sample that was available or unavailable as of March 15, 2020, for each of age-in-months. Since the cutoff value 89 corresponds to the age-in-months (evaluated in August 2020) for children to move from preschool to elementary school, the probability of school closure becomes suddenly 100% at this threshold.

Fig. 1(b) presents the RD estimate of the impact of being an elementary school student (becoming 89 age-in-months or older) on the probability of "non-schooling": It is 0.623, and it is significantly positive even at the 1% significance level.

Thus, we can utilize the large increase in the probability of "non-schooling" around the threshold (caused by the difference in the status of school closures) to identify the effects of schooling.

Since we already confirmed that there was no gap in observable and unobservable factors between mothers who have children barely below and barely above the threshold in the Online Appendix B, if there was some gap in the outcome variables at the threshold, it should have been caused by the discontinuity in the probability of schooling as shown in Fig. 1(b).

5.3. Impact of school closure on family

5.3.1. Empirical results related to children

First, we explore the impact of the national school closure on variables related to children. To estimate the impact on variables related to children, we created yes-or-no questions about the changes in the respondent due to the COVID-19 outbreak. By asking questions focused on changes caused by the COVID-19 outbreak, it is expected that we can exclude the possibility that the estimates capture the effect of the difference in the children's lifestyles over the past 12 months around the threshold.[12]

Both Fig. 2 and Table 2 report the results related to children from the yes-or-no questions about the changes due to the COVID-19 outbreak. Table 2 presents estimates from the conventional local-linear regression and those from the robust bias-corrected inference method (Calonico et al., 2014, 2020). Although as indicated in the previous section, the framework of our study is the fuzzy RD design, we also report sharp-RD estimates as well as fuzzy RD estimates here because reporting the sharp estimates is helpful to compare the magnitude of the estimates with the mean values of each dependent variable as well as the magnitude of the discontinuity in Fig. 2. In contrast, the fuzzy estimates restore the causal effect of not going to school in March 2020.

First, according to Fig. 2, the discontinuity in children's weight gain is very conspicuous, and the existence of significant discontinuity at the threshold is undeniable. To see the exact magnitude of the discontinuity and statistical significance level, we will also check Table 2.

According to the mean value of the dependent variable from Table 2, about 15 % of respondents answered that their child gained weight. During the period of school closure, children were basically

[12] On the parents' side, we also checked the effect of having an elementary school student, using the outcome from the previous year in the same RD setting, but we did not see any impact of having an elementary school student on variables related to parents in the previous year, when COVID-19 did not occur, which implies the validity of our identification methodology and the robustness of our results.

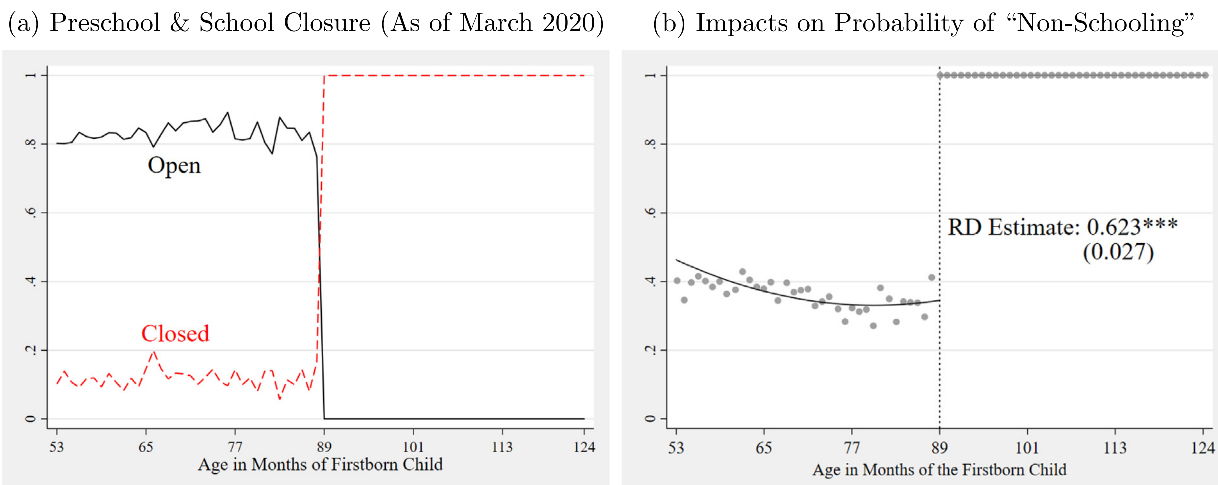


Fig. 1. The Impact of School Closures on "Non-Schooling". *Notes:* The childcare facilities that were available as of March 15, including those with requests for voluntary restraint in the use of the childcare facility, are categorized as "Open" in Fig. 1(a). In Fig. 1(b), observations are averaged within bins using the mimicking variance evenly-spaced method described in Calonico et al. (2015). Fig. 1(b) also includes second-order global polynomial fits represented by the solid lines. The estimate reported inside the figure is a sharp-RD estimate obtained from the conventional local-linear regressions. Conventional heteroskedasticity-robust standard errors are reported in parenthesis. The CCT bandwidth selector proposed by Calonico et al. (2014) is used to calculate the optimal bandwidth. The same bandwidth is applied to the areas below and above the cutoff. A triangular kernel function is used to construct the estimators. The selected optimal bandwidth is 9.634, and the number of observations within the bandwidth is 4,003. $***$ p<0.01, $**$ p<0.05, and $*$ p<0.1.

required to stay home, thus, this result makes sense. The sharp-RD estimates in Table 2 also suggest that there exists a very large discontinuity on children's weight gain by 9.2 to 9.6 percentage points around the threshold and that the estimated coefficients are significantly positive even at the 1% significance level. The fuzzy estimates suggest "non-schooling" due to school closure increases the fraction of respondents whose child (children) gained weight by 14.4 to 15.4 percentage points, which are statistically significant even at the 1% significance level. Note that this estimate means the counter-factual effect of school closure, that is, if the probability of school closure was 0 for the left side of the threshold and 1 for the right side, the effect of school closure would be 14.4 to 15.4 percentage points.

The other conspicuous result is about the following item: "I began to worry about how to raise my child (children) more frequently." The mean value of the dependent variable in Table 2 and the magnitude of the coefficient are the largest for this item. Concerning the mean value of the dependent variable, about one-fifth of respondents answered yes to this item. The fuzzy estimates indicate that mothers' worrying over how to raise their children increased by 17.8 to 20.2 percentage points, which is also statistically significant even at the 1% significance level. From this result, we can see how parents became confused and were not ready for the school closure in March without any instruction on how to raise children who do *not* go to school. Accordingly, we find a large discontinuity in this variable at the cutoff in Fig. 2.

Regarding the remaining two variables: "I began to leave my child home alone for a longer period of time (per day)" and "I began to worry about my relationship with my child (children) more frequently," although the magnitude of the discontinuity seems somewhat smaller than the first two variables, children's weight gain and mothers' worrying over how to raise their child (children), the discontinuities at the threshold are also clear in Fig. 2.

To check the statistical significance, we will move on to Table 2. Concerning the variable "I began to worry about my relationship with my child (children) more frequently," the mean value of the dependent variable indicates that 15.7 % of respondents answered yes to this question, the number of which is almost the same level as that for children's weight gain. Comparing this number with the sharp-RD estimates, it can be said that 6.4 to 7.5% of the increase in the fraction of those who answered yes to this question should have been caused by the school closure in March. In contrast, the fuzzy RD estimates indicate that a compulsory school closure that changes a probability of schooling from zero to one would increase the fraction who answer "I began to worry about my relationship with my child (children) more frequently" by 10.1–12.0 percentage points, which is not a small amount.

The modest but still significant effect among the four variables is the variable about leaving a child home alone, which potentially induces delinquency among children (Aizer, 2004; Blau and Currie, 2006). In Japan, since the school closure was announced abruptly and implemented within a short period, there should have been many parents who were not ready for it. For this reason, especially among working mothers who could not find any support for their children, there should have been many people who began to leave their child (or children) home alone. In Table 2, the mean value of the dependent variable is 7 %, and the sharp-RD estimates suggest that 4.6 to 5.2 % of the increase in the fraction of those who answered yes to the question,"I began to leave my child home alone for a longer period of time (per day)" can be due to the school closure that happened in March. In contrast, the fuzzy RD estimates suggest that if being a preschool child had exactly meant a "full schooling," the magnitude of the effect of school closures on mothers' leaving their child home alone would have been 7.3 to 8.4 percentage points.

5.3.2. Empirical results on parents

Note that in the Online Appendix A, we introduced evidence of a sharp increase in Google searches for the word "divorce" on March 2, when the school closure was suddenly announced. Does this lead to a situation in which marital relationships worsened in response to the confusion of the school closure and/or too great of a burden of childcare on parents?.

To answer this question, first, we examine the impact on DVs.[13] Fig. 3 presents the results of total scores of DV behavior in August. Obviously, the discontinuity at the threshold for this variable is "hard-to-see," and the scatter plots look very "noisy." Indeed,

[13] For the definition of each marital relationship measure in this section is reported in Section 3.3.

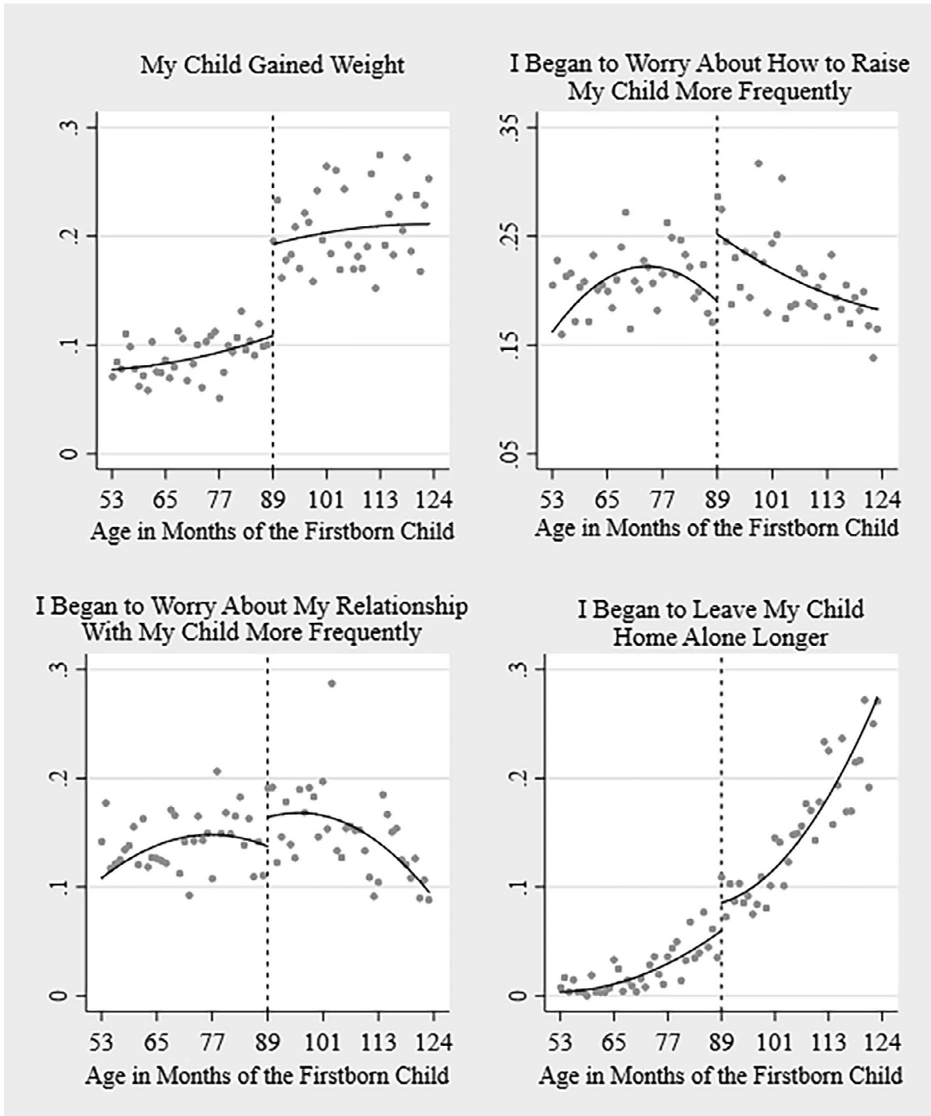


Fig. 2. RD Estimates on Changes Related to Children Caused by the COVID-19 Outbreak. *Notes:* Observations are averaged within bins using the mimicking variance evenly-spaced method described in Calonico et al. (2015). Each plot includes second-order global polynomial fits represented by the solid lines.

Table 2
RD estimates for the impact of school closures on variables related to children.

	(1)	(2) Sharp (Reduced Form)	(3)	(4) Fuzzy (IV)	(5)	(6)	(7)
Dependent Variable: 1 (Yes) or 0 (No)	Mean of Dep. Var.	Conventional	Bias-corrected	Conventional	Bias-corrected	Optimal Bandwidth	N
My child gained weight							
	0.151	0.092***	0.096***	0.144***	0.154***	11.570	4,728
		(0.022)	(0.026)	(0.035)	(0.042)		
I began to worry about how to raise my child more frequently							
	0.218	0.110***	0.124***	0.178***	0.202***	7.805	3,189
		(0.032)	(0.036)	(0.053)	(0.059)		
I began to worry about my relationship with my child more frequently							
	0.157	0.064***	0.075***	0.101***	0.120***	9.769	4,003
		(0.025)	(0.028)	(0.039)	(0.044)		
I began to leave my child home alone for a longer period of time (per day)							
	0.070	0.046***	0.052***	0.073***	0.084***	9.898	4,003
		(0.017)	(0.019)	(0.027)	(0.031)		

Notes: Table 2 presents estimates from the conventional local-linear regressions as well as estimates to which the robust bias-corrected inference method (Calonico et al., 2014, 2020) is applied. Conventional heteroskedasticity-robust standard errors are reported in parentheses. For the estimates from the robust bias-corrected inference method, robust standard errors are reported. The CCT bandwidth selector proposed by Calonico et al. (2014) is used to calculate the optimal bandwidth. The same bandwidth is applied to the areas below and above the cutoff. A triangular kernel function is used to construct the estimators.*** $p<0.01$, ** $p<0.05$, and * $p<0.1$.

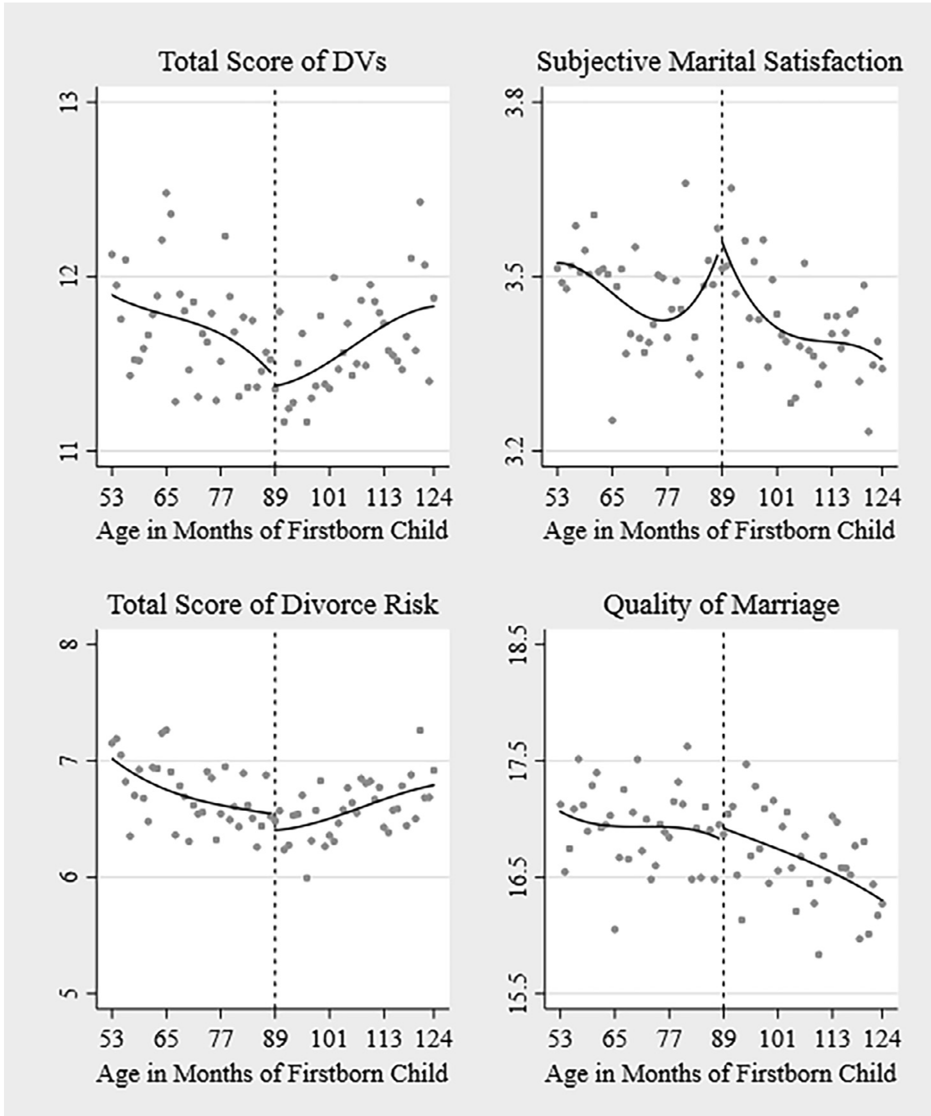


Fig. 3. RD Estimates for the Impact of School Closures on Parents in August. *Notes:* Observations are averaged within bins using the mimicking variance evenly-spaced method described in Calonico et al. (2015). Each plot includes second-order global polynomial fits represented by the solid lines.

according to Table C1 in the Online Appendix, this discontinuity turned out to be statistically insignificant.

Next, we will see how other measures of marital relationships were affected by the sudden school closure. Fig. 3 also reports the results on several measures of marital relationships, and we do not see clear discontinuities from any measure of marital relationships. The estimation results for these variables are presented in Table C1, and we confirmed that the estimates are all insignificant for these variables.[14] Note that although the bias-corrected RD estimate of the impact on subjective marital satisfaction is close to zero, it should not be characterized as precisely zero, because the standard error is too large to rule out economically significant effects. The same trend can be seen for the other estimates in Table C1 as well.

Thus, although in Fig. 3, all the discontinuities at the threshold seem very tiny or negligible, indeed, we have confirmed that these "hard-to-see" or "invisible" gaps in the figures are truly statistically insignificant by Table C1. Thus, we have not obtained any significant results in marital relationship measures.

There might be a concern that the timing of the survey was too late to capture the impact on these measures. Thus, we also asked about situations related to marital relationship in March for each variable except for "Quality of Marriage Index," which results in the March result being missed. Table C2 reports the comparison of impacts on DVs between August and March. As can be confirmed from the table, we can see the increase in the mean value of the dependent variable in March, but we do not observe any statistically significant results. Although due to the limited space, we do not include the comparison between March and August for other variables, and other measures of marital relationships also show a similar pattern, that is, we do not see any statistically significant results even in March.

5.3.3. Robustness check

In Online Appendix D, we present some robustness checks. First, in Table D1, we report results with local-quadratic specification. Second, in Table D2, we report results with another bandwidth

[14] This result might be a bit surprising if we recall the sharp increase of Google searches for the word "divorce" on March 2 shown in the Online Appendix A. However, in reality, we do not see any evidence of a significant increase even in the risk of divorce. This is probably because Google searches measure the trends of divorce risk only in a rough manner.

selector type that focuses on delivering confidence intervals with optimal coverage error rates proposed by Calonico et al. (2018). From both robustness checks, we can confirm that our main results have been preserved, which indicates how robust our main results are.

6. Sub-sample analysis

In this subsection, by utilizing rich individual-level information included in our survey, we will explore a sub-sample analysis of children's outcomes. We first focus on children who had the greatest potential to be negatively affected by the COVID-19 pandemic (Bacher-Hicks et al., 2020; Chetty et al., 2020; Adams-Prassl et al., 2020) —that is, those with mothers and fathers with low educational attainment. In this sub-sample analysis, mothers and fathers are categorized into the "high" education group if they graduated from college, and otherwise into the "low" group. Next, we explore the heterogeneity based on working status as of February 2020 and availability of informal support from grandparents as of February 2020 because sudden school closure may have detrimental effects on families with low availability of alternative childcare resources—for example, dual-income couples whose parents did not live nearby.

Finally, we additionally split the sample according to (1) prefecture of residence, (2) mother's age, (3) gender of the child, and (4) the number of sibling(s). Out of (1) prefecture of residence, we constructed two groups according to whether the respondents lived in one of the seven prefectures where the state of emergency was declared proactively on April 7 because of the rapid spread of COVID-19. While elementary schools were closed nationwide, the local spread of COVID-19 may have affected how they coped with new daily life. For example, children in low infection regions could play together during March and April, but those in high infection regions could not, so they had to play alone, and the childcare burden, especially for mothers, might have been enormous.

Subsample results on changes in children due to the COVID-19 outbreak are reported in Fig. E1. In Fig. E1(a), we find a significant increase in home-alone hours among two-income households and households with boys. While the coefficient is statistically significant among households with children's grandparents living nearby, the point estimates do not differ substantially by the grandparents' proximity. In Fig. E1(b), we find suggestive evidence that the extent of children's weight gain due to the COVID-19 outbreak differs according to the educational attainment of the fathers. When fathers have graduated from college, the fuzzy RD estimate is smaller by about 5 percentage points than that for children with non-college-educated fathers. This directly suggests that school closures have negative effects on children's health, especially among those from low socioeconomic backgrounds. In addition to this, we found strong adverse effects among children without support from grandparents and without sibling(s). While there may be numerous stories to account for these findings consistently, one possibility is that school closure made children much more physically inactive when they had no close relatives. In fact, during March and April, it was generally difficult for children to meet and play with non-relatives so that the absence of relatives, especially siblings, leads to weight gain through a sharp reduction of physical activities.

Finally, we found a large increase in childcare anxiety measured by the item "I began to worry about how to raise my child more frequently" among mothers who lived in the seven prefectures with a high infection rate. On the subsample results on other outcome variables such as DVs and QMI, see Online Appendix E.

7. Conclusions

This study provides the first evidence in a comprehensive study design of how school closures without a strong lockdown policy affected children and parents. Unlike countries that implemented lockdown, the Japanese government did not implement strong anti-COVID-19 policies except for school closures. In addition to this, our research design enables us to compare the two groups of children and parents who faced totally different statuses of school closures even with a very small difference in the timing of the children's birth. Due to the very small difference in the timing of their birth, they are likely to have similar characteristics and experiences of the same alternative anti-COVID-19 policies. Thus, this study successfully estimates the *pure* impacts of school closure on comprehensive outcome variables related to families.

As the most pronounced results, we have found a clear increase in children's body weight, time spent home alone by children, and mothers' worrying over how to raise their children. Quantitative impacts are also sizable: The fraction of mothers whose child (children) gained weight increased by 14.4 to 15.4 percentage points and mothers' worrying over how to raise their children increased by 17.8 to 20.2 percentage points. Regarding the increase in body weight, the effects were prominent among children from low socioeconomic backgrounds.

Overall, this study implies that the school closure increased time spent home alone, and the reduction in physical activities might directly have resulted in the large weight gain among children. Furthermore, because of these negative effects on children, mothers began to worry about how to raise their children more frequently, which may lead to further deterioration of a healthy parent–child relationship in the long run.

Concerning the current policy debate on school closure, this paper provides clear insights on what we should know before we close schools during a pandemic. First, given the results presented in this paper, school closure may have unexpected side effects on non-academic health outcomes (i.e., weight gain). It is obvious that every day walking on school roads with friends itself is an exercise and that children are naturally kept away from eating unhealthy snacks by schools. School meal programs generally provide children with well-balanced meals. As a result of the sudden loss of these in their daily lives, many children experienced weight gain. Even if online education could offer a complete substitute for real in-person education in the future, in an academic sense, it should not be ignored that schools do not solely give academic education to children, but contribute to children's healthy lives. Therefore, this aspect of real in-person schooling should not be ignored. Given that some epidemiological studies consistently find that school closure is not an effective tool to control COVID-19 (Armbruster and Klotzbücher, 2020; Iwata et al., 2020), we should pay closer attention to the adverse side effects of school closures carefully.

Second, if we have to close schools again due to the overwhelming spread of COVID-19, schools should provide families with adequate online education as well as guidelines on how children and parents can spend their time at home in more healthy and productive ways. Throughout the school closure during March and April, many parents in Japan worried about their relationships with their children because elementary schools did nothing except provide homework. It was really surprising that during March and April 2020, interactive online learning was provided in only five percent of all schools in Japan (MEXT, 2020a). Consistent with a lack of policy to support children's education, our results show that school closure had negative effects on mothers' worrying about how to raise their children. While this should be confirmed carefully, the deterioration of the mother–child relationship might have been alleviated if policymakers had provided much more effective measures to compensate for the sudden stop of schooling.

We hope to examine effects on children from other perspectives as well. For example, high-quality schools (Dobbie and Fryer, 2011) and intensive compulsory education (Kawaguchi, 2016) both contribute to equalizing the academic performance of children from different socioeconomic backgrounds; thus there is a possibility that this school closure may lead to the widening of the inequality of academic performance of children and hence the inequality of their future economic outcomes. To uncover these long run effects is left to future studies.

8. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.jpubeco.2020.104364.

References

Adams-Prassl, A., Boneva, T., Golin, M., Rauh, C., 2020. Inequality in the impact of the coronavirus shock: Evidence from real time surveys. J. Public Econ. 189 (104), 245. https://doi.org/10.1016/j.jpubeco.2020.104245. http://www.sciencedirect.com/science/article/pii/S0047272720301092.

Aizer, A., 2004. Home alone: Supervision after school and child behavior. J. Public Econ. 88 (9–10), 1835–1848. https://ideas.repec.org/a/eee/pubeco/v88y2004i9-10p1835-1848.html.

Ando, M., Furukawa, C., Nakata, D., Sumiya, K., et al., 2020. Fiscal responses to the COVID-19 crisis in Japan: the first six months. Natl. Tax J. 73 (3), 901–926.

Armbruster, S., Klotzbücher, V., 2020. Lost in lockdown? COVID-19, social distancing, and mental health in Germany. Diskussionsbeiträge 2020–04, Freiburg i. Br., http://hdl.handle.net/10419/218885.

Bacher-Hicks, A., Goodman, J., Mulhern, C., 2020. Inequality in household adaptation to schooling shocks: Covid-induced online learning engagement in real time. Tech. rep., National Bureau of Economic Research.

Baron, E.J., Goldstein, E.G., Wallace, C.T., 2020. Suffering in silence: How COVID-19 school closures inhibit the reporting of child maltreatment. J. Public Econ. 190 (104), 258. https://doi.org/10.1016/j.jpubeco.2020.104258. http://www.sciencedirect.com/science/article/pii/S0047272720301225.

Blau, D., Currie, J., 2006. Pre-school, day care, and after-school care: who's minding the kids? In: Hanushek, E., Welch, F., (eds.), Handbook of the Economics of Education, Handbook of the Economics of Education, vol. 2, Elsevier, chap 20, pp 1163–1278, https://ideas.repec.org/h/eee/educhp/2-20.html.

Brodeur, Abel, Clark, Andrew E., Fleche, Sarah, Powdthavee, Nattavudh, 2021. COVID-19, lockdowns and well-being: Evidence from Google Trends. Journal of Public Economics 193. https://doi.org/10.1016/j.jpubeco.2020.104346.

Cabinet Office, 2020. Prime Minister Abe's Press Conference On February 29 https://www.kantei.go.jp/jp/98_abe/statement/2020/0229kaiken.html, . accessed: 2020-09-14.

Calonico, S., Cattaneo, M.D., Titiunik, R., 2014. Robust nonparametric confidence intervals for regression-discontinuity designs. Econometrica 82 (6), 2295–2326.

Calonico, S., Cattaneo, M.D., Titiunik, R., 2015. Optimal data-driven regression discontinuity plots. J. Am. Stat. Assoc. 110 (512), 1753–1769.

Calonico, S., Cattaneo, M.D., Farrell, M.H., 2018. On the effect of bias estimation on coverage accuracy in nonparametric inference. J. Am. Stat. Assoc. 113 (522), 767–779.

Calonico, S., Cattaneo, M.D., Farrell, M.H., 2020. Optimal bandwidth choice for robust bias corrected inference in regression discontinuity designs. Econometrics J. 23 (2), 192–210.

Campedelli, G.M., Aziani, A., Favarin, S., 2020. Exploring the immediate effects of COVID-19 containment policies on crime: an empirical analysis of the short-term aftermath in Los Angeles. Am. J. Criminal Justice. https://doi.org/10.1007/s12103-020-09578-6.

Canaan, S., 2020. The long-run effects of reducing early school tracking. J. Public Econ. 187 (104), 206. https://doi.org/10.1016/j.jpubeco.2020.104206. http://www.sciencedirect.com/science/article/pii/S0047272720300700.

Cauchemez, S., Valleron, A.J., Boelle, P.Y., Flahault, A., Ferguson, N., 2008. Estimating the impact of school closure on influenza transmission from sentinel data. Nature 452, 750–754. https://doi.org/10.1038/nature06732.

Chetty, R., Friedman, J.N., Hendren, N., Stepner, M., Team TOI, 2020. How did COVID-19 and stabilization policies affect spending and employment? A new real-time economic tracker based on private sector data. Working Paper 27431, National Bureau of Economic Research, doi:10.3386/w27431, http://www.nber.org/papers/w27431.

Dobbie, W., Fryer, J Roland G., 2011. Are high-quality schools enough to increase achievement among the poor? Evidence from the Harlem children's zone. Am. Econ. J.: Appl. Econ. 3 (3), 158–187. https://doi.org/10.1257/app.3.3.158. https://www.aeaweb.org/articles?id=10.1257/app.3.3.158.

Hidrobo, M., Fernald, L., 2013. Cash transfers and domestic violence. J. Health Econ. 32 (1), 304–319. https://doi.org/10.1016/j.jhealeco.2012.11.002. http://www.sciencedirect.com/science/article/pii/S0167629612001750.

Iwata, K., Doi, A., Miyakoshi, C., 2020. Was school closure effective in mitigating coronavirus disease 2019 (COVID-19)? Time series analysis using bayesian inference. Int. J. Infect. Diseases 99, 57 – 61, doi: 10.1016/j.ijid.2020.07.052, http://www.sciencedirect.com/science/article/pii/S1201971220305981.

Kawaguchi, D., 2011. Actual age at school entry, educational outcomes, and earnings. J. Jpn. Int. Econ. 25 (2), 64–80.

Kawaguchi, D., 2016. Fewer school days, more inequality. J. Jpn. Int. Econ. 39, 35–52.

Kessler, R.C., Andrews, G., Colpe, L.J., Hiripi, E., Mroczek, D.K., Normand, S.L.T., Walters, E.E., Zaslavsky, A.M., 2002. Short screening scales to monitor population prevalences and trends in non-specific psychological distress. Psychol. Med. 32, 959–976.

Kolesár, M., Rothe, C., 2018. Inference in regression discontinuity designs with a discrete running variable. Am. Econ. Rev. 108 (8), 2277–2304.

Lee, D.S., Lemieux, T., 2010. Regression discontinuity designs in economics. J. Econ. Lit. 48 (2), 281–355.

Leslie, E., Wilson, R., 2020. Sheltering in place and domestic violence: Evidence from calls for service during COVID-19. J. Public Econ. 189 (104), 241. https://doi.org/10.1016/j.jpubeco.2020.104241. http://www.sciencedirect.com/science/article/pii/S0047272720301055.

MEXT, 2020. School Basic Survey https://www.mext.go.jp/b_menu/toukei/chousa01/kihon/1267995.html, . accessed: 2020-11-14.

MEXT, 2020a. School closures to combat new coronavirus infections and learning guidance in public schools https://www.mext.go.jp/content/20200421-mxt_kouhou01-000006590_1.pdf, . accessed: 2020-11-24.

MEXT, 2020b. What is the mext and Japanese government doing in response to covid-19? https://www.mext.go.jp/a_menu/coronavirus/index.html, . accessed: 2020-09-14.

Mohler, G., Bertozzi, A.L., Carter, J., Short, M.B., Sledge, D., Tita, G.E., Uchida, C.D., Brantingham, P.J., 2020. Impact of social distancing during COVID-19 pandemic on crime in los angeles and indianapolis. J. Criminal Justice 68 (101), 692. https://doi.org/10.1016/j.jcrimjus.2020.101692. http://www.sciencedirect.com/science/article/pii/S0047235220301860.

Moroi, K., 1996. Perceptions of equity in the division of household labor. Jpn. J. Family Psychol. 10, 15–30 (in Japanese with English abstract).

New York Times, 2020. Japan shocks parents by moving to close all schools over coronavirus https://www.nytimes.com/2020/02/27/world/asia/japan-schools-coronavirus.html, . accessed: 2020-09-14.

Norton, R., 1983. Measuring marital quality: a critical look at the dependent variable. J. Marriage Family, 141–151.

Payne, J.L., Morgan, A., 2020. COVID-19 and violent crime: A comparison of recorded offence rates and dynamic forecasts (ARIMA) for March 2020 in Queensland, Australia. SocArXiv g4kh7, Center for Open Science, doi:10.31219/osf.io/g4kh7, https://ideas.repec.org/p/osf/socarx/g4kh7.html.

Pietrobelli, A., Pecoraro, L., Ferruzzi, A., Heo, M., Faith, M., Zoller, T., Antoniazzi, F., Piacentini, G., Fearnbach, S.N., Heymsfield, S.B., 2020. Effects of COVID-19 lockdown on lifestyle behaviors in children with obesity living in Verona, Italy: A longitudinal study. Obesity 28 (8), 1382–1385. https://doi.org/10.1002/oby.22861. https://onlinelibrary.wiley.com/doi/abs/10.1002/oby.22861, https://onlinelibrary.wiley.com/doi/pdf/10.1002/oby.22861.

Piquero, A.R., Jordan, Riddell SANC, Bishopp, Rand, Reid, J.A., Piquero, N.L., 2020. Staying home, staying safe? a short-term analysis of COVID-19 on dallas domestic violence. Am. J. Criminal Justice 45 (4), 601–635.

Sanga, S., McCrary, J., 2020. The impact of the coronavirus lockdown on domestic violence. Working paper.

Sibley, C., Greaves, C., Satherley, L., Wilson, N., Lee, C., Osborne, J., Barlow, Sibley C., Greaves, L., Satherley, N., Wilson, M., Overall, N., Lee, C., Milojev, P., Bulbulia, J., Osborne, D., Milfont, T., Houkamau, C., Duck, I., Mueller, W., 2020. Effects of the COVID-19 pandemic and nationwide lockdown on trust, attitudes towards government, and wellbeing. Am. Psychol. 75. https://doi.org/10.1037/amp0000662.

Straus, M.A., Douglas, E.M., 2004. A short form of the revised conflict tactics scales, and typologies for severity and mutuality. Violence Vict. 19 (5), 507–520.

Takada, M., Kondo, N., Hashimoto, H., 2014. Japanese study on stratification, health, income, and neighborhood: study protocol and profiles of participants. J. Epidemiol. 24 (4), 334–344.

UNESCO, 2020. Covid-19 response https://en.unesco.org/covid19, . accessed: 2020-09-14.

Online Appendix

A Schools and School Closure in Japan

This appendix provides a brief explanation of school systems in Japan and a comprehensive description of school closures from March 2.

A.1 Schools in Japan

In Japan, choices of childcare change according to the child's age. For children less than 3 years old, parents, especially mothers, generally take care of their children as full-time childcare providers. If mothers work outside the home, their children can be left in a daycare center. A typical daycare center in Japan keeps preschool children from 8:00 or 9:00 o'clock to 18:00 o'clock for low fees. Once the children of full-time housewives reach 3 years old, they can go to kindergarten. Japanese kindergartens typically have short business hours (e.g., 4 hours per day), which is mostly similar to the school hours for first graders. Also, kindergartens are closed during long seasonal vacations, just as elementary schools are, while there are no seasonal vacations at daycare centers. Because home care or home education for children aged 3 to 5 years old is not popular in Japan, preschool children are either in daycare centers or kindergarten before starting school. According to Cabinet Office (2015), 33.3% and 63.8% of 5-year-olds attend daycare and kindergartens, respectively.

Children enter elementary school at age 6. The school year starts April 2 and all children who turn 7 during the year from April 2 to April 1 in the next year enroll in first grade. While some students go to private schools, this proportion is only 1 percent. Finally, as is stated in the main text, the timing of school entry cannot be manipulated since school admission dates are strictly enforced with almost complete compliance (Kawaguchi, 2011).

A.2 General Description of Timeline of School Closure

As noted in the main text, elementary schools were closed on March 2. This sudden and unpredictable request for school closure was one of the most prominent features of Japan's anti-Covid-19 measures. From an international perspective, as is shown in Table A1, many developed countries closed schools only gradually. For example, in the UK, school closures started on February 20 in some regions, and nationwide closures were implemented about one month after the first closures. However, nationwide closure was abruptly implemented in Japan despite the fact that the number of accumulated Covid-19 deaths on the day of the announcement was only three people.

While the true reason behind Prime Minister Abe's declaration of school closures is not completely clear, hosting the Tokyo Olympics in August 2020 as planned would be a strong motivation to implement school closure. The official explanation in the statement from the Cabinet Office seems to be based on the belief that infectious disease spreads rapidly among

Table A1: Timings of School Closures and Reopenings in Selected Countries

	Localized Closure	National Closure	Localized Again	Reopen
Japan	-	March 2	April 6	June 1
US	February 28	April 10[a]	-	-
UK	February 28	March 20	June 1	-
France	March 3	March 16	May 11	May 25
Germany	March 3	March 18	May 4	-
Italy	February 24	March 10[b]	-	-
Canada	March 13	March 23	-	-
Sweden	March 18[c]	-	-	June 15
Finland	-	March 18[d]	-	-
Denmark	-	March 16	April 15	May 27
Switzerland	March 12	March 16	May 11	June 8
Korea	-	March 2	May 20	June 8
China	February 16	February 21	April 27	
Australia	March 24	-	-	June 9

Notes:
a. The majority of states mandated school closures, including until the end of the academic year. Some states, however, recommended but did not mandate school closures.
b. On May 11, the government announced school closures until the end of the academic year. Classes continued through distance learning.
c. Closure of all upper secondary educational institutions and universities.
d. Education in the early grades continued in some cases, as well as in general for children with special needs and if considered necessary for the completion of studies.
Source: (UNESCO, 2020)

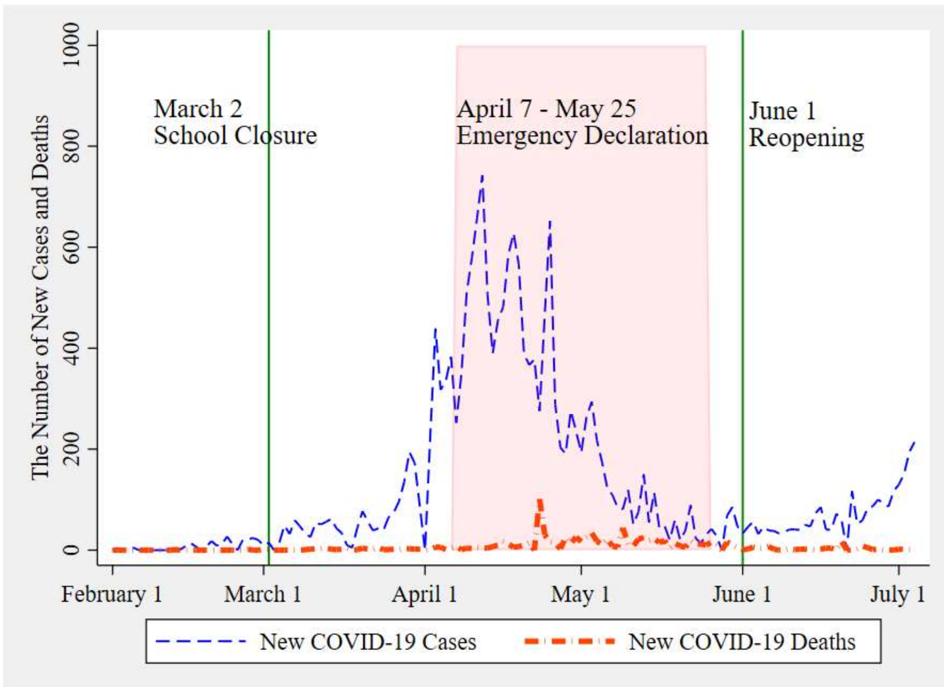
young children, as with influenza, and the government should save children from this new disease. Indeed, Prime Minister Abe explained why schools should be closed as follows: (Cabinet Office, 2020):

When schools are closed, it will be a great burden for families with small children. Nevertheless, above all, the health and safety of children must be the first priority, and we should avoid having many children and school teachers gather for long hours on a daily basis. Also, we must prepare for the risk of infection by avoiding gathering in the same space.

Shinzo Abe, February 29, 2020

The timeline of major events after March 2 is summarized in Figure A1(a). Since the number of new COVID-19 cases dramatically increased, the declaration of emergency was made on April 7 for seven prefectures that had especially large numbers of COVID-19 infections. However, the situation was worsening and it was finally expanded nationwide on April 16. Note that the declaration of emergency gives local governments the power to enforce preventive steps and allows them to request school and business closures, though there are no legal penalties for noncompliance. Since the epidemic was temporally over, the state of emergency was lifted on May 25. The date of schools reopening was June 1.

Figure A1: The COVID-19 Outbreak in Japan



Notes: The area shaded in red represents the duration of the declaration of emergency. The solid vertical line represents the timing of the announcement of school closure on March 2 and the reopening of schools on June 1.

A.3 Proportion of Schools Closed

Since the official document of the Ministry of Education, Culture, Sports, Science and Technology (MEXT) summarized how many schools opened or closed at each time point (MEXT, 2020b), we could follow the influence of major events such as the declaration of a state of emergency more accurately. Soon after school closure began on March 2, 99.9% of elementary schools were closed by March 4, as shown in Table A2. Note that there are no official statistics on how many preschools opened in March, but preschools were generally exempted from the request of school closure because of concerns about preschool children staying home alone when their parents were at work outside the home. Our survey confirmed this point directly; Figure 1 shows that most children aged less than 89 months, namely preschool children as of March 2020, could actually go to preschool.

In the Japanese school calendar, in which the new year starts in April, spring vacation generally starts March 25–26 and ends April 5–6 in 2020. Thus school-aged children were

Table A2: Major Events and the Proportion of Schools Closed

Date	Major Events	Preschool (1)	Elementary School (2)
March 2	School Closures Started		
March 4		N/A	99%
March 16		N/A	99%
March 25	Spring Vacation Stared		
April 6		27%	36%
April 7	Declaration of Emergency in 7 Prefectures		
April 10		46%	67%
April 16	Nationwide Declaration of Emergency		
April 22		73%	95%
May 11		77%	88%
May 25	End of Declaration of Emergency		
June 1		2%	1%

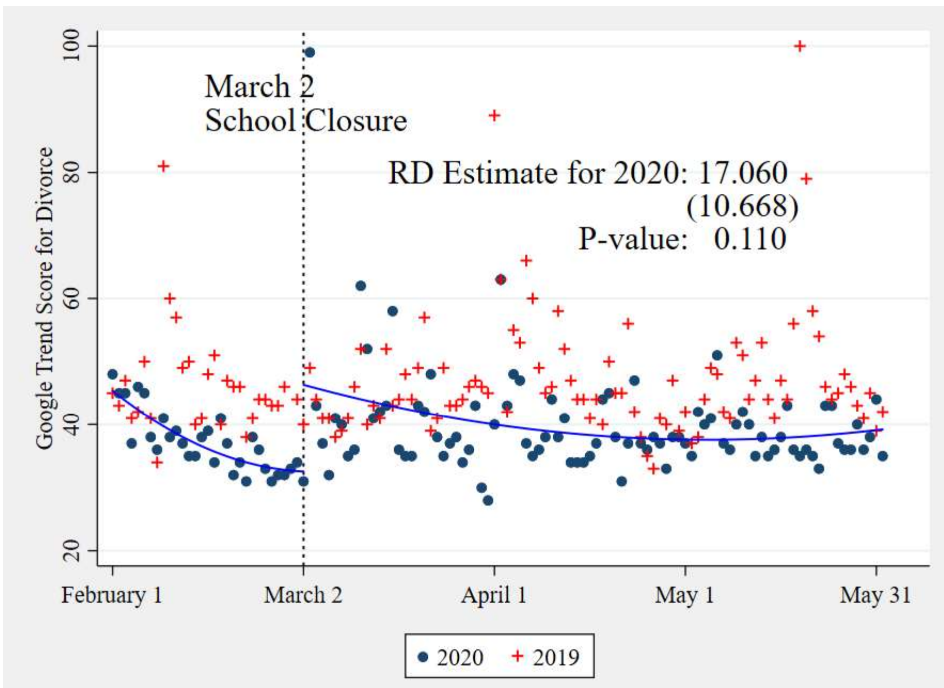
Notes: Rows shaded in gray represent the date when the major event occurred. Chiba, Saitama, Tokyo, Kangawa, Osaka, Hyogo, and Fukuoka represent the seven prefectures where the state of emergency was declared on April 7. The data in Column (1) does not include daycare centers.
Source: (MEXT, 2020b)

out of school for a month even at the beginning of a new school year. While some schools in the area with low infection rates opened gradually beginning on April 6, this trend reversed again on April 7 when the state of emergency was declared in seven large prefectures (i.e., Chiba, Sitama, Tokyo, Kanagawa, Osaka, Hyogo, and Fukuoka). The proportion of closed schools suddenly increased from 36% to 67% by April 10. When the nationwide declaration of emergency went into effect on April 16, almost all the schools were closed again. About 70% of preschools were also closed due to the nationwide declaration of emergency. Even as of May 11, 85% of elementary schools and 77% of preschools were still closed. These streams of events suggest that most elementary school students were deprived of access to education for at least two months, but preschool children experienced school closure for only about one month. Eventually, school closure ended when the declaration of emergency was lifted on May 25.

A.4 Google Search Trend for the Word "Divorce"

Because the school closure on March 2 was too proactive, it caused substantial confusion for families. As suggestive evidence of this confusion, a new term "corona divorce" became commonly used on Japanese social media to describe the surge of divorce risk during March-April 2020. Figure A2(b) reports a Google search trend, which represents the relative popularity of the word in Google searches. The number 100 on the y-axis indicates very frequent searches and 0 indicates the opposite. The dots and the plus (+) symbols represent the data in 2020 and 2019, respectively. The RD gap for 2020 data estimated by the local-linear regression is positive, although the estimate is somewhat noisy. The gap for 2019 data is invisible at around March 2, which means that the observed jump in March 2 in 2020 is not driven by the seasonal trend of the divorce search because we can see no significant jump at the threshold when using the data of the previous year (i.e. 2019).

Figure A2: Google Search Trends for the Word "Divorce"



Notes: In the upper figure, the area shaded in red represents the duration of the declaration of emergency. The solid vertical line represents the timing of the announcement of school closure on March 2 and the reopening of schools and daycare centers on May 25. The RD estimate reported in the figure was obtained in the same way as that of Figures 1(b) and Figure B2.

B Tests for the Continuity Assumption of the RD Design

B.1 Check for Continuity of Potential Covariates

In this subsection, we conduct a parallel RD analysis on potential covariates to determine whether they are continuous at the age of 89 months. Figure B1 presents the average number of children, the fraction of those whose firstborn child was a girl, the fraction of college graduates, the fraction of those whose parents' support was available at least as of Feb. 2020, the fraction of those who worked as regular workers as of Feb. 2020, and the fraction of those who did not work as of Feb. 2020 by each age-in-months of firstborn children.

According to Figure B1, there seems to be no discontinuity at the threshold age-in-months of 89, and indeed, it has been confirmed that the gap is statistically insignificant for all the potential covariates. This result demonstrates that observable factors were all continuous at the threshold, assuring the similarity of the two groups around the threshold.

In the next subsection, we will see the unobservable factors in the error term were also continuous at the threshold.

B.2 Check for Continuity in Unobservable Factors

As already explained in the data section, before moving to the main survey, respondents were required to answer a question about their willingness to participate in the main survey after having been explained that the main survey included some sensitive questions, such as questions about mental health and marital relationships, as well as some questions about negative impacts that the childcare burden could cause.[15]

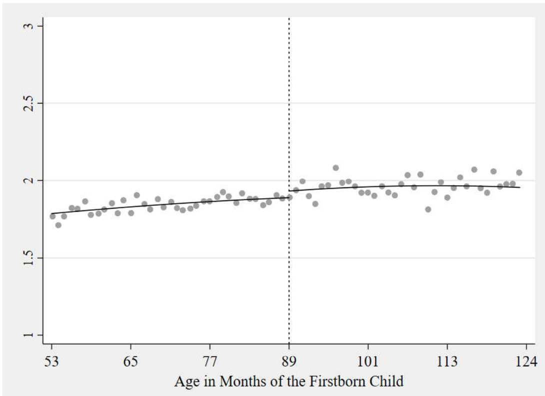
The respondents who refused to move on to the main survey were then dropped from the sample at this stage, and it was assumed that they felt uncomfortable in answering sensitive questions due to some negative experience related to the questions.[16]

[15] Through this question, about 20% of the respondents who satisfied all the required conditions about their attributes chose not to move on to the main survey because of the existence of sensitive questions.

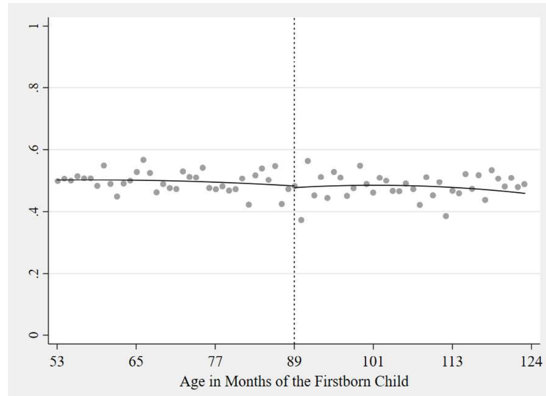
[16] Here, we will clarify that these respondents chose to quit the survey at this stage not because they had no time or had no interest in this research but because they had some problems or felt uneasy in answering sensitive questions. Before the last screening question, there were 11 questions, including questions asking for information about their children. In addition to this, the respondents had originally decided to participate in our survey after understanding the topic of the survey. Thus, if the issue was that they simply had no time to move on to the main survey or no interest in our survey, they would not have reached the stage of the last question of the screening test. In addition to this, by company rule, it was determined that remuneration would be paid only if a respondent finished all the questions of the main survey after passing the screening test. Thus, there was no benefit in proceeding to the screening test only in terms of the reward, which implies that the those who dropped out of the screening test at the last question were those who felt significantly uncomfortable about answering the potentially sensitive questions.

Figure B1: Check for Continuity of Potential Covariates

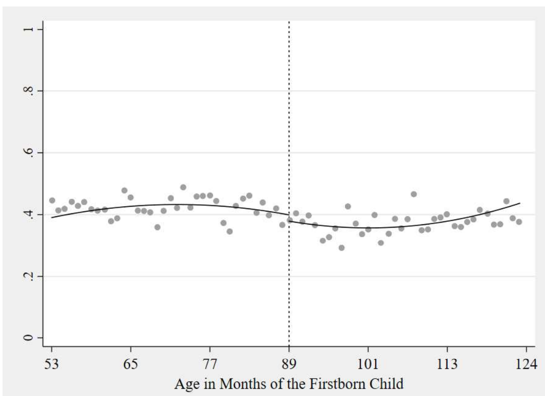
(a) Number of Children



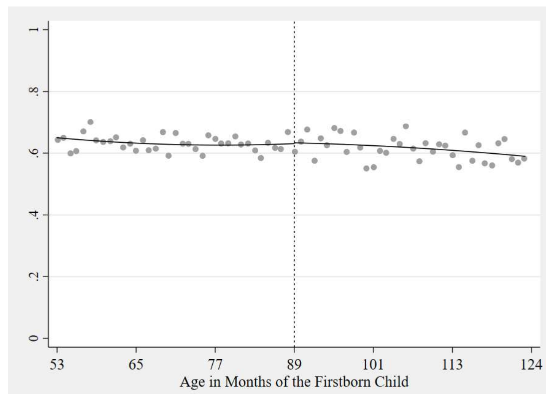
(b) Firstborn Child Is a Girl



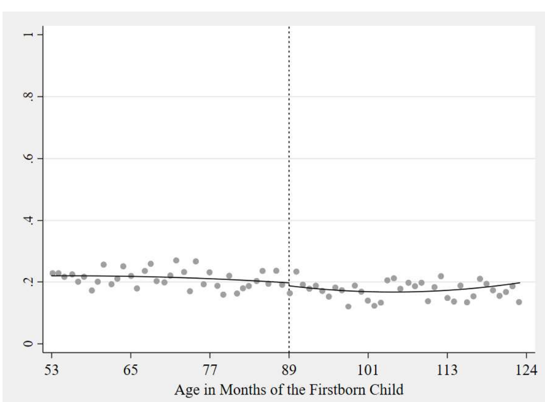
(c) 4-Year College Graduates



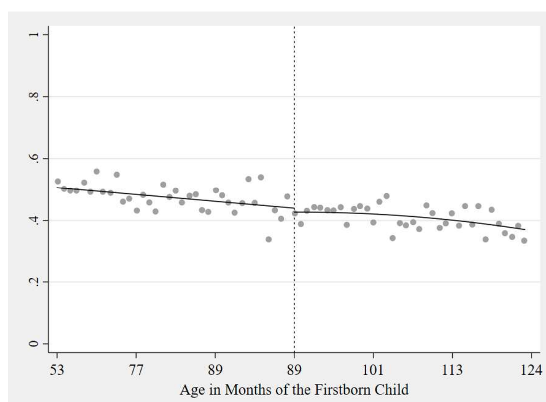
(d) Support from Grandparents Available (As of Feb.)



(e) Regular Worker (As of Feb. 2020)



(f) Not Working (As of Feb.2020)



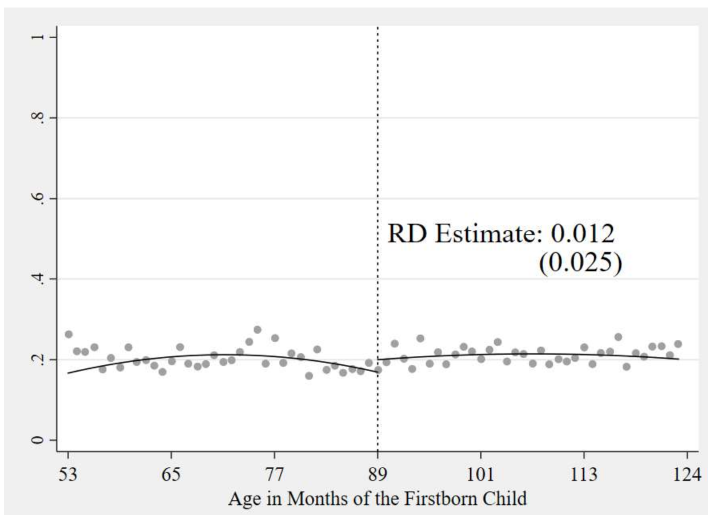
Notes: Observations are averaged within bins using the mimicking variance evenly-spaced method described in Calonico et al (2015). Each plot includes second-order global polynomial fits represented by the solid lines. Except for Figure B1(a), the y-axis represents the fraction of each category.

Since our running variable is the age of months, which is uncontrollable, originally, there should be no general manipulation problem around the threshold. However, it could be possible for the groups around the threshold to be different if mothers who had experienced school closures were more likely to feel uncomfortable in answering sensitive questions and tended to drop from the sample of the main survey. In that case, it would lead to underestimating the magnitude of the effect of the school closure because those seriously affected were not included in the sample.

Thus, we tested whether sample drops at the sensitive question were "truly" more likely to occur for mothers barely above the threshold, that is, those who experienced the sudden school closure in March 2020, which confirmed that there is no such trend in this subsection.

Figure B2 presents the result for the fraction of those who answered no to moving forward to the main survey. According to Figure B2, we do not find any statistically significant gap around the threshold.

Figure B2: Check for Continuity in Sample Drop



Notes: Observations are averaged within bins using the mimicking variance evenly-spaced method described in Calonico et al (2015). Each plot includes second-order global polynomial fits represented by the solid lines. The estimate reported inside the figure is a sharp-RD estimate obtained from the conventional local-linear regressions. Conventional heteroskedasticity-robust standard errors are reported in parenthesis. The CCT bandwidth selector proposed by Calonico et al (2014) is used to calculate the optimal bandwidth. The same bandwidth is applied to the areas below and above the cutoff. A triangular kernel function is used to construct the estimators. The selected optimal bandwidth is 8.288, and the number of observations within the bandwidth is 5,022. *** p<0.01, ** p<0.05, and * p<0.1.

Thus, we do not see any significant gap at the threshold in the fraction of sample drops caused by the existence of sensitivity questions in the main survey between the right and the left sides of the threshold, which does support the validity of our RD strategy.

Furthermore, because of limited space, we omitted presenting the results, however, there was also no significant difference between the two groups around the threshold in the sample drop among 17,860 samples who agreed to move on to the main survey to 15,836 samples who finished the main survey as well.

Note that the continuity test here was different from Figure B1 in the sense that the gap between the two groups in the sample drop potentially captured the difference in unobservable factors in the error term between the two groups, while Figure B1 captures the difference in observable factors. Thus, by showing the result of Figure B2 as well, we have confirmed that unobservable factors are thought to satisfy the continuous assumption too.

The results both from Figures B1 and B2 indicate that the people at the right-hand side of the threshold are considered to be a good counter-factual for those at the left-hand side, and vice versa.

C RD Estimation Results for Parents

Table C1: RD Estimates for the Impact of School Closures on Parents in August

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
		Sharp (Reduced Form)		**Fuzzy (IV)**			
Dependent Variable:	Mean of Dep. Var.	Conventional	Bias-corrected	Conventional	Bias-corrected	Optimal Bandwidth	N
Total Score of DVs: 10(Low) -30(High)							
	11.507	-0.048	-0.007	-0.075	-0.013	13.452	5466
		(0.202)	(0.239)	(0.316)	(0.375)		
Subjective Marital Satisfaction: 1(Not at all) -5(Very Satisfied)							
	3.474	0.011	-0.001	0.017	-0.001	13.698	5466
		(0.065)	(0.077)	(0.101)	(0.121)		
Total Score of Divorce-Risk Indexes: 4(Low Risk) -20(High Risk)							
	6.492	-0.121	-0.089	-0.190	-0.144	13.623	5466
		(0.191)	(0.228)	(0.299)	(0.357)		
Quality of Marriage Index: 6(Poor) -24(Excellent)							
	16.913	0.181	0.253	0.285	0.403	11.928	4728
		(0.310)	(0.369)	(0.488)	(0.580)		

Notes: Table C1 presents estimates from the conventional local-linear regressions as well as estimates to which the robust bias-corrected inference methods are applied. Conventional heteroskedasticity-robust standard errors are reported in parentheses. For the estimates from the robust bias-corrected inference methods, robust standard errors are reported. The CCT bandwidth selector proposed by Calonico et al (2014) is used to calculate the optimal bandwidth. The same bandwidth is applied to the areas below and above the cutoff. A triangular kernel function is used to construct the estimators.*** $p<0.01$, ** $p<0.05$, and * $p<0.1$.

Table C2: RD Estimates for the impacts of School Closures on DVs

Dependent Variable:	(1)	(2)	(3)	(4)	(5)	(6)	(7)
		Sharp (Reduced Form)		**Fuzzy (IV)**			
Total Scores of Each DV Dummy Variable	Mean of Dep. Var.	Conventional	Bias-corrected	Conventional	Bias-corrected	Optimal Bandwidth	N
Panel A. August							
•More Than Once = 1	1.217	-0.055	-0.021	-0.086	-0.035	13.171	5466
		(0.141)	(0.167)	(0.221)	(0.261)		
•Frequently = 1	0.283	0.008	0.016	0.013	0.026	12.107	5116
		(0.086)	(0.103)	(0.135)	(0.162)		
Panel B. March							
•More Than Once = 1	1.265	0.187	0.220	0.296	0.353	9.583	4003
		(0.175)	(0.207)	(0.276)	(0.328)		
•Frequently = 1	0.341	0.126	0.109	0.200	0.175	10.248	4390
		(0.103)	(0.123)	(0.164)	(0.195)		

Notes: Table C2 presents estimates from the conventional local-linear regressions as well as estimates to which the robust bias-corrected inference methods are applied. Conventional heteroskedasticity-robust standard errors are reported in parentheses. For the estimates from the robust bias-corrected inference methods, robust standard errors are reported. The CCT bandwidth selector proposed by Calonico et al (2014) is used to calculate the optimal bandwidth. The same bandwidth is applied to the areas below and above the cutoff. A triangular kernel function is used to construct the estimators.*** $p<0.01$, ** $p<0.05$, and * $p<0.1$. The total score of DVs is 10 at maximum because we asked 10 questions on DVs, and here we used dummies for each item of DV. 10 DV items consisted of five elements (e.g., "ignoring" and "hitting") and who did it (i.e., wife or husband). We measured the frequency of DVs in three categories (i.e., Never, Sometimes, and Frequently). Thus, for the results of "More Than Once = 1," we counted the number of DVs in which the respondent chose "Sometimes" or "Frequently." For the results of "Frequently = 1," we counted the number of DVs in which the respondent chose "Frequently" only.

D Robustness Checks

Table D1: Robustness Check Using Local-quadratic Specification

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
		Sharp (Reduced Form)		Fuzzy (IV)			
Dependent Variable: 1 (Yes) or 0 (No)	Mean of Dep. Var.	Conventional	Bias-corrected	Conventional	Bias-corrected	Optimal Bandwidth	N
My child gained weight							
	0.152	0.105***	0.109***	0.176***	0.187***	10.845	4390
		(0.034)	(0.040)	(0.058)	(0.068)		
I began to worry about how to raise my child more frequently							
	0.226	0.132***	0.132***	0.222***	0.229***	10.392	4390
		(0.042)	(0.050)	(0.073)	(0.086)		
I began to worry about my relationship with my child more frequently							
	0.158	0.095***	0.097**	0.160**	0.169**	10.277	4390
		(0.037)	(0.043)	(0.063)	(0.074)		
I began to leave my child home alone for a longer period of time (per day)							
	0.072	0.053***	0.058***	0.085***	0.094***	16.613	6681
		(0.019)	(0.021)	(0.031)	(0.034)		

Notes: Table D1 presents estimates from the conventional local-linear regression and those from local-linear regressions with robust bias-corrected confidence intervals and inference procedures following the approach developed in Calonico et al (2014, 2020). Conventional heteroskedasticity-robust standard errors are reported in parentheses. For the estimates from the robust bias-corrected inference methods, robust standard errors are reported. The CCT bandwidth selector proposed by Calonico et al (2014) is used to calculate the optimal bandwidth. The same bandwidth is applied to the areas below and above the cutoff. A triangular kernel function is used to construct the estimators.*** p<0.01, ** p<0.05, and * p<0.1.

Table D2: Robustness Check Using Another Type of Bandwidth Selector

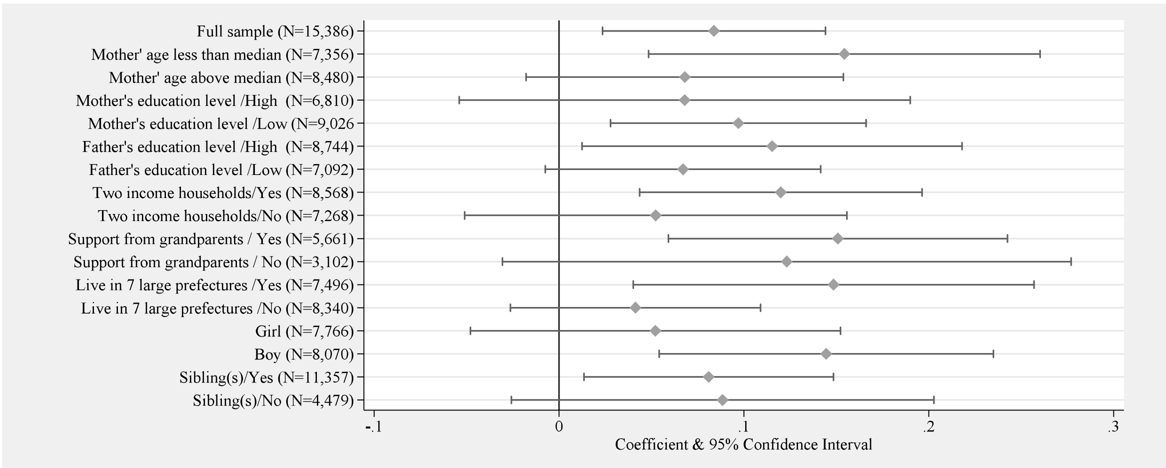
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
		Sharp (Reduced Form)		Fuzzy (IV)			
Dependent Variable: 1 (Yes) or 0 (No)	Mean of Dep. Var.	Conventional	Bias-corrected	Conventional	Bias-corrected	Optimal Bandwidth	N
My child gained weight							
	0.153	0.099*	0.101***	0.162*	0.166***	7.134	3189
		(0.028)	(0.030)	(0.047)	(0.050)		
I began to worry about how to raise my child more frequently							
	0.222	0.135***	0.141***	0.231***	0.241***	4.813	1882
		(0.043)	(0.044)	(0.074)	(0.077)		
I began to worry about my relationship with my child more frequently							
	0.150	0.079**	0.084**	0.130**	0.139**	6.024	2730
		(0.032)	(0.033)	(0.054)	(0.056)		
I began to leave my child home alone for a longer period of time (per day)							
	0.073	0.060***	0.062***	0.099***	0.103***	6.103	2730
		(0.022)	(0.023)	(0.036)	(0.037)		

Notes: Table D2 presents estimates from the conventional local-linear regression and those from local-linear regressions with robust bias-corrected confidence intervals and inference procedures following the approach developed in Calonico et al (2014, 2020). Conventional heteroskedasticity-robust standard errors are reported in parentheses. For the estimates from the robust bias-corrected inference methods, robust standard errors are reported. A bandwidth choice that focuses on delivering confidence intervals with optimal coverage error rates proposed by Calonico et al (2018) is used to calculate the optimal bandwidth. The same bandwidth is applied to the areas below and above the cutoff. A triangular kernel function is used to construct the estimators.*** p<0.01, ** p<0.05, and * p<0.1.

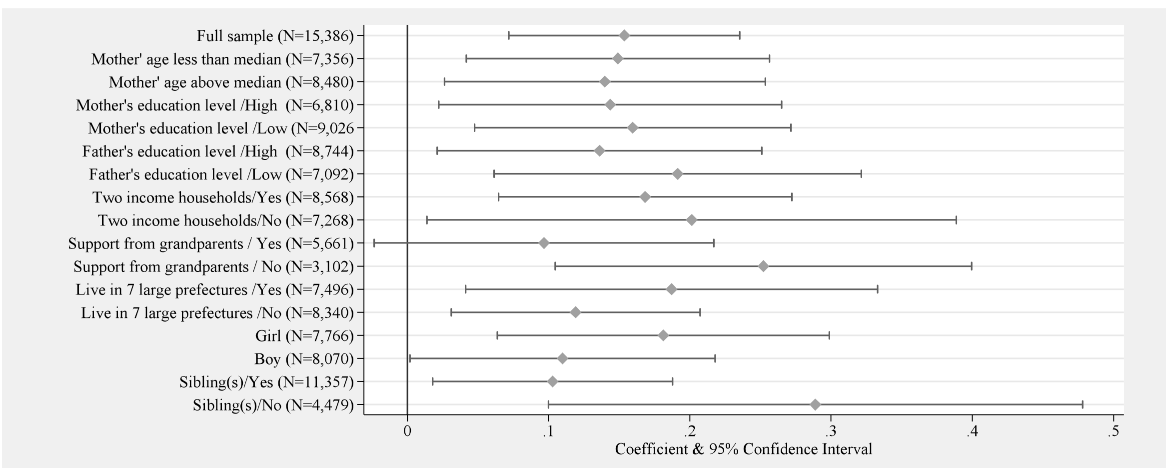
E Heterogeneous Effects

Figure E1: Heterogeneity: Changes of Children's Outcomes and Daily Lives Due to COVID-19

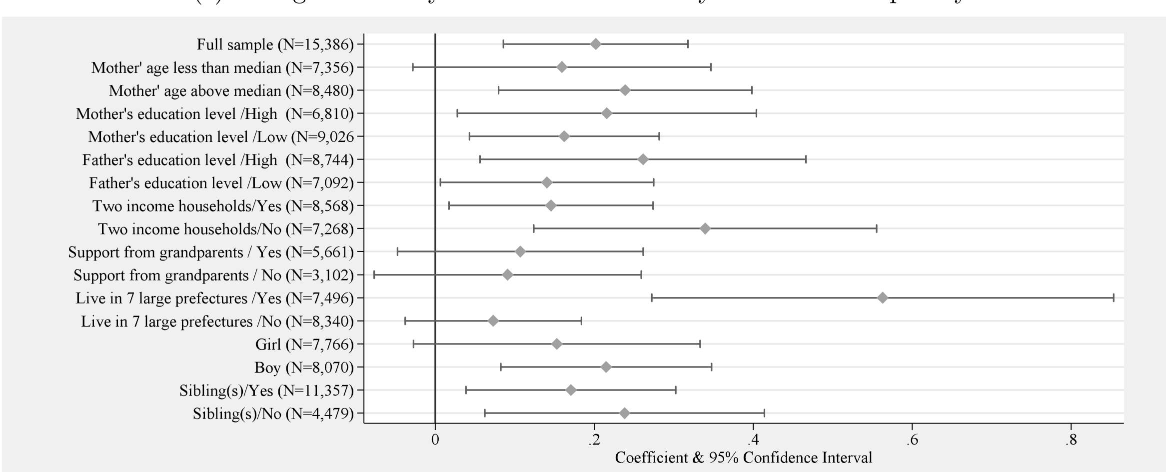
(a) "I began to leave my child home alone for a longer period of time (per day)."



(b) "My child gained weight."



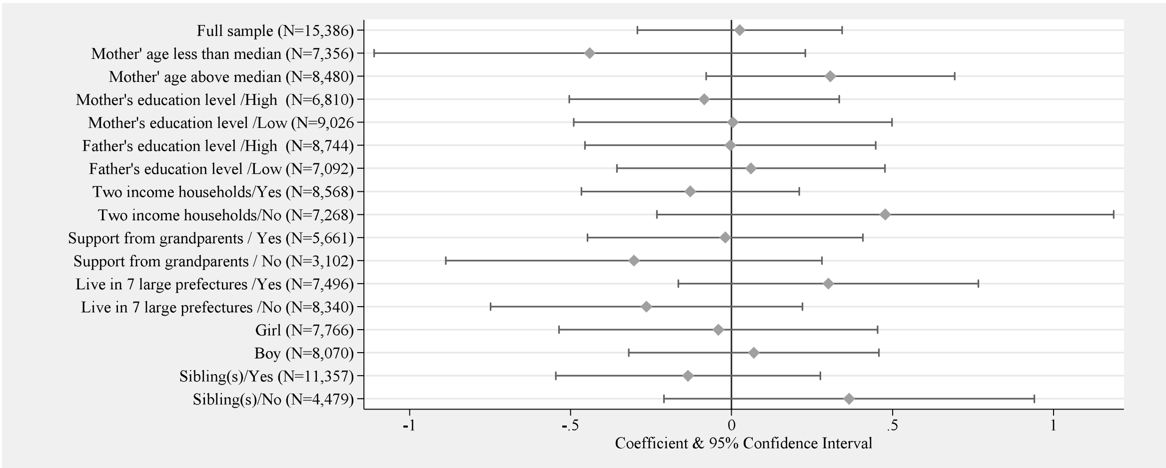
(c) "I began to worry about how to raise my child more frequently."



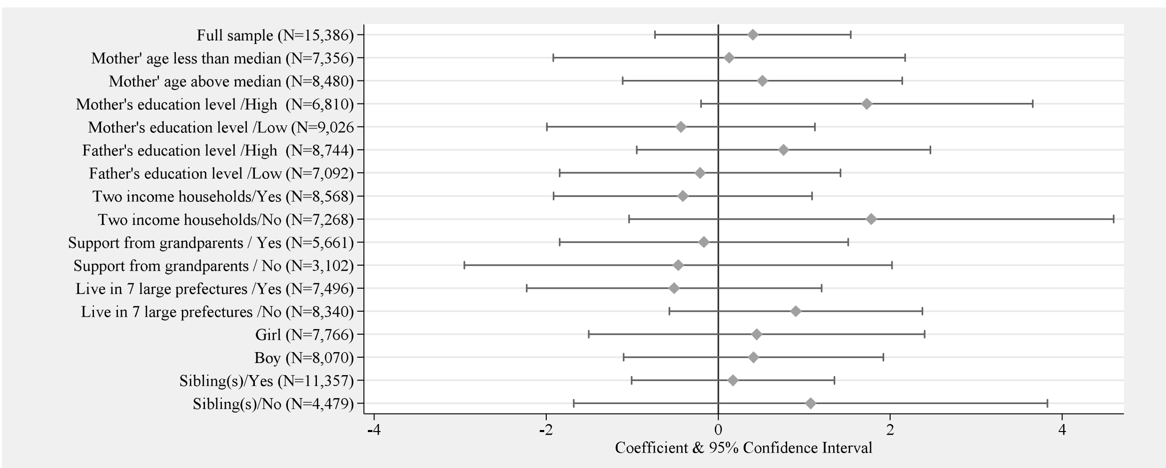
Notes: See the notes of Figure E2.

Figure E2: Heterogeneity: Marital Relationships

(a) Total DV Score (Frequently=1)



(b) Quality of Marriage Index



Notes: Fuzzy RD estimate and a 95% confidence interval are plotted according to subsamples that are explained in the label of the vertical axis. Fuzzy RD estimate is obtained from a local-linear regression with robust bias-corrected confidence intervals. The bias-corrected coefficient and a standard derived error from a robust variance estimator are reported (Calonico et al, 2014). The CCT bandwidth selector proposed by Calonico et al (2014) is used to calculate the optimal bandwidth. Heteroskedasticity-robust standard errors are used to construct a 95% confidence interval. For the results of "Frequently = 1," we counted the number of DVs in which the respondent chose "Frequently" only.