INSTITUTE for REPLICATION

No. 16
I4R DISCUSSION PAPER SERIES

Direct Replication and Additional Sensitivity Analyses for Altindag et al. (2022): A Replication Report from the Oslo Replication Games

Carl Bonander

Gabriella Chauca Strand

Niklas Jakobsson

February 2023

I4R Discussion Paper Series

I4R DP No. 16

Direct Replication and Additional Sensitivity Analyses for Altindag et al. (2022): A Replication Report from the Oslo Replication Games

Carl Bonander[1], Gabriella Chauca Strand[1], Niklas Jakobsson[2]

[1]*University of Gothenburg/Sweden*
[2]*Karlstad University/Sweden*

February 2023

ISSN: 2752-1931

Direct replication and additional sensitivity analyses for Altindag et al. (2022): A replication report from the Oslo Replication Games

Carl Bonander (University of Gothenburg), Gabriella Chauca Strand (University of Gothenburg), and Niklas Jakobsson (Karlstad University) *

Abstract

This report presents a replication of Altindag et al. (2022) performed at the Olso Replication Games in 2022. Altindag et al. (2022) estimate the effects of an age-specific lockdown on mental health outcomes and mobility among adults aged 65 and older in Turkey, using a regression discontinuity design. The authors find a decline in mobility with a one-day decrease in the number of days being outside and an increase in the probability of never going out by 30 percentage points. These point estimates are statistically significant at the 1% level. The mobility restrictions lead to a worsening in mental health outcomes of approximately 0.2 standard deviations, statistically significant at the 10% level in their preferred specification. In this paper we accomplish two things. First, we successfully reproduce Altindag et al.'s main findings. Second, we test the robustness of the results to a small number of changes to their preferred estimations by (1) not clustering the standard errors on the running variable, (2) not including control variables, and (3) calculating the optimal bandwidth using another technique. Point estimates for mobility outcomes are stable to all three manipulations, and standard errors only change marginally. Point estimates and standard errors for the mental health outcomes are somewhat more sensitive, especially to changing the optimal bandwidth selection method. However, the observed changes are reasonably expected when applying data-driven model selection methods to noisy data (to avoid overfitting, it is likely preferable to apply a less data-driven approach like the original authors did). Our general impression is that the original analyses and results are both theoretically plausible and credible, despite some defensible model dependencies.

* Carl Bonander (University of Gothenburg, Sweden), carl.bonander@gu.se. Gabriella Chauca Strand (University of Gothenburg, Sweden), gabriella.chauca.strand@gu.se. Niklas Jakobsson (Karlstad Business School, Karlstad University, Sweden). Corresponding author, niklas.jakobsson@kau.se. The authors did not receive any financial support and report no conflict of interest.

1. Introduction

Altindag et al. 2022 estimate the impact of an age-specific lockdown and stay-at-home order on mental health outcomes and mobility among the elderly in Turkey. Between March- and June 2020, the Turkish government enforced a curfew on all adults aged 65 years or older. The curfew was a stay-at-home order with financial penalties imposed for noncompliance. Using the age cutoff of the curfew, the authors adopt a regression discontinuity design, comparing individuals just below and above the age cutoff to estimate the causal impact of the curfew on mental health outcomes and mobility. Data were collected through a phone survey conducted on individuals aged 59-70 between May and July 2020. The survey was based on a validated questionnaire, SRQ-20, as well as additional questions developed by the authors. The main results on mobility are described on p. 331 and indicate an effect of a 1 – 1.1 day decrease in the “number of days outside last week” as well as a 30-percentage point increase in the probability of never going out during the period of lockdown. The estimates were all statistically significant at the 5% significance level. The main results on mental health outcomes are described on pp. 334-335 as follows:

> “Remarkably, the RD estimates show a substantial impact of the curfew on all measures of mental distress: the first-row estimates imply a 0.21 standard deviation increase in the mental distress index. We estimate similar effects for the more objective measure of depression – the somatic symptom index (0.18 standard deviation) … The corresponding effect size of the nonsomatic symptom index is an increase of 0.16 standard deviation. Finally, the RD estimates indicate that the curfew had a positive impact of 0.7 on the sum of “yes” answers in SRQ-20 inventory reported to the respondents”.

The analyses were conducted using clustered standard errors based on the month-year of birth and the selection of the optimal bandwidth was made based on Imbens and Kalyanaraman (2012), as described on p.329. The authors conduct robustness checks over several different bandwidths and control for multiple testing.

This report presents a replication of Altindag et al. (2022) conducted at the Olso Replication Games in 2022. The purpose of the replication exercise was to (i) verify that we could reproduce all primary analyses in the paper using the original authors’ codes and instructions, and (ii) make reasonable changes to some modelling choices while, per instructions, keeping others constant to assess the sensitivity of the primary estimates to model specification. The scope of our analyses is not to test the robustness of their finding to all potential manipulations, but to by simple means find out how sensitive the main results are to three straight-forward and reasonable estimations that we would have likely performed if we had done the study ourselves. With primary analyses, we here refer to the results that are brought up in the abstract and the results that are highlighted in the text in the results section.[1]

Following our own usual approach to regression discontinuity estimation and motivated by recent methods papers, we determined that three modelling choices could be interesting assess further: Specifically, we assess the robustness of the results to (1) not clustering the standard errors on the month-year of birth (using Eicker-Huber-White heteroscedasticity-robust standard errors instead) as clustering may overstate statistical certainty due to small sample bias with discrete running variables, especially with small bandwidths (few clusters) close to the cutoff (Kolésar & Rothe, 2018); (2) not including control variables, which we believe is a sensible default to present even though we recognize that including covariates may improve statistical power; (3) calculating the optimal bandwidth using a technique that conducts estimation, bandwidth selection, and bias-corrected statistical inference simultaneously (using the default settings of rdrobust in Stata; Calonico, Cattaneo and Titiunik, 2014a), using both a uniform kernel and a triangular kernel to calculate the optimal bandwidth (Calonico, Cattaneo and Titiunik, 2014b). Without bias-correction, the bias-variance tradeoff involved in data-driven bandwidth selection may lead to biased inferences (Calonico, Cattaneo and Farrell, 2020).

We successfully reproduced the main tables (3 and 4) in Altindag et al. (2022) using their codes and data, made available as a replication package. We uncovered one

[1] The suggested ways to select claims to replicate during the Replication Games were: “(1) select claims for all "hypotheses tests" in the original study, (2) select claims mentioned in the abstract or (3) select claims for what is considered the main result in the paper as stated by the original author(s)”. We opted for a combination of (2) and (3), focusing on mobility and mental health outcomes, which we interpret to be the study’s main findings (although we note that the abstract also contains results about potential channels). In the text, Altindag et al. (2022) primarily focus on the results for the mean-squared-error-optimal estimated using the first-stage outcome “outside last week”, which is ±45 months from the cutoff (Altindag et al., 2022, pp. 330). We therefore focus on these results in our report. However, we note that they consistently present results for other cutoffs in their tables; a more comprehensive robustness check could also include these.

minor error in the survey questionnaire (Appendix C), where one question (about unhappiness) in the SRQ-20 instrument was not included. We contacted the authors, who told us that this was an error in the Appendix, and that the question was in fact used in constructing the index used in the article. In our sensitivity analyses, we find that the point estimates for mobility outcomes are stable to all three manipulations. The mental health outcomes are somewhat more sensitive to our modelling choices, but the changes are reasonably expected given the relative amount of noise in these outcomes when combined with the alternative methods we use.

2. Reproducibility

In this section, we describe one minor error that we uncovered while reproducing the study. We noticed one minor error in the survey questionnaire (Appendix C), where one question (about unhappiness) in the SRQ-20 instrument was not included. After contacting the authors, we got the information that this was an error in the Appendix, and that the question was in fact used in constructing the index used in the article.

3. Replication

We now turn our attention to our replication, and test the robustness of the results to not clustering the standard errors, excluding control variables from the model, and changing the method of calculating the optimal bandwidth. We use Eicker-Huber-White heteroscedasticity-robust standard errors instead of clustering the standard errors on the month-year of birth (i.e., the running variable). We calculate the optimal bandwidth and perform bias-corrected estimation using another technique (using the default settings of rdrobust in Stata), using both uniform kernel and triangular kernels to weight observations away from the cutoff. A uniform kernel weights all observations equally, whereas a triangular kernel down weights observations linearly away from the cutoff until the end of the selected bandwidth. Altindag et al. (2022) present results using a uniform kernel, but mention in a footnote that the results were similar with a triangular kernel (footnote 30, p. 329). Triangular kernels are the default option in rdrobust for Stata (Calonico, Cattaneo and Titiunik, 2014a). Unlike Altindag et al. (2022), we also perform the rdrobust bandwidth selection for each outcome and model instead of performing the bandwidth selection on one outcome, as the mean-squared-error-optimal bandwidth depends on both the outcome and model specification.

3.1 Regression model

For our analysis, we rely on the same specifications and a sharp regression discontinuity design comparing those just above the age threshold to those just below the age threshold. See the original study for more details and equations.

3.2 Results on mobility outcomes

3.2.1 Clustering

We first investigate whether changing the clustering technique affects the estimated precision of the mobility outcome estimates. Our findings are reported in Table 1. The dependent variables are days outside last week, under curfew, and never goes out, as defined in Altindag et al. (2022). Column 1 reports the preferred estimates from the original study using individuals born within 45 months of the age threshold, December 1955, and is a reproduction of the results presented in Table 3, Column 3 in Altindag et al. (2022). Column 2 is the same estimation without clustering the standard errors on the month-year cohort level (using Eicker-Huber-White heteroscedasticity-robust standard errors instead). The specifications include month fixed effects, province fixed effects, surveyor fixed effects, as well as indicator variables for education levels, ethnicity, and gender. Corresponding p-values and Anderson's (2008) sharpened q-values are presented, as in Altindag et al. (2022). We find that the standard errors for the point estimates (Column 2) only change marginally.

3.2.2 Removing covariates

We then investigate whether not controlling for any other variables affects the point estimates and precision of the regression discontinuity model. Our findings are reported in Table 2. Column 1 reports the preferred estimates from the original study, including month fixed effects, province fixed effects, surveyor fixed effects, as well as indicator variables for education levels, ethnicity, and gender as covariates. Column 2 is the same estimation without including the above control variables. We find that the point estimates for the preferred specifications (Column 2) and that the standard errors only change marginally.

3.2.2 Method for bandwidth selection

Finally, we investigate whether calculating the optimal bandwidth using another technique (rdrobust in Stata), using both a uniform kernel and a triangular kernel to calculate the optimal bandwidth, affects the point estimates and statistical significance of the model. Our findings are reported in Table 3. Column 1 reports the preferred estimates from the original study. Column 2 uses rdrobust and a uniform kernel to calculate the optimal bandwidth, Column 3 uses rdrobust and a triangular kernel to calculate the optimal bandwidth. Again, we find that point estimates do not change much and that the standard errors only change marginally.

3.3 Results on mental health outcomes

3.3.1 Clustering

In this subsection, we turn to the second set of outcome variables, mental health outcomes. Our findings are reported in Table 4. The dependent variables are mental distress index, somatic symptoms of distress index, non-somatic symptoms of distress index, and sum of yes answers in SRQ-20, as defined in Altindag et al. (2022). Column 1 reports the preferred estimates from the original study using individuals born within 45 months of the age threshold, December 1955, and is a reproduction of the results presented in Table 4, Column 3 in Altindag et al. (2022). Column 2 is the same estimation without clustering the standard errors on month-year cohort level (using Eicker-Huber-White heteroscedasticity-robust standard errors instead). The specifications include month fixed effects, province fixed effects, surveyor fixed effects, as well as indicator variables for education levels, ethnicity, and gender. Corresponding p-values and Anderson's (2008) sharpened q-values are presented, as in Altindag et al. (2022). We find that the standard errors for the point estimates (Column 2) increase somewhat.

3.3.2 Removing covariates

We then investigate whether not controlling for any other variables affects the point estimates and statistical significance of the regression discontinuity model. Our findings are reported in Table 5. Column 1 reports the preferred estimates from the original study, including month fixed effects, province fixed effects, surveyor fixed effects, as well as indicator variables for education levels, ethnicity, and gender as covariates. Column 2 is the same estimation without including the above control variables. We find that the point estimates for the preferred specifications (Column 2) change somewhat and that the standard errors generally increase (which is expected given that covariates can improve power).

3.3.3 Methods for bandwidth selection

Finally, we investigate whether calculating the optimal bandwidth using another technique (rdrobust in Stata), using both a uniform kernel and a triangular kernel to calculate the optimal bandwidth, affects the point estimates and precision of the model. Our findings are reported in Table 6. Column 1 reports the preferred estimates from the original study. Column 2 uses rdrobust and a uniform kernel to calculate the optimal bandwidth, Column 3 uses rdrobust and a triangular kernel to calculate the optimal bandwidth. We find that point estimates change a lot (both up and down), as do the standard errors, indicating noisiness in the data.

4. Conclusion

We test the robustness of the results in Altindag et al. (2022) to a small number of changes to their preferred estimations by (1) not clustering the standard errors on the running variable, (2) not including control variables, and (3) calculating the optimal bandwidth using another technique. Even though these manipulations do not give a full picture of the robustness of the findings, they are reasonable, straightforward, and simple to implement, and we would typically prefer the main results to be stable to these manipulations to judge the findings as robust to model selection.

We find that point estimates for mobility outcomes are robust to all three manipulations, and that standard errors only change marginally. The point estimates and estimated precision of the mental health estimates appear to be more sensitive. However, in defense of the original analyses, we note that the identified sensitivities may have natural explanations and do likely not reflect that the original results are wrong. For instance, using covariates may increase power if they are strongly predictive of the outcome, and the variance in the mental health outcomes is generally higher. Thus, it may be preferable to include observed pre-treatment covariates to increase power, as is done in the original paper. Overall, we think that the sensitivity in the mental health results may be due to the relatively high variance in these outcomes compared to the mobility outcomes, and applying the rdrobust algorithm to these data may result in overfitting compared to a less data-driven approach. Per instructions for the replication exercise, we only changed one aspect per analysis, which meant that we kept the clustering on the running variable when performing the bandwidth selection using the rdrobust algorithm. We note that the model selection in this program depends on the specification, including the standard error computation, and applying clustering on the running variable in combination with noisy data may cause the algorithm to select a bandwidth that is much too small and further exacerbate problems related to overfitting. With this in mind, it would be reasonable to prefer the approach in Altindag et al. (2022), which is less data-driven.

In conclusion, we commend Altindag et al. (2022) on their timeliness in collecting important data during the lockdown period. Despite some (defensible) model dependencies identified in the small number of sensitivity analyses we have run here, our general impression is that the original analyses and results are both theoretically plausible and credible.

References

Altindag, O., Erten, B. and Keskin, P. (2022). Mental health costs of lockdowns: evidence from age-specific curfews in Turkey. *American Economic Journal: Applied Economics*, 14(2), 320-343.

Calonico, Cattaneo and Titiunik (2014a): Robust Data-Driven Inference in the Regression-Discontinuity Design. *Stata Journal* 14(4): 909-946.

Calonico, Cattaneo and Titiunik (2014b): Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. *Econometrica* 82(6): 2295-2326.

Calonico, Cattaneo and Farrell (2020): Optimal Bandwidth Choice for Robust Bias Corrected Inference in Regression Discontinuity Designs. *Econometrics Journal* 23(2): 192-210.

Kolesár, M., & Rothe, C. (2018). Inference in regression discontinuity designs with a discrete running variable. *American Economic Review*, 108(8), 2277-2304.

Tables

Table 1 – Changing clustering (Mobility outcomes)

	Original Study **(1)**	**Without clustering** **(2)**
Days outside last week	-1.090	-1.090
Standard errors	0.180	0.189
p-value	0.000	0.000
q-value	0.001	0.001
Bandwidth	±45	±45
Under curfew	0.708	0.708
Standard errors	0.045	0.039
p-value	0.000	0.000
q-value	0.001	0.001
Bandwidth	±45	±45
Never goes out	0.297	0.297
Standard errors	0.036	0.051
p-values	0.000	0.000
q-values	0.001	0.001
Bandwidth	±45	±45

Note: Estimates of the effect of being born before December 1955 on the mobility outcomes of individuals. The variable descriptions are provided in Altindag et al. (2022). The first column presents results for individuals born within 45 months of the age threshold, December 1955, and is a reproduction of the results presented in Table 3, Column 3 in Altindag et al. (2022). The second column is the same estimation without clustering the standard errors on month-year cohort level. The specification includes month fixed effects, province fixed effects, surveyor fixed effects, as well as indicator variables for education levels, ethnicity, and gender. Corresponding p -values and Anderson's (2008) sharpened q –values are presented, as in Altindag et al. (2022).

Table 2 – Removing control variables (Mobility outcomes)

	Original Study (1)	Without covariates (2)
Days outside last week	-1.090	-0.963
Standard errors	0.180	0.190
p-value	0.000	0.000
q-value	0.001	0.001
Bandwidth	±45	±45
Under curfew	0.708	0.696
Standard errors	0.045	0.040
p-value	0.000	0.000
q-value	0.001	0.001
Bandwidth	±45	±45
Never goes out	0.297	0.281
Standard errors	0.0360	0.037
p-values	0.000	0.000
q-values	0.001	0.001
Bandwidth	±45	±45

Note: Estimates of the effect of being born before December 1955 on the mobility outcomes of individuals. The variable descriptions are provided in Altindag et al. (2022). The first column presents results for individuals born within 45 months of the age threshold, December 1955, and is a reproduction of the results presented in Table 3, Column 3 in Altindag et al. (2022). The specification includes month fixed effects, province fixed effects, surveyor fixed effects, as well as indicator variables for education levels, ethnicity, and gender. The second column is the same estimation without including the above control variables. Corresponding p-values and Anderson's (2008) sharpened q-values are presented, as in Altindag et al. (2022).

Table 3 – Changing bandwidth selection/estimation (Mobility outcomes)

	Original Study (1)	RD robust uniform (2)	RD robust triangular (3)
Days outside last week	-1.089	-1.019	-1.087
Standard errors	0.179	0.202	0.168
p-value	0.000	0.000	0.000
q-value	0.001	0.001	0.001
Bandwidth	±45	±9.478	±17.139
Under curfew	0.708	0.531	0.588
Standard errors	0.044	0.035	0.031
p-value	0.000	0.000	0.000
q-value	0.001	0.001	0.001
Bandwidth	±45	±11.440	±16.666
Never goes out	0.297	0.297	0.233
Standard errors	0.036	0.054	0.038
p-values	0.000	0.000	0.000
q-values	0.001	0.001	0.001
Bandwidth	±45	±14.103	±17.614

Note: Estimates of the effect of being born before December 1955 on the mobility outcomes of individuals. The variable descriptions are provided in Altindag et al. (2022). The first column presents results for individuals born within 45 months of the age threshold, December 1955, and is a reproduction of the results presented in Table 3, Column 3 in Altindag et al. (2022). Column 2 uses rdrobust and a uniform kernel to calculate the optimal bandwidth, Column 3 uses rdrobust and a triangular kernel to calculate the optimal bandwidth. The specification includes month fixed effects, province fixed effects, surveyor fixed effects, as well as indicator variables for education levels, ethnicity, and gender. Corresponding p-values and Anderson's (2008) sharpened q-values are presented, as in Altindag et al. (2022).

Table 4 – Changing clustering (Mental health outcomes)

	Original Study (1)	Without clustering (2)
Mental distress index	0.205	0.205
Standard errors	0.094	0.119
p-value	0.032	0.086
q-value	0.095	0.239
Bandwidth	±45	±45
Somatic symptoms of distress index	0.175	0.175
Standard errors	0.085	0.120
p-value	0.043	0.145
q-value	0.095	0.239
Bandwidth	±45	±45
Non-somatic symptoms of distress index	0.158	0.158
Standard errors	0.092	0.114
p-values	0.088	0.168
q-values	0.095	0.239
Bandwidth	±45	±45
Sum of "yes" answers in SRQ-20	0.734	0.734
Standard errors	0.440	0.563
p-values	0.098	0.193
q-values	0.095	0.239
Bandwidth	±45	±45

Note: Regression discontinuity estimates of the effect of the curfew on mental health outcomes. The variable descriptions are provided in Altindag et al. (2022). The first column presents results for individuals born within 45 months of the age threshold, December 1955, and is a reproduction of the results presented in Table 4, Column 3 in Altindag et al. (2022). The second column is the same estimation without clustering the standard errors on month-year cohort level. The specification includes month fixed effects, province fixed effects, surveyor fixed effects, as well as indicator variables for education levels, ethnicity, and gender. Corresponding p-values and Anderson's (2008) sharpened q-values are presented, as in Altindag et al. (2022).

Table 5 – Removing covariates (Mental health outcomes)

	Original Study (1)	Without covariates (2)
Mental distress index	0.205	0.169
Standard errors	0.094	0.135
p-value	0.032	0.215
q-value	0.095	0.472
Bandwidth	±45	±45
Somatic symptoms of distress index	0.175	0.221
Standard errors	0.085	0.113
p-value	0.043	0.054
q-value	0.095	0.278
Bandwidth	±45	±45
Non-somatic symptoms of distress index	0.158	0.100
Standard errors	0.092	0.125
p-values	0.088	0.427
q-values	0.095	0.472
Bandwidth	±45	±45
Sum of "yes" answers in SRQ-20	0.734	0.560
Standard errors	0.440	0.589
p-values	0.098	0.345
q-values	0.095	0.472
Bandwidth	±45	±45

Note: Estimates of the effect of being born before December 1955 on mental health outcomes. The variable descriptions are provided in Altindag et al. (2022). The first column presents results for individuals born within 45 months of the age threshold, December 1955, and is a reproduction of the results presented in Table 4, Column 3 in Altindag et al. (2022). The specification includes month fixed effects, province fixed effects, surveyor fixed effects, as well as indicator variables for education levels, ethnicity, and gender. The second column is the same estimation without including the above control variables. Corresponding p-values and Anderson's (2008) sharpened q-values are presented, as in Altindag et al. (2022)

Table 6 – Changing bandwidth selection/estimation (mental health outcomes)

	Original Study (1)	RD Robust Uniform (2)	RD Robust Triangular (3)
Mental distress index	0.205	0.146	0.053
Standard errors	0.094	0.142	0.074
p-value	0.032	0.304	0.471
q-value	0.095	0.436	0.458
Bandwidth	±45	±15.246	±11.536
Somatic symptoms of distress index	0.175	0.375	0.199
Standard errors	0.085	0.144	0.097
p-value	0.043	0.009	0.040
q-value	0.095	0.038	0.193
Bandwidth	±45	±16.009	±17.687
Non-somatic symptoms of distress index	0.158	0.065	0.118
Standard errors	0.092	0.143	0.074
p-values	0.088	0.648	0.111
q-values	0.095	0.681	0.201
Bandwidth	±45	±14.848	±11.480
Sum of "yes" answers in SRQ-20	0.734	0.699	-0.039
Standard errors	0.440	0.636	0.352
p-values	0.098	0.272	0.911
q-values	0.095	0.436	0.836
Bandwidth	±45	±16.151	±11.150

Note: Estimates of the effect of being born before December 1955 on mental health outcomes. The variable descriptions are provided in Altindag et al. (2022). The first column presents results for individuals born within 45 months of the age threshold, December 1955, and is a reproduction of the results presented in Table 4, Column 3 in Altindag et al. (2022). Column 2 uses rdrobust and a uniform kernel to calculate the optimal bandwidth, Column 3 uses rdrobust and a triangular kernel to calculate the optimal bandwidth. The specification includes month fixed effects, province fixed effects, surveyor fixed effects, as well as indicator variables for education levels, ethnicity, and gender. Corresponding p-values and Anderson's (2008) sharpened q-values are presented, as in Altindag et al. (2022).