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

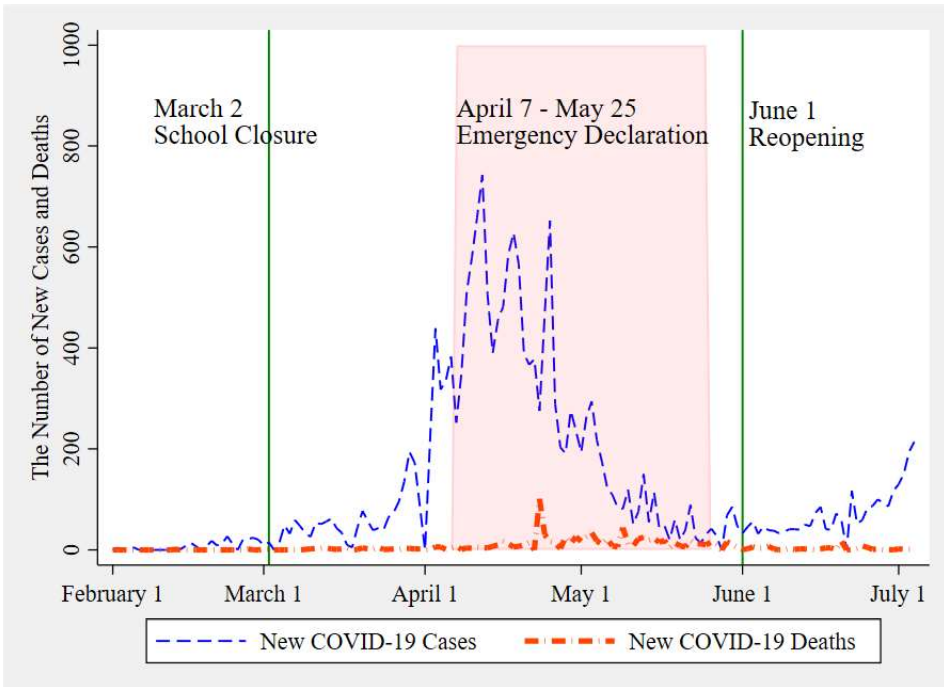
Notes:
a. The majority of states mandated school closures, including until the end of the academic year. Some states, however, recommended but did not mandate school closures.
b. On May 11, the government announced school closures until the end of the academic year. Classes continued through distance 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%

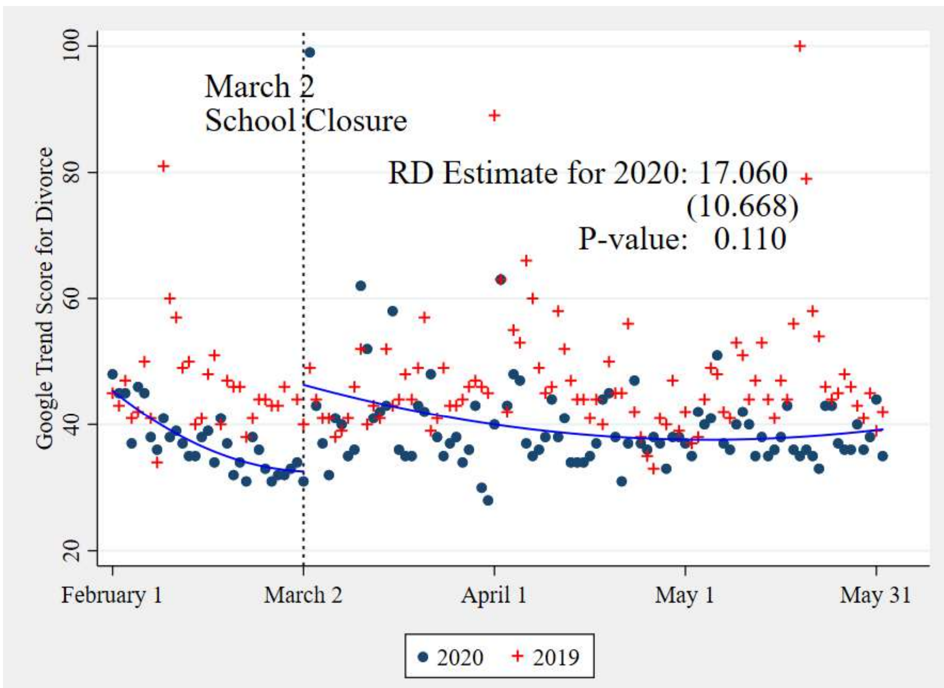
Notes: Rows shaded in gray represent the date when the major event occurred. Chiba, Saitama, Tokyo, Kangawa, Osaka, Hyogo, and Fukuoka represent the seven prefectures where the state of emergency was declared on April 7. The data in Column (1) does not include daycare centers.
Source: (MEXT, 2020b)

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

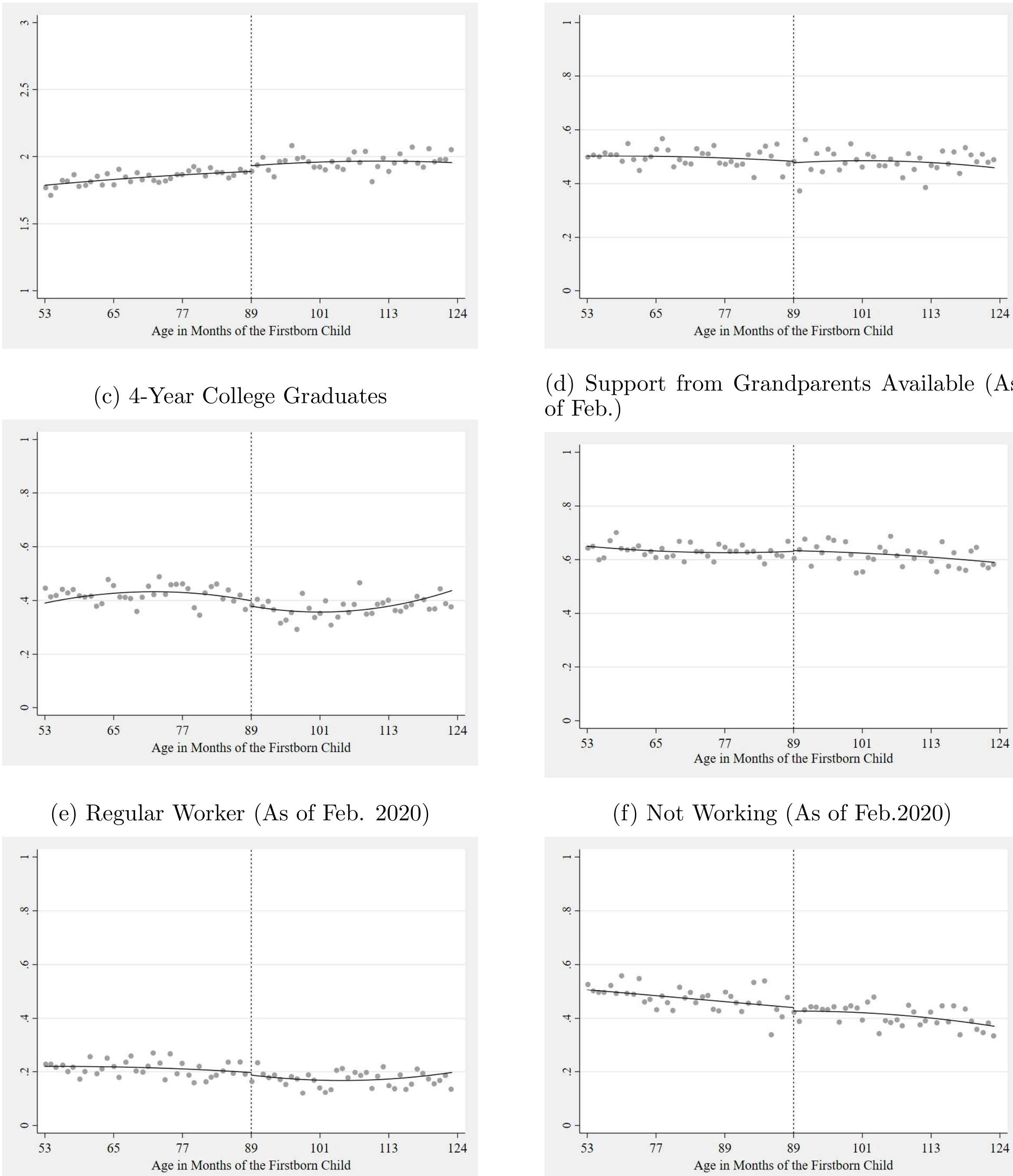
As already explained in the data section, before moving to the main survey, respondents were required to answer a question about their willingness to participate in the main survey after having been explained that the main survey included some sensitive questions, such as questions about mental health 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)

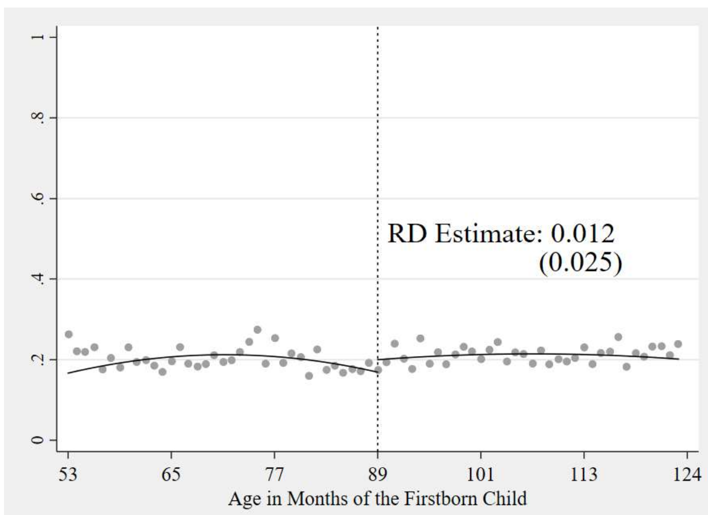
Notes: Observations are averaged within bins using the mimicking variance evenly-spaced method described in Calonico et al (2015). Each plot includes second-order global polynomial fits represented by the solid lines. Except for Figure B1(a), the y-axis represents the fraction of each category.

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 under-estimating 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

Dependent Variable:	(1) Mean of Dep. Var.	(2) Sharp (Reduced Form) Conventional	(3) Sharp (Reduced Form) Bias-corrected	(4) Fuzzy (IV) Conventional	(5) Fuzzy (IV) Bias-corrected	(6) Optimal Bandwidth	(7) 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: Total Scores of Each DV Dummy Variable	(1) Mean of Dep. Var.	(2) Sharp (Reduced Form) Conventional	(3) Sharp (Reduced Form) Bias-corrected	(4) Fuzzy (IV) Conventional	(5) Fuzzy (IV) Bias-corrected	(6) Optimal Bandwidth	(7) 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

Dependent Variable: 1 (Yes) or 0 (No)	(1) Mean of Dep. Var.	(2) Sharp (Reduced Form) Conventional	(3) Sharp (Reduced Form) Bias-corrected	(4) Fuzzy (IV) Conventional	(5) Fuzzy (IV) Bias-corrected	(6) Optimal Bandwidth	(7) 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

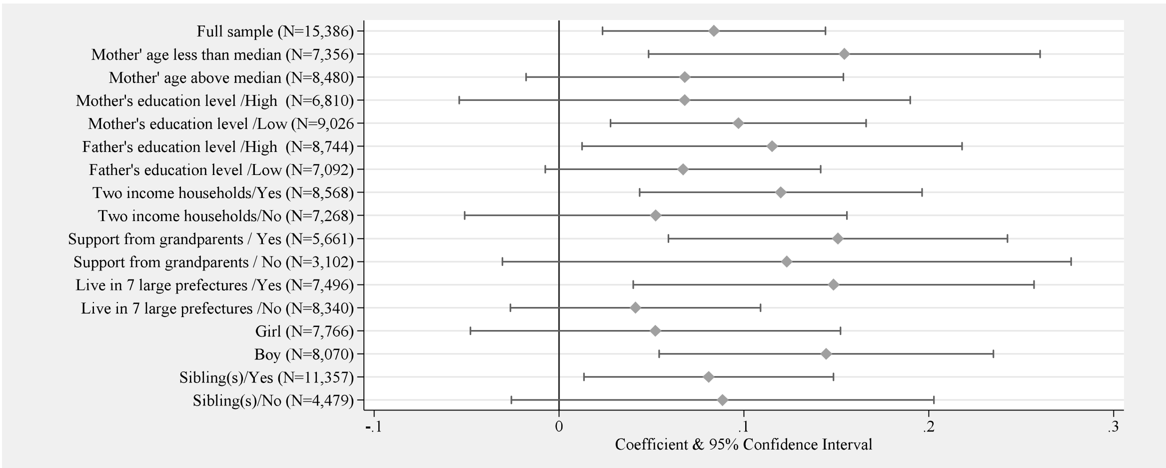
Dependent Variable: 1 (Yes) or 0 (No)	(1) Mean of Dep. Var.	(2) Sharp (Reduced Form) Conventional	(3) Sharp (Reduced Form) Bias-corrected	(4) Fuzzy (IV) Conventional	(5) Fuzzy (IV) Bias-corrected	(6) Optimal Bandwidth	(7) N
My child gained weight							
	0.153	0.099*	0.101***	0.162*	0.166***	7.134	3189
		(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.

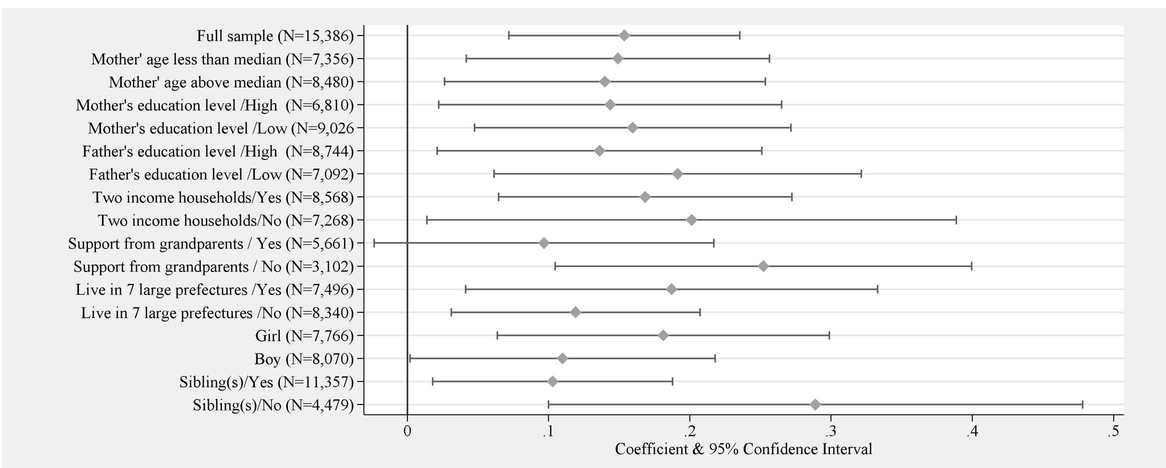
E Heterogeneous Effects

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

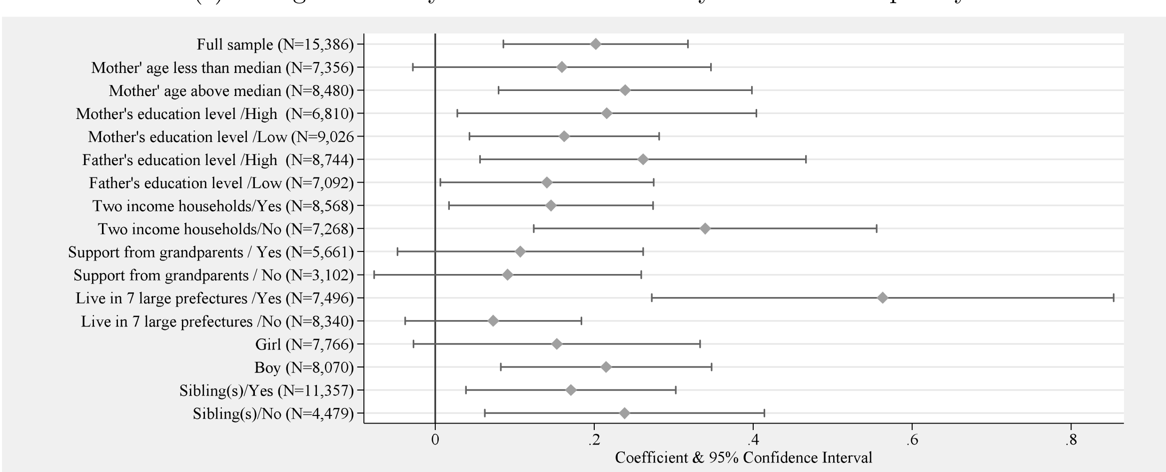
(a) "I began to leave my child home alone for a longer period of time (per day)."



(b) "My child gained weight."



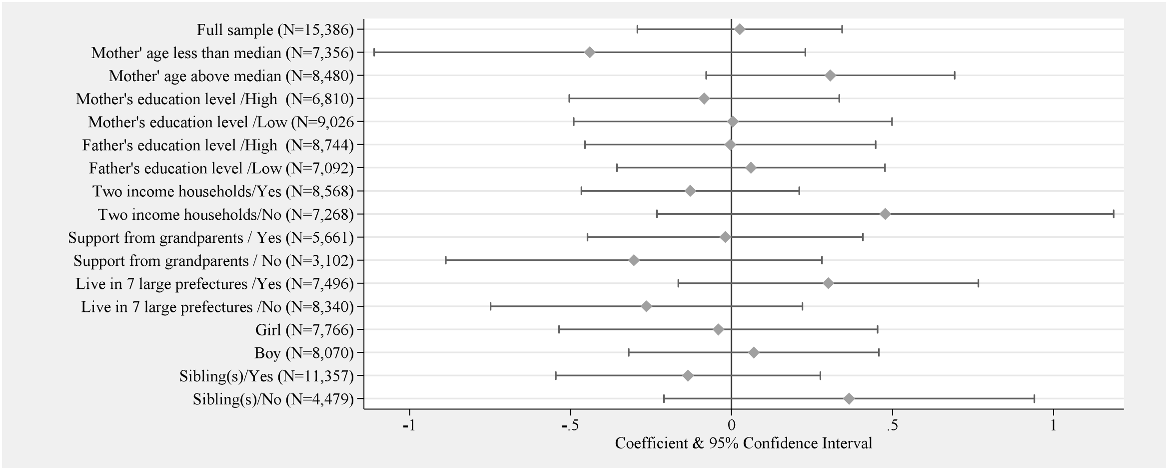
(c) "I began to worry about how to raise my child more frequently."



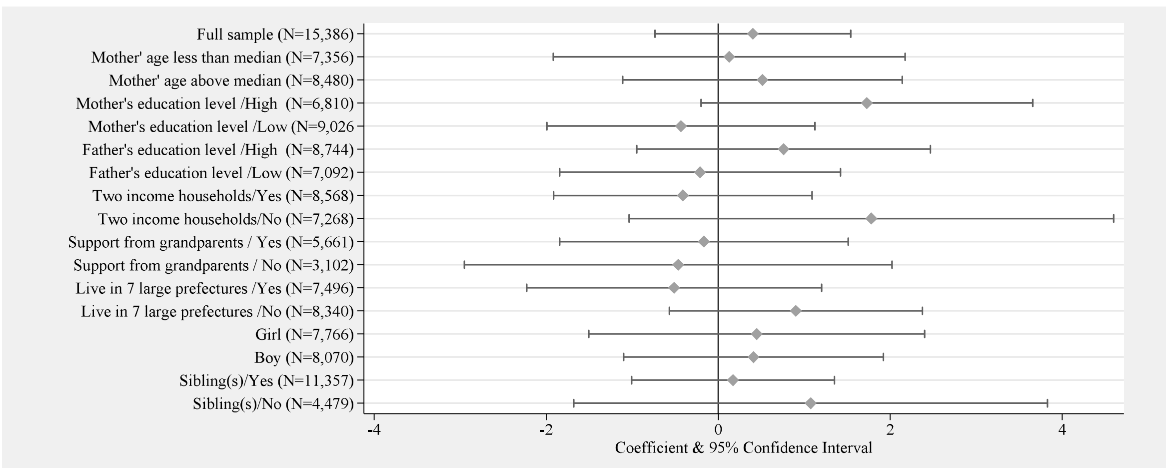
Notes: See the notes of Figure E2.

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.