

Gallegos, Sebastian

Working Paper

Replication: "Physical Disability and Labor Market Discrimination: Evidence from a Video Résumé Field Experiment"

I4R Discussion Paper Series, No. 256

Provided in Cooperation with:

The Institute for Replication (I4R)

Suggested Citation: Gallegos, Sebastian (2025) : Replication: "Physical Disability and Labor Market Discrimination: Evidence from a Video Résumé Field Experiment", I4R Discussion Paper Series, No. 256, Institute for Replication (I4R), s.l.

This Version is available at:

<https://hdl.handle.net/10419/324167>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



No. 256

I4R DISCUSSION PAPER SERIES

**Replication:
“Physical Disability and Labor Market
Discrimination: Evidence from a Video
Résumé Field Experiment”**

Sebastian Gallegos

August 2025

I4R DISCUSSION PAPER SERIES

I4R DP No. 256

Replication: “Physical Disability and Labor Market Discrimination: Evidence from a Video Résumé Field Experiment”

Sebastian Gallegos^{1,2,3}

¹Universidad Adolfo Ibáñez, Santiago/Chile

²IZA – Institute of Labor Economics, Bonn/Germany

³HCEO/ The University of Chicago/USA

AUGUST 2025

Any opinions in this paper are those of the author(s) and not those of the Institute for Replication (I4R). Research published in this series may include views on policy, but I4R takes no institutional policy positions.

I4R Discussion Papers are research papers of the Institute for Replication which are widely circulated to promote replications and meta-scientific work in the social sciences. Provided in cooperation with EconStor, a service of the [ZBW – Leibniz Information Centre for Economics](#), and [RWI – Leibniz Institute for Economic Research](#), I4R Discussion Papers are among others listed in RePEc (see IDEAS, EconPapers). Complete list of all I4R DPs - downloadable for free at the I4R website.

I4R Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

Editors

Abel Brodeur
University of Ottawa

Anna Dreber
Stockholm School of Economics

Jörg Ankel-Peters
RWI – Leibniz Institute for Economic Research

Replication: “Physical Disability and Labor Market Discrimination: Evidence from a Video Résumé Field Experiment”*

Sebastian Gallegos

July 9, 2025

Abstract

[Bellemare et al. \(2023b\)](#) examine discrimination against individuals with physical disabilities in the labor market in Quebec, Canada. Their findings indicate that callbacks from potential employers decrease by 25 percentage points if physical disability is (randomly) revealed. Callbacks increase by 10 percentage points if there is a video resume (randomly) sent to potential employers. In this document, we first conduct a computational reproduction using the replication package. Then, we test the robustness of the findings to the inclusion of different covariates, selecting them with a Double Selection Lasso approach. We complement the analysis estimating heterogeneous treatment effects using Causal Forests, which allow us to uncover data-driven subgroups and test their responses to the treatment. We find that [Bellemare et al. \(2023b\)](#)’s estimates are stable across these robustness checks.

*Gallegos (corresponding author): Assistant Professor of Economics, Business School, Universidad Adolfo Ibañez, Chile, IZA Research Fellow and member of the Human Capital and Economic Opportunity Global Working Group (HCEO) at The University of Chicago. Email: sebastian.gallegos@uai.cl. Constanza Gonzalez provided superb research assistance.

1 Introduction

[Bellemare et al. \(2023b\)](#) examines discrimination against individuals with physical disabilities in the labor market through an experiment conducted in the province of Quebec, Canada. The purpose of the study is to measure the gap in callbacks between disabled and non-disabled applicants and to investigate whether this gap is reduced by including a video resume which demonstrates the applicant's unrestricted mobility for the job.

The authors conduct two field experiments sequentially. In the first experiment they generated a set of otherwise identical fake resumes sent to real positions of secretary-receptionist, computer programmer, and accounting clerk. A randomly chosen half of the applicants revealed that the job seekers required the use of a wheelchair on the job, as mentioned in the cover letter.

The second experiment –only for the secretary-receptionist job– included a video sent with the application, showing how the applicant could perform common tasks required for the job while on a wheelchair. This approach allows the authors to study whether the gap observed in the first experiment is reduced when it is demonstrated that there are no differences in performing the job's required activities.

The main dataset was collected between 2016 and 2017 for the first experiment, and between 2018 and 2019 for the second one. The authors find that “*revealing a disability decreases callback rates by 25 percentage points for applications both with and without videos, even after controlling for accessibility constraints. We also find that the video résumés increase callback rates by 10 percentage points for disabled and nondisabled applicants alike.*” ([Bellemare et al. 2023b](#))

We conducted a reproducibility exercise using the replication package available online in the *American Economic Journal: Applied Economics* website, which contains the data and code used by the authors to implement their empirical analysis. ([Bellemare et al. 2023a](#))

The results confirm the robustness of the original findings. We first executed the Stata codes provided by the authors, successfully reproducing all results in the the original published document as well as those in the online appendix. We detected a minor error in an outcome computation which does not affect the authors' results. We then examined the robustness of the results by applying Double Selection

Lasso, which uses the data to automatically select the covariates to be included in the estimation. Finally, we supplemented our analysis by implementing a Causal Forest to estimate Conditional Average Treatment Effects (CATEs) and test for heterogeneous treatment effects across data-driven subgroups.

2 Computational Reproducibility

We used the available replication package [online](#). The replication package includes the code used for cleaning and preparing the raw data, as well as the code used to generate the tables and figures. We report on its contents in [Table 1](#). The shared raw data includes only the variables used in the analyses, rather than the full range of data collected. This is why we label the provision of raw data and cleaning code as partial in column 2. However, while the limited dataset restricts what could be a broader set of robustness checks, the main results are indeed reproducible.

Table 1: Replication Package Contents and Reproducibility

Replication Package Item	Fully	Partial	No
Raw data provided		✓	
Analysis data provided	✓		
Cleaning code provided		✓	
Analysis code provided	✓		
Reproducible from raw data	✓		
Reproducible from analysis data	✓		

Notes: [Table 1](#) summarizes the replication package contents from [Bellemare et al. \(2023b\)](#).

We successfully reproduced all main results. We reproduced [Table 2](#) below, which contains the main estimates presented in Table 2 in the original paper. The only difference is that we report p-values alongside the estimates.¹

¹See appendix for additional replicated tables that were reported in the main analysis of the original paper.

Table 2: Reproduction of Main Results, Linear Probability Models, Dependent Variable: Callback for an Interview or for Requesting Additional Information

	Video experiment			All data
	(1)	(2)	(3)	(4)
Disab	-0.250 (0.038) [0.000]	-0.256 (0.048) [0.000]	-0.250 (0.038) [0.000]	-0.250 (0.039) [0.000]
Video	0.104 (0.036) [0.004]	0.104 (0.038) [0.006]	0.104 (0.036) [0.004]	0.110 (0.034) [0.001]
Disab \times Video	-0.030 (0.046) [0.517]	-0.026 (0.048) [0.585]	-0.037 (0.049) [0.449]	-0.037 (0.040) [0.355]
Access		0.052 (0.034) [0.121]		0.036 (0.031) [0.251]
Disab \times Access		-0.001 (0.044) [0.981]		0.008 (0.040) [0.843]
Video \times Disab \times Late reveal			0.014 (0.033) [0.673]	
Video exp. dummy				0.052 (0.028) [0.063]
Constant	0.450 (0.031) [0.000]	0.420 (0.038) [0.000]	0.450 (0.031) [0.000]	0.371 (0.036) [0.000]
Observations	2021	1788	2021	2236

Notes: [Table 2](#) reproduces Table 2 from [Bellemare et al. \(2023b\)](#). Robust standard errors are in parentheses and p-values are reported in brackets.

Despite the successful reproduction, we identified a minor inconsistency when replicating the construction of the main outcome variable, callbacks. We show how the original outcome variable differs from our constructed outcome variable in [Table 3](#). According to the authors, the main outcome takes the value one if the applicant received either an invitation for an interview or a request for additional information, indicating a successful application. Upon inspecting the data, we found one observation classified as a successful application despite not meeting either criterion, and two observations marked as unsuccessful even though a request for more information was recorded. This discrepancy does not affect the results, as all

three observations belong to the benchmark experiment and are not job applications for the secretary-receptionist position. In other words, they are not part of the sample analyzed in the video experiment nor in the benchmark-video experiment comparison. Thus, there were no differences between the reproduced tables and the tables shown in the original article.

Table 3: Comparison Between Original and Constructed Outcome Variable

Firm Response (Original Outcome)	Firm Response (Constructed)		Total
	0 (No Callback)	1 (Callback)	
0 (No Callback)	2,410	2	2,412
1 (Callback)	1	1,085	1,086
Total	2,411	1,087	3,498

Notes: **Table 3** shows the frequency distribution of the main outcome in the original paper (rows) and our constructed version of the main outcome (columns), based on the specification provided in the paper. The constructed outcome takes a value one if the applicant received an information request or an invitation for an interview—either by phone or in person—using the variables available in the provided dataset.

3 Robustness Reproduction

3.1 Double Selection Lasso

We implement a Double Selection Lasso analysis (Belloni et al. 2013) as a robustness check. The original study does not rely on a pre-analysis plan to define control variables. As a result, there is flexibility in covariate selection which may affect the robustness of the estimated effects. In particular, there is a risk of including only balanced covariates while omitting important unbalanced ones. Omitting covariates could also produce a larger standard error for the coefficient of interest. Overall, the omissions can affect both the magnitude and precision of the estimated causal effects. Double Selection Lasso helps address this issue by automatically selecting covariates that significantly predict the outcome variable, while controlling for potential confounders and avoiding overfitting.

We use the same pool of covariates presented in Table C4 of the original work. **Table 4** replicates the four specifications from the original study to estimate the callback, and **Table 5** compares the estimations obtained using the original specification

(with and without controls) alongside the one proposed in this section, focusing on the effects of disability disclosure in the benchmark and video experiments.

Table 4: Double-Selection LASSO Estimation on Callbacks

	Video experiment			All data
	(1)	(2)	(3)	(4)
Disab=1	-0.269 (0.040) [0.000]	-0.271 (0.050) [0.000]	-0.269 (0.040) [0.000]	-0.264 (0.044) [0.000]
Video=1	0.105 (0.039) [0.007]	0.111 (0.041) [0.007]	0.105 (0.039) [0.007]	0.120 (0.038) [0.002]
Disab=1 × Video=1	-0.013 (0.048) [0.781]	-0.016 (0.051) [0.754]	-0.031 (0.051) [0.549]	-0.034 (0.045) [0.453]
Access=1		0.071 (0.036) [0.045]		0.045 (0.033) [0.181]
Disab=1 × Access=1		-0.002 (0.046) [0.958]		0.019 (0.043) [0.659]
Video=1 × Disab=1 × Late reveal=1			0.034 (0.035) [0.337]	
Observations	1832	1629	1832	1949

Notes: This table reports coefficient estimates from double selection LASSO regressions. The outcome variable is positive callbacks response. “Disab.” is a dummy variable indicating that the application disclosed a disability, either through the cover letter or in the video. “Video” is a dummy variable indicating that the cover letter included a hyperlink to a video résumé. “Access” is a dummy variable indicating that the firm is considered accessible for wheelchairs. “Late reveal” is a dummy variable indicating that the wheelchair (or regular chair) was disclosed later. This variable is set to zero for applications without a video. “Video exp. dummy” is a dummy variable indicating that the application was sent during our second experiment. Robust standard errors are reported in parentheses. Columns (1) - (3) present estimations on only the video experiment data, while column (4) presents the estimation that pools data from the benchmark experiment’s secretaries-receptionists observations with data from the video experiment. Robust standard errors are in parentheses and p-values are reported in brackets.

Table 5: Estimated Effects of Disability Disclosure Across Model Specifications

Estimation	Coefficient		
	Disab.	Video	Disab. \times Video
Baseline	-0.250 (0.039) [0.000]	0.110 (0.034) [0.001]	-0.037 (0.040) [0.355]
All controls	-0.265 (0.045) [0.000]	0.119 (0.038) [0.002]	-0.038 (0.045) [0.402]
Double selection	-0.264 (0.044) [0.000]	0.120 (0.038) [0.002]	-0.034 (0.045) [0.453]

Notes: This table summarizes the main coefficients of interest under three different model specifications. The first two correspond to the estimation of Model (d) in Table 2 and Table C4 in Bellemare et al. (2023b), that is, the specification that pools data from the benchmark experiment’s secretaries-receptionists observations with data from the video experiment. The third specification applies the Double Selection Lasso method, as implemented in this report, selecting among the variables used in the “All controls” results. The sample size for the “Baseline” specification is $N = 2,236$, whereas both the “All Controls” and “Double Selection” estimations are based on $N = 1,949$. Robust standard errors are reported in parentheses and p-values are reported in brackets.

We find that the results are consistent across specifications. The effect of disclosing a disability remains statistically significant and, in magnitude, similar to the one obtained using all controls. The same holds for the inclusion of the video variable.

3.2 Heterogeneous Treatments Effects

While we acknowledge that treatment effect heterogeneity does not mean the average effects are not robust, exploring Conditional Average Treatment Effects (CATEs) helps us see if the results are driven by certain subgroups or if important differences across individuals are being masked. To investigate this, we employ a Causal Forest approach using the Generalized Random Forest (GRF) framework, which enables data-driven estimation of treatment effect variation across covariates.² First, we use a training sample to fit the causal forest and then predict CATES on a separate test sample for each treatment (see Figure A1). Second, we select the most common root split across the trees in the forest and determine the median value of that splitting variable to define subgroups (see Figure A2). Third, we follow the

²Since the original paper does not specify a pre-analysis plan, we conduct additional analyses to examine potential heterogeneous treatment effects.

subgroup analysis approach proposed by [Davis and Heller \(2017\)](#). We compare the estimated treatment effects and standard errors for individuals above and below the mean of the most frequently selected root variable in the forest. In our case, this variable is the business registration year of the firm to which the applicant applied, as well as the timing of the application, depending on the treatment condition. Additionally, we contrast treatment effects between individuals above and below the median estimated CATE.

Tables 6, 7, and 8 present treatment effect heterogeneity in positive callback responses across different treatment conditions. Table 6 reports the treatment effects of the disability condition across subgroups defined by the firm's business registration year (Panel A) and by whether the estimated CATE is above or below the median (Panel B). In both panels, the subgroup coefficients remain negative and statistically significant; however, we find no significant differences in treatment effects across either subgroup. Table 7 uses disability with video as the treatment, and similar conclusions are observed in Panel A. By contrast, when defining subgroups based on whether the predicted CATE is above or below the median (Panel B), a statistically significant difference is observed only in the training sample. This result is not consistent in the hold-out test sample.

Table 6: Treatment Effects of Disability by Groups

	(1) In Sample	(2) Out Sample
Panel A		
Year \geq 2009	-0.337 (0.056)	-0.327 (0.056)
Year $<$ 2009	-0.254 (0.040)	-0.286 (0.040)
Observations	814	814
H0: subgroups equal, p =	0.230	0.551
Panel B		
$\hat{\tau} \geq p50$	-0.314 (0.046)	-0.258 (0.047)
$\hat{\tau} < p50$	-0.247 (0.047)	-0.333 (0.046)
Observations	814	814
H0: subgroups equal, p =	0.309	0.251

Notes: This table presents intent-to-treat (ITT) estimates of the effect of the disability treatment on positive callbacks response, conditional on data-driven subgroups selected using the causal forest. In Panel A, subgroups are defined by the most common root split in the forest—the year of business registration of the employer. In Panel B, subgroups are defined by whether the individual’s predicted CATE lies above or below the sample median. Column (1) reports estimates using the training sample (used to fit the causal forest), while Column (2) shows estimates for the hold-out test sample not used in model training.

Table 7: Treatment Effects of Disability and Video by Groups

	(1) In Sample	(2) Out Sample
Panel A		
Year \geq 2007	-0.209 (0.053)	-0.212 (0.055)
Year $<$ 2007	-0.156 (0.044)	-0.185 (0.043)
Observations	814	814
H0: subgroups equal, p =	0.440	0.690
Panel B		
$\hat{\tau} \geq p50$	-0.276 (0.048)	-0.211 (0.049)
$\hat{\tau} < p50$	-0.075 (0.048)	-0.187 (0.047)
Observations	814	814
H0: subgroups equal, p =	0.003	0.736

Notes: This table presents intent-to-treat (ITT) estimates of the effect of the disability and video treatment on positive callbacks response, conditional on data-driven subgroups selected using the causal forest. In Panel A, subgroups are defined by the most common root split in the forest—the year of business registration of the employer. In Panel B, subgroups are defined by whether the individual’s predicted CATE lies above or below the sample median. Column (1) reports estimates using the training sample (used to fit the causal forest), while Column (2) shows estimates for the hold-out test sample not used in model training.

Table 8 presents heterogeneous treatment effects using the video-only condition. Here, subgroups are defined based on whether the résumé was submitted in March 2019. As in previous cases, we find no statistically significant differences across subgroups in Panel A. However, in Panel B, we observe a statistically significant difference between subgroups defined by the CATE—once more, only in the training sample—suggesting a lack of robustness.

Table 8: Treatment Effects of Video by Groups

	(1) In Sample	(2) Out Sample
Panel A		
Sent in March 2019	0.092 (0.105)	0.165 (0.113)
Not sent in March 2019	0.097 (0.041)	0.117 (0.039)
Observations	814	814
H0: subgroups equal, $p =$	0.966	0.686
Panel B		
$\hat{\tau} \geq p50$	-0.071 (0.060)	0.131 (0.056)
$\hat{\tau} < p50$	0.212 (0.049)	0.086 (0.050)
Observations	814	814
H0: subgroups equal, $p =$	0.000	0.552

Notes: This table presents intent-to-treat (ITT) estimates of the effect of the video treatment on positive callbacks response, conditional on data-driven subgroups selected using the causal forest. In Panel A, subgroups are defined by the most common root split in the forest-time of application. In Panel B, subgroups are defined by whether the individual’s predicted CATE lies above or below the sample median. Column (1) reports estimates using the training sample (used to fit the causal forest), while Column (2) shows estimates for the hold-out test sample not used in model training.

4 Conclusion

In this report, we successfully reproduce the findings of [Bellemare et al. \(2023b\)](#). We extend the analysis by examining the robustness of the results using Double Selection Lasso and by exploring heterogeneous treatment effects through data-driven methods. The results from the Double Selection Lasso procedure are consistent with those reported in the original study. The estimated treatment effects remain similar regardless of which control variables are selected, which supports the robustness of the original findings. In addition, when we use Causal Forests to estimate hetero-

geneous treatment effects (CATEs) and test for differences across subgroups, we do not find statistically significant variation across subgroups selected in a data-driven method. This suggests that the average treatment effects are stable across different types of individuals, further reinforcing the robustness of the results.

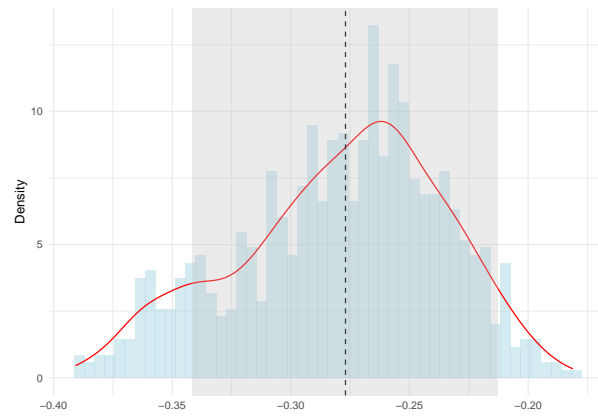
References

- Bellemare, C., Goussé, M., Lacroix, G. and Marchand, S.: 2023a, Data and code for: Physical disability and labor market discrimination: Evidence from a video resume field experiment, Ann Arbor, MI: Inter-university Consortium for Political and Social Research (ICPSR). Version: 2023-09-12.
URL: <https://doi.org/10.3886/E175061V1>
- Bellemare, C., Goussé, M., Lacroix, G. and Marchand, S.: 2023b, Physical disability and labor market discrimination: Evidence from a video résumé field experiment, *American Economic Journal: Applied Economics* **15**(4), 452–76.
URL: <https://www.aeaweb.org/articles?id=10.1257/app.20210633>
- Belloni, A., Chernozhukov, V. and Hansen, C.: 2013, Inference on treatment effects after selection among high-dimensional controls[†], *The Review of Economic Studies* **81**(2), 608–650.
URL: <https://doi.org/10.1093/restud/rdt044>
- Davis, J. M. and Heller, S. B.: 2017, Using causal forests to predict treatment heterogeneity: An application to summer jobs, *American Economic Review* **107**(5), 546–50.
URL: <https://www.aeaweb.org/articles?id=10.1257/aer.p20171000>

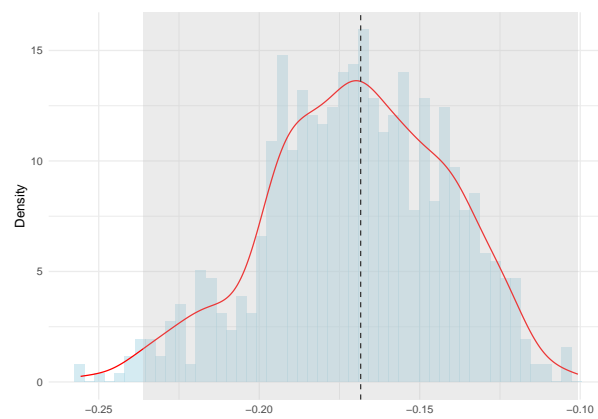
5 Appendix

5.1 Appendix Figures

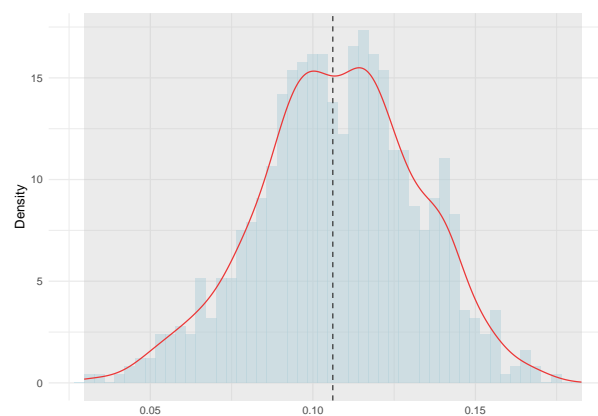
Figure A1: Estimated distribution of the CATE by Causal Forest



(a) Disabled



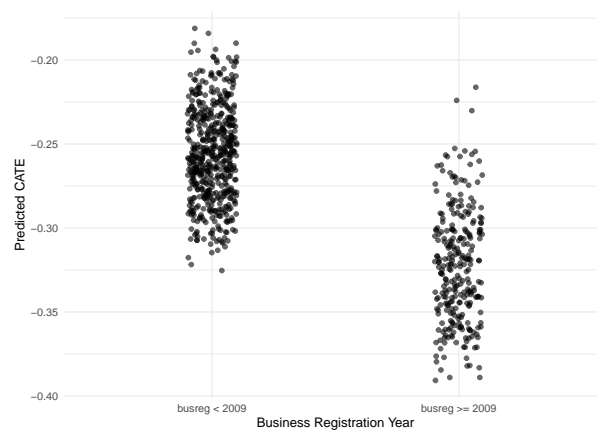
(b) Disabled \times Video



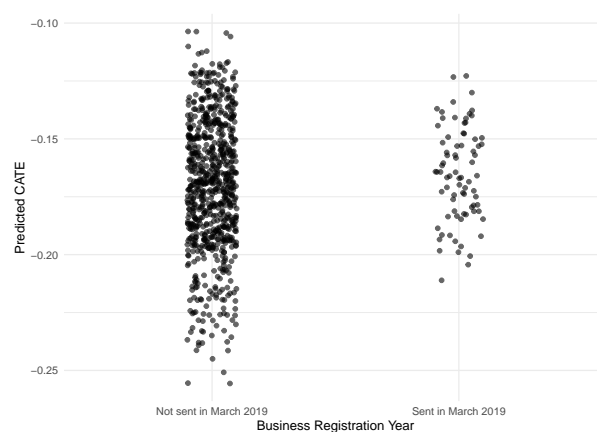
(c) Video

Notes: **Figure A1** presents histograms and density plots illustrating the distribution of Conditional Average Treatment Effects (CATEs) for each treatment condition: disability only, disability with video, and video only. The dashed vertical line represents the estimated Average Treatment Effect (ATE), and the shaded grey area corresponds to its 95% confidence interval.

Figure A2: Distribution of predicted CATEs for the subgroup



(a) Disabled

(b) Disabled \times Video

(c) Video

Notes: **Figure A2** presents the distribution of Conditional Average Treatment Effects (CATEs) across data-driven subgroups for each treatment condition: disability only, disability with video, and video only. Subgroups were defined based on the most common root split variable selected in the causal forest estimation.

5.2 Appendix Tables

Table A1: Reproduction of Viewing Activity and Video Types, OLS Regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Attention								
	Early quitting								
	Focus on revelation zone								
Disab=1 × Late reveal=0	-24.790 (10.688) [0.021]	-0.190 (0.128) [0.139]	-0.122 (0.046) [0.007]	-0.326 (0.099) [0.001]	-0.129 (0.039) [0.001]	0.093 (0.033) [0.005]	0.136 (0.039) [0.001]	0.060 (0.046) [0.189]	0.027 (0.046) [0.554]
Disab=1 × Late reveal=1	-3.680 (10.716) [0.731]	-0.089 (0.126) [0.482]	-0.010 (0.046) [0.823]	-0.177 (0.100) [0.076]	-0.065 (0.043) [0.129]	0.002 (0.033) [0.947]	0.018 (0.038) [0.637]	0.052 (0.037) [0.158]	0.126 (0.067) [0.062]
Late reveal=1	-0.075 (10.437) [0.994]	0.073 (0.123) [0.556]	-0.039 (0.046) [0.394]	0.099 (0.101) [0.325]	0.028 (0.043) [0.519]	0.035 (0.031) [0.261]	0.031 (0.037) [0.396]	-0.089 (0.038) [0.020]	0.034 (0.048) [0.480]
Constant	173.652 (7.239) [0.000]	2.086 (0.084) [0.000]	0.598 (0.031) [0.000]	1.205 (0.068) [0.000]	0.316 (0.030) [0.000]	0.111 (0.020) [0.000]	0.180 (0.025) [0.000]	0.221 (0.030) [0.000]	0.102 (0.027) [0.000]
Observations	946	946	946	946	946	946	946	946	946

Notes: This table reproduces Table 4 from the work of [Bellemare et al. \(2023b\)](#). Robust standard errors are in parentheses and p-values are reported in brackets.