

Legewie, Nicolas; Shiffer-Sebba, Doron; Jacobsen, Jannes; Goldstein, Yoav;  
Dollmann, Jörg

**Working Paper**

## A comment on "Seeing racial avoidance on New York City streets"

I4R Discussion Paper Series, No. 250

**Provided in Cooperation with:**

The Institute for Replication (I4R)

*Suggested Citation:* Legewie, Nicolas; Shiffer-Sebba, Doron; Jacobsen, Jannes; Goldstein, Yoav; Dollmann, Jörg (2025) : A comment on "Seeing racial avoidance on New York City streets", I4R Discussion Paper Series, No. 250, Institute for Replication (I4R), s.l.

This Version is available at:

<https://hdl.handle.net/10419/323244>

**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. 250  
I4R DISCUSSION PAPER SERIES

# **A Comment on “Seeing racial avoidance on New York City streets”**

Nicolas Legewie  
Jannes Jacobsen  
Jörg Dollmann

Doron Shiffer-Sebba  
Yoav Goldstein

**July 2025**

## I4R DISCUSSION PAPER SERIES

I4R DP No. 250

### **A comment on “Seeing racial avoidance on New York City streets”**

**Nicolas Legewie<sup>1</sup>, Doron Shiffer-Sebba<sup>2</sup>, Jannes Jacobsen<sup>3</sup>,  
Yoav Goldstein, Jörg Dollmann<sup>3,4,5</sup>**

*<sup>1</sup>University of Münster/Germany*

*<sup>2</sup>Northwestern University, Evanston/USA*

*<sup>3</sup>German Center for Integration and Migration Research (DeZIM), Berlin/Germany*

*<sup>4</sup>Mannheim Centre for European Social Research (MZES), Mannheim/Germany*

*<sup>5</sup>University of Mannheim/Germany*

**JULY 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*

E-Mail: [joerg.peters@rwi-essen.de](mailto:joerg.peters@rwi-essen.de)  
RWI – Leibniz Institute for Economic Research

Hohenzollernstraße 1-3  
45128 Essen/Germany

[www.i4replication.org](http://www.i4replication.org)

ISSN: 2752-1931

## A comment on “Seeing racial avoidance on New York City streets”<sup>1</sup>

Nicolas Legewie (University of Münster, [nlegewie@uni-muenster.de](mailto:nlegewie@uni-muenster.de)) (corresponding author)

Doron Shiffer-Sebba (Northwestern University, [doron@northwestern.edu](mailto:doron@northwestern.edu))

Jannes Jacobsen (DeZIM Institute, [jacobsen@dezim-institut.de](mailto:jacobsen@dezim-institut.de))

Yoav Goldstein (Independent Researcher, [yoavgoldstein1@gmail.com](mailto:yoavgoldstein1@gmail.com))

Jörg Dollmann (DeZIM Institute and Mannheim Centre for European Social Research (MZES), University of Mannheim, [dollmann@dezim-institut.de](mailto:dollmann@dezim-institut.de))

### Abstract

Dietrich and Sands (2023) used New York City traffic camera footage to experimentally examine the effect of a pair of racialized confederate bystanders on the distance pedestrians maintained from those bystanders as they passed them on the sidewalk. Across their block-randomized experimental conditions, the authors found that pedestrians deviated by around 4 inches on average, maintaining larger distances from Black individuals as opposed to White individuals. Their point estimate was significant at the 5% level. In this *conceptual replication* we use data from a new context and new data collection technique, using 3D videos and computer vision models to estimate the effect of minoritized bystanders on pedestrian distance in Berlin, Germany. Despite these differences to the original study, we directionally reproduce Dietrich and Sands’ main claim. Writ large, pedestrians maintain a larger distance from Muslim bystanders of Middle Eastern descent wearing jalabiyas (a religious Muslim garment), as opposed to White bystanders wearing jeans and t-shirts, in the German context. Using bootstrapped sub-samples from our data, our point estimates range from 0.39 inches (1 cm) to 9.84 inches (25 cm), with an average difference in distance of about 5.9 inches (15 cm). This roughly mirrors Dietrich and Sands’ finding of a 4-inch difference. However, our replication finds greater heterogeneity across locations, where different types of areas and different streets show opposite patterns. Our pooled results are only statistically significant at the 5% level when using a wild block bootstrap but are not significant when using clustered standard errors.

<sup>1</sup> All data, code, and outputs are available at <https://github.com/nlegewie/Replication-Dietrich-and-Sands-2023>

## 1. Introduction

Dietrich & Sands (2023),<sup>1</sup> henceforth DS, tested the hypothesis that the phenotypic race of bystanders will affect the distance pedestrians maintain from them on city sidewalks. The authors tested this hypothesis on two streets in New York City in 2017. They positioned pairs of standing confederates on each street, one in Midtown Manhattan and one on Manhattan's Upper East Side. One confederate pair was phenotypically White whereas the other was phenotypically Black. DS then used video footage from traffic cameras to measure the distance maintained between pedestrians who walked down the street and each confederate pair.

DS tested the effect of phenotypic race on pedestrian distance from confederates experimentally using block randomization. In each block, they measured the distance pedestrians maintained from confederates as a proportion of the sidewalk, calculating an Average Treatment Effect (ATE) using bivariate ordinary least squares (OLS), which reflected a difference-in-means across confederates who were either phenotypically Black or phenotypically White. To account for dependence within each 15-minute experimental condition, in their main approach DS used wild block bootstrapping to estimate confidence intervals. In their supplementary material, they also refer to clustered standard errors, with clustering based on 15-minute experimental conditions. As DS state in their abstract, they find that, “on average, pedestrians [...] give a wider berth to Black confederates, as compared with white non-Hispanic confederates.” As they state on page 1276, when pooling across both neighborhoods, pedestrians deviate on average 3.43% more of the sidewalk when passing phenotypically Black confederates as opposed to phenotypically White confederates. This difference, which amounts to around 4 inches, is significant at the 5% level. DS's main findings were statistically significant in Midtown, where the sample size was larger, but not statistically significant at the 5% level in the more residential Upper East Side, where their sample was substantially smaller.

In the present report, prepared as part of a collaboration between the Institute for Replication and Nature Human Behaviour (Brodeur et al., 2024),<sup>2</sup> we *conceptually replicate* DS, testing the effect of the minoritized status of bystanders on the distance maintained by pedestrians on the streets of Berlin, Germany, in 2024. Our criteria for

replication are an effect in the same direction as the original study (with less distance maintained from racialized White bystanders than minoritized bystanders) and a p-value  $\leq 0.05$ . We adopted DS's block randomization and confederate pair design. The original study captured 3,419 pedestrians across the two relevant experimental conditions. The ATE was  $\beta = 0.032$  without controls and  $\beta = 0.034$  with controls when pooling across locations. Based on a difference in average raw distance of 4.2 inches and standard deviation of 30.7 inches (reported by DS through E-Mail communication), with a power of 80%, we need 839 pedestrians per condition to detect a significant difference at  $\alpha = 0.05$  for a two-tailed test (1123 pedestrians per condition with 90% power, respectively). Based on a pilot at our filming locations, a 15-minute block at the lower pedestrian-traffic location captured 79 pedestrians. Therefore, 10 blocks per experimental condition per location should suffice with a total over 2,800. Like DS, we also filmed one commercial area and one residential area, using 15-minute experimental blocks. Due to practical complications, we filmed in two different residential streets (see Appendix I).

Beyond these parallels, our analysis departs from DS in several ways: our approach 1) captures a different population – Berlin pedestrians in 2024 (specifically two areas: Frankfurter Allee close to the local train station “Frankfurter Allee”, and close to local train station Ostkreuz at Gürtelstraße and Neue Bahnhofstraße). Both locations are in East-Berlin (at the border between Friedrichshain and Lichtenberg districts) and characterized by a comparably low share of non-White inhabitants. 2) We examine pedestrian behaviors toward bystanders of phenotypically Middle Eastern descent (in comparison to phenotypically White bystanders), as this group is one of the major minoritized groups in Germany. 3) Accordingly, we include two distinct minoritized experimental conditions – one with confederates who self-identify as Middle Eastern wearing clothing that does not mark them as Muslim, and one where the same confederates wear jalabiyas, a traditional Muslim garment. 4) Our approach uses an improved measurement strategy that relies on 3D video,<sup>3</sup> which enables us to 5) examine raw distances between individuals rather than proportions of sidewalk widths (DS's original intent as stated in their pre-registration<sup>4</sup>). Finally, (6) we filmed on two different residential streets due to practical concerns that arose during filming (see Appendix I).

We collected 3D videos unobtrusively using small (highlighter marker-sized) 3D cameras. We then applied the 3D Social Research pipeline (3DSR),<sup>3</sup> which enables the precise measurement of distance between individuals in video (our pre-registration is available through the Open Science Foundation Registries<sup>5</sup>). Below, we outline the steps we took to process and analyze the 3D video data. DS assisted our replication by providing additional statistics of their analysis.

Our study substantiates DS's main finding: pedestrians maintain wider berths from bystanders racialized as non-White than they do from bystanders racialized as White. However, our results are less robust to estimation approaches of the confidence intervals and, unlike DS, we find opposite patterns across neighborhoods, driven by one of the residential streets (Neue Bahnhofstraße). Moreover, due to the legal restrictions around data collection in our setting, we were not able to replicate DS's robustness analysis that used pedestrian demographic characteristics (which supported their main claim).

## **2. Conceptual Replication with New Data**

Our analysis replicates DS's analysis in Berlin, Germany. We replicate many of the authors' specifications, calculating ATEs reflecting difference-in-means between experimental conditions that vary bystander minoritized status using bivariate OLS regression. Like DS, we use both a wild block bootstrap and clustered standard errors, and specify both unconditional and conditional models. In the conditional model, where we report Controlled Treatment Effects (CTEs), we control for the 45-minute filming block (three conditions x 15-minute block; see Appendix II for discussion of model controls). We also estimate ATEs for a commercial neighborhood and a residential neighborhood separately, in addition to pooling the two samples (see sample comparisons by location and experimental condition in Table 1).

However, our replication differs from DS in several ways. First and foremost, since we focus on Muslims of Middle Eastern descent as one of the major minoritized groups in the German context (analogous to Black individuals in the United States), we decided

to examine an additional experimental condition. DS's study included one condition with confederates racialized as White and another condition with Black confederates (and a third condition with no confederates that was not used for calculating their main results). In our case, because Muslims of Middle Eastern descent may be othered due to either religion or racialized phenotypical elements, we include one condition where Muslim confederates of Middle Eastern descent wear clear Muslim religious markers (jalabiyas, a religious Muslim garment) and another where the same confederates wear clothing similar to the White confederates.

Our approach also uses a more precise measurement strategy that allows measuring absolute distances, as opposed to distance as a proportion of sidewalk width, following DS's original intent as per their pre-registration.<sup>4</sup>

Additionally, due to the stringent data protection laws in the European Union and Germany, we were required to delete the raw videos immediately after 3DSR processing on the day of capture. Furthermore, we were prohibited from collecting demographic information on pedestrians without their consent, as phenotypic characteristics or perceived gender are classified as "special categories of personal data" under the General Data Protection Regulation (GDPR) and are subject to stringent protection. Therefore, we could not rely on image data to identify pedestrian characteristics (see Appendix II for model specification) or to differentiate between pedestrians.

Because our approach was not able to reliably distinguish between individual pedestrians at the time of data collection, we estimate distances using a "corridor method" (visualized in Figure 1) that utilizes all observed pedestrian locations close to confederates. First, for each video, we modeled the linear path that most pedestrians took walking down the street (the red line in Figure 1). Second, we drew a perpendicular line to that path that crosses the mid-point between the two confederates (blue line). We then defined a "corridor" around the perpendicular line using a constant width for all videos, and used all pedestrian observations that fall within that corridor (the light green data points in Figure 1). The corridor method benefits from the fact that confederates stood in fixed locations marked by tape on the sidewalk.



This corridor method captures the same pedestrian multiple times, which leads to an unknown number of duplicate pedestrian observations in the dataset. While it seems plausible to assume that this adds an equal amount of noise within each experimental condition and location, using this raw data would artificially inflate our sample size. To address this issue, we drew 50 random bootstrap sub-samples with  $N = 3000$ , based on our estimations that the locations will result in a sample size around  $N = 3000$  (see above). We then calculate estimates and standard errors using both a wild block bootstrap and clustered standard errors. We conduct the analysis using these techniques within each of the 50 sub-samples. In the following, we will report the ranges and averages for our point estimates from these sub-sample analyses, as well as the share of estimates that are significant at the 5% confidence level. We do not expect the corridor method to interfere with the accuracy of point estimates because it does not affect block randomization.

## 2.1 Results

Overall, our results directionally replicate DS's main claim in this new population (Figure 2). Specifically, the ATEs in our pooled unconditional sample range between 0.39 inches (1 cm) and 9.84 inches (25 cm), with an average difference in standardized distance of about 5.9 inches (15 cm). In other words, averaged within each experimental condition and across our 50 sub-samples, pedestrians maintain a distance that is about 15 cm greater from Muslim bystanders wearing jalabiyas relative to White bystanders (but not from Muslim bystanders without jalabiyas). The corresponding CTEs from the pooled sample show a range from -0.39 inches (-1 cm) to 7.48 inches (19 cm), with an average of 4.3 inches (11 cm). Our point estimates are not directly comparable with DS as the two studies measure distance differently. DS used "standard pedestrian deviation", which measures distance as a proportion of sidewalk width, while our study uses absolute distance. Nevertheless, DS state that their main results roughly translate to 4 inches (p. 1277), which is 1.9 inches less than the average ATE across our sub-samples and well within the range of our sub-samples.

However, unlike DS, our pooled ATE results are only statistically significant at the 5% level (for 44 out of 50 sub-samples, or 88 percent) when calculating standard errors using a wild block bootstrap. When using clustered standard errors, results are not statistically significant for any sub-sample without controlling for the block (our only covariate; see Appendix II), and are statistically significant for 6 out of 50 sub-samples (12 percent) when controlling for block. DS's pooled results were robust at the 5% level with either measure of uncertainty. Table 2 compares DS's pooled results with our pooled results across model specifications and measures of uncertainty.

## 2.2 Different Filming Locations

The larger confidence intervals in our replication stem from greater heterogeneity across locations. We followed DS in observing two neighborhoods, one commercial and one residential. Like DS, we collected a larger sample in the commercial neighborhood, though our sample was more balanced across neighborhoods (see Table 1). We intended to follow DS in filming one street per neighborhood, but due to logistical challenges during filming this proved impossible. We therefore filmed in two residential streets that were in close proximity (Neue Bahnhofstraße and Gürtelstraße) and one commercial street (Frankfurter Allee).

Unlike DS, when we examine each neighborhood type separately, we find opposite patterns (Figure 3). DS's point estimates in the residential neighborhood were smaller and not statistically significant at the 5% level, but still positive. In our replication, point estimates in the residential neighborhood are overwhelmingly *negative* with close to half of the sub-samples negative and statistically significant CTEs. In other words, in the residential neighborhood pedestrians kept a smaller distance to minoritized confederates. However, most sub-samples do not show significant differences in distance in the residential neighborhood. Our estimates in the commercial setting are much closer to DS, and are almost all significant across sub-samples. Even though unconditional models using clustered standard errors are still not statistically significant at the 5% level in the model without controls, all estimates are positive. Unlike the pooled models, residential

models are even statistically significant in CTE models when using clustered standard errors. Appendix I demonstrates that differences across neighborhood types are driven by a particular location, Neue Bahnhofstraße.

Our change of residential location after the first filming day (see Appendix I) hence reveals that distancing behavior in public spaces, as operationalized by the distance pedestrians maintain from bystanders, may be highly sensitive to the choice of street-level location, calling for future study designs in which 1) more locations are filmed, and 2) locations are randomly selected.

### 3. Conclusion

Pedestrians, on average, maintain a greater distance from minoritized bystanders than non-minoritized individuals, both in New York City and Berlin. Yet, our findings suggest that these patterns may differ street-by-street. DS showed similar point estimates across different neighborhood types (though one neighborhood did not produce statistically significant results). Our findings point to the need for future research to test whether differences across residential areas are an artefact of sample selection within Berlin and New York or stem from substantial differences across cities or countries. Our replication study calls for new approaches that include more and randomly selected filming micro-locations to account for between-neighborhood and within-neighborhood heterogeneity.

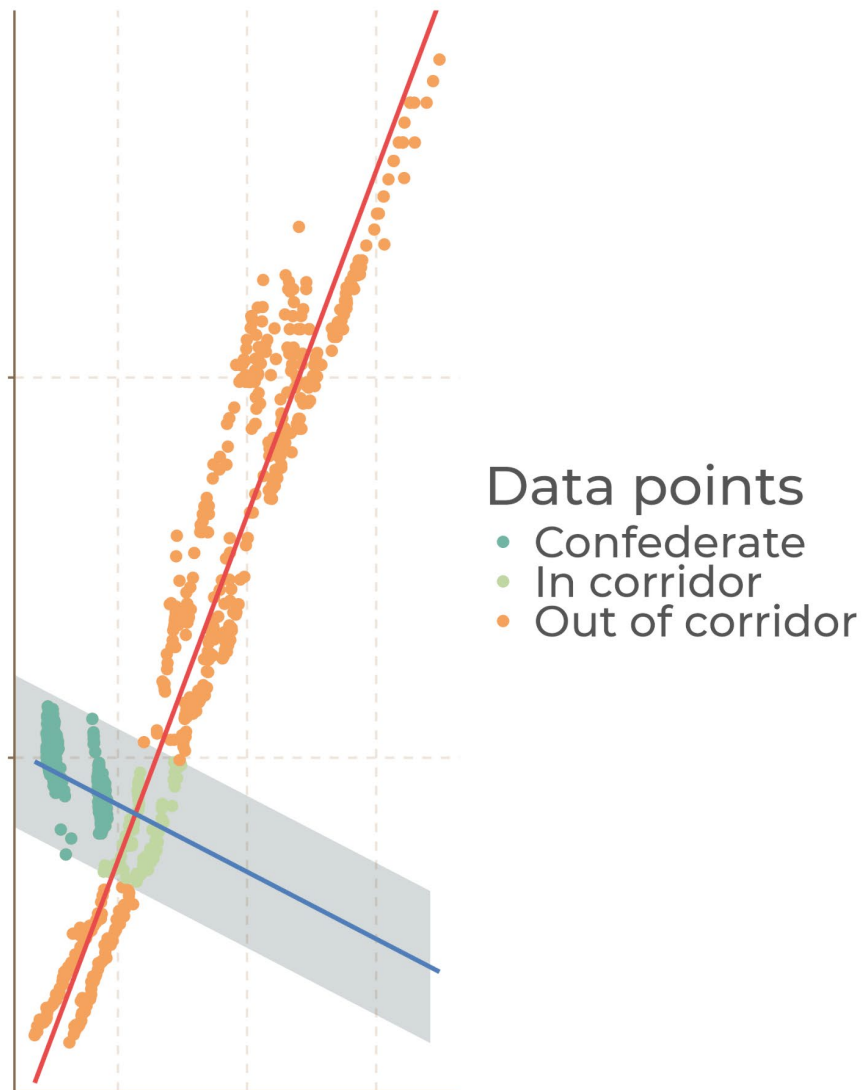
### References

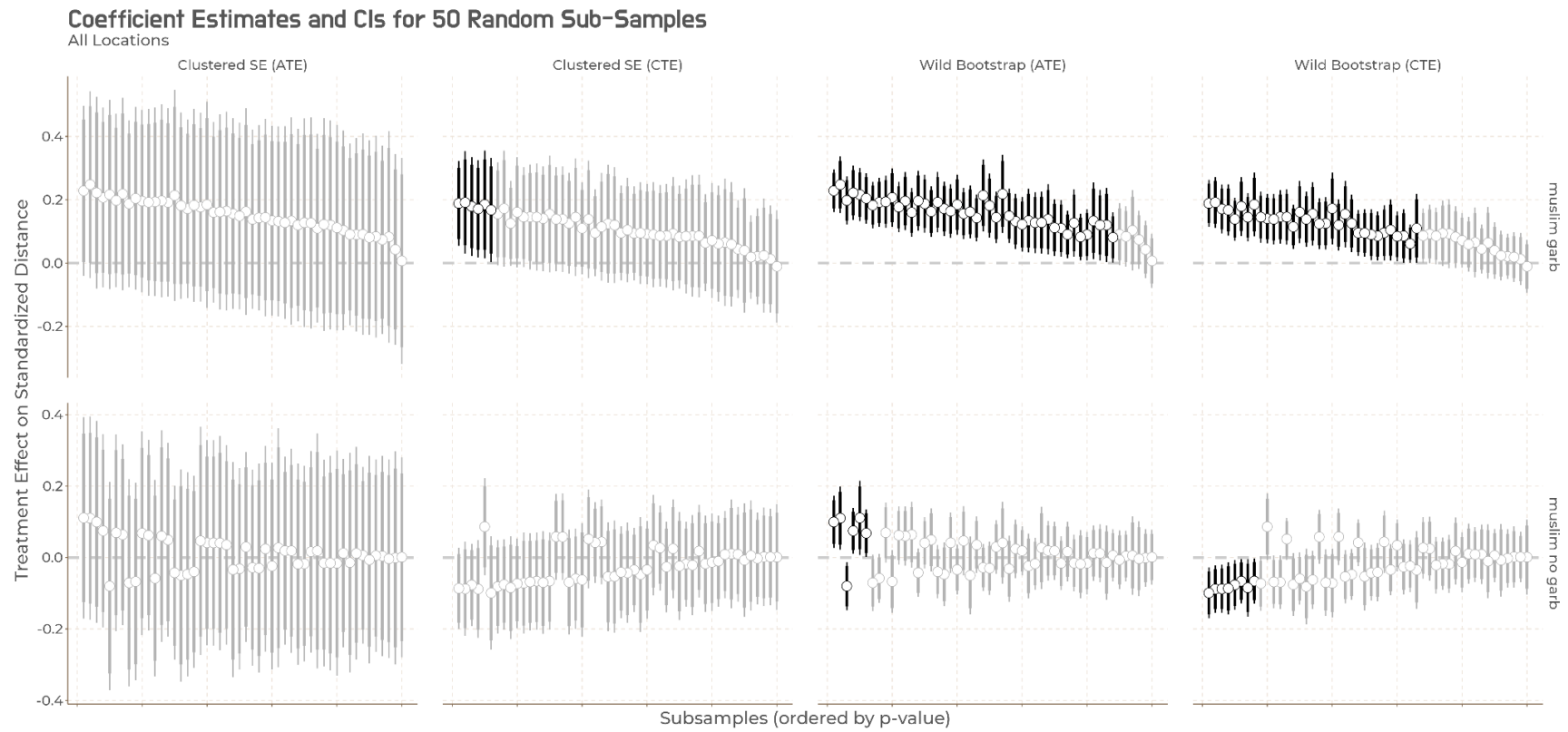
1. Dietrich, B. J. & Sands, M. L. Seeing racial avoidance on New York City streets. *Nat Hum Behav* 7, 1275–1281 (2023).
2. Brodeur, A., Dreber, A., Hoces de la Guardia, F. *et al.* Reproduction and replication at scale. *Nat Hum Behav* 8, 2–3 (2024). <https://doi.org/10.1038/s41562-023-01807-2>

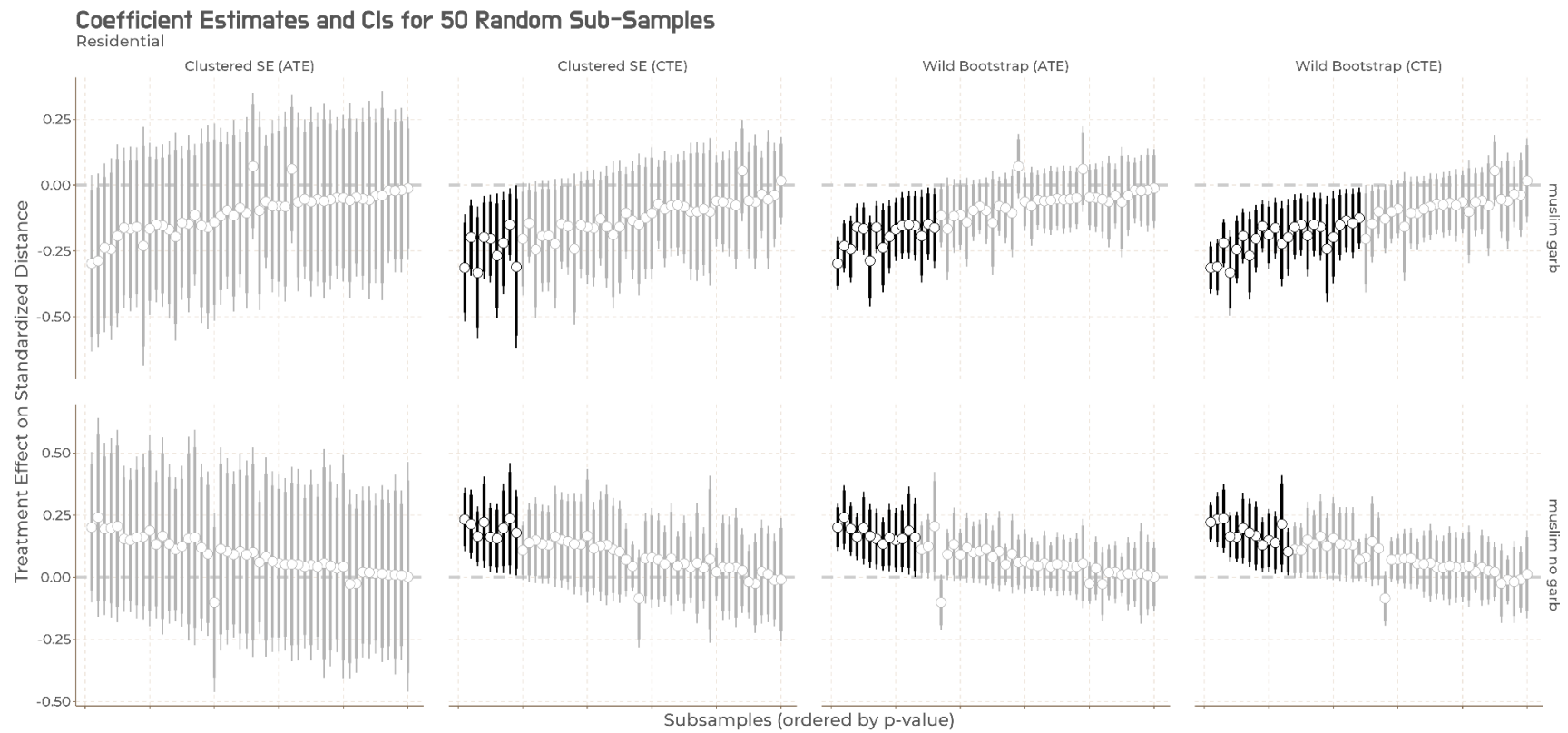
3. Goldstein, Y., Legewie, N. M. & Shiffer-Sebba, D. 3D social research: Analysis of social interaction using computer vision. *Sociological Methods & Research* 52, 1201–1238 (2023).
4. Dietrich, B. J. & Sands, M. L. Using Public Video Feeds to Understand Intergroup Exposure on the Streets of New York. (2017) doi:10.17605/[OSF.IO/VTQEZ](https://doi.org/10.17605/OSF.IO/VTQEZ).
5. Shiffer-Sebba, D., Jacobsen, J., Dollmann, J., Legewie, N., & Goldsein, Y. (2024, June 9). 3DSR-DeZIM: How Pedestrians React to Males with Muslim Markers in Berlin. <https://doi.org/10.17605/OSF.IO/9745E>

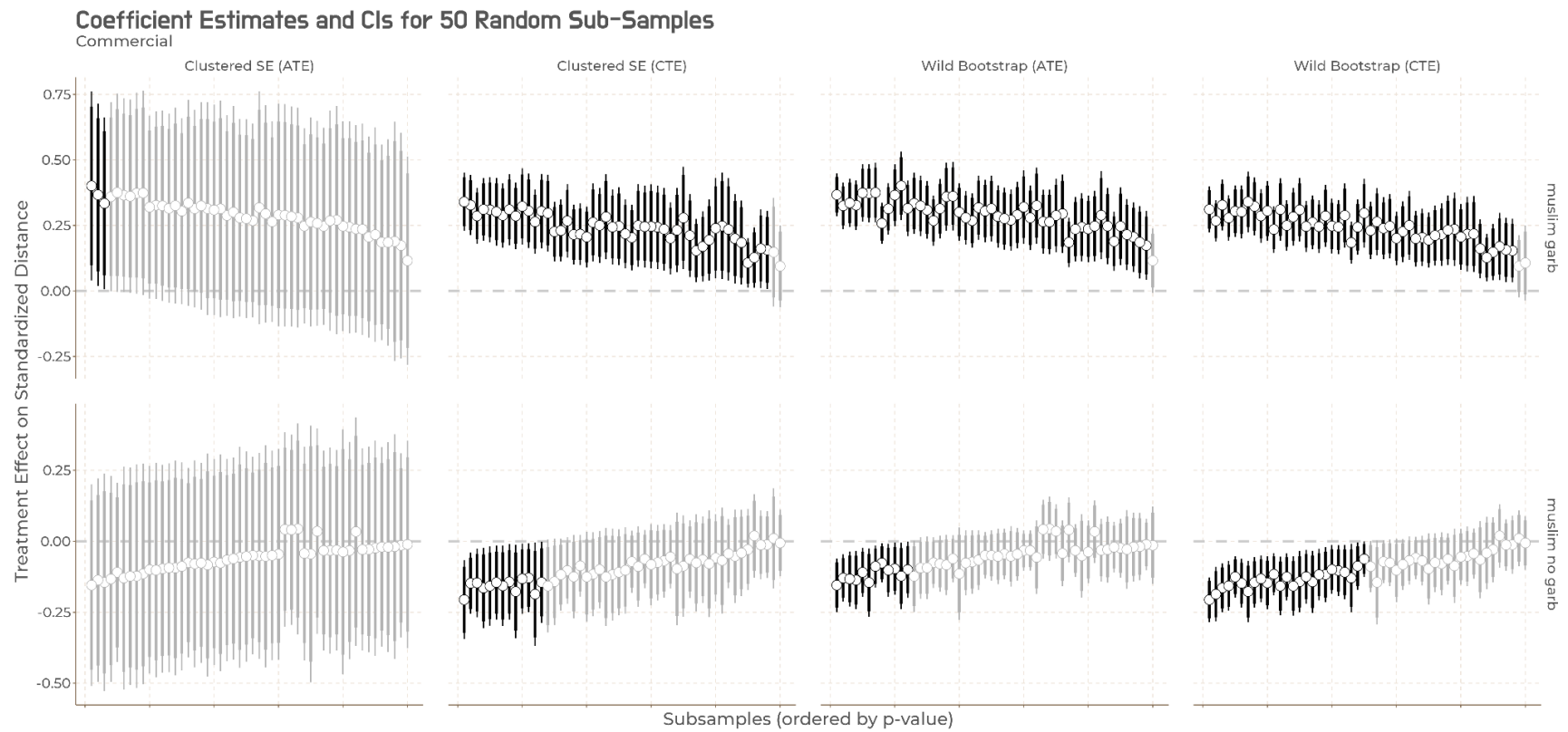
## Figures

Figure 1 – Visualizing the Corridor Method in One Video



**Figure 2 – Experimental Results, Pooled Locations**

**Figure 3A – Experimental Results, Residential Location**

**Figure 3B – Experimental Results, Commercial Location**



## Tables

**Table 1 – Number of Observations in Original and Replication Studies by Experimental Condition and Location**

	<b>Original - Individual Observations (%)</b>	<b>Replication - Data Point Observations (%)</b>
<b>Experimental Condition</b>		
Non-Minoritized	1,719 (50.3%)	33,509 (29.4%)
Black/ Muslim Jalabiyas	1,700 (49.7%)	44,752 (39.3%)
Muslim no Jalabiyas		35,620 (31.3%)
<i>Total</i>	3,419 (100%)	113,881 (100%)
<b>Location</b>		
Commercial	2,903 (84.9%)	64,242 (56.4%)
Residential	516 (15.1%)	49,639 (43.6%)
<i>Total</i>	3,419 (100%)	113,881 (100%)

Note: The original study used pedestrians as the unit of analysis whereas the current replication uses pedestrian-observations. The table excludes the original study's "no confederate" condition as it was not used to compute the main findings. The replication's residential sample includes two locations – Neue Bahnhofstraße ( $N = 32,934$ ) and Gürtelstraße ( $N = 16,705$ ).

**Table 2 – Average Treatment Effects for Original and Replication Studies**

	<b>Original Study, ATE (1)</b>	<b>Replication, ATE (2)</b>	<b>Original Study, CTE (3)</b>	<b>Replication, CTE (4)</b>
Pooled Sample	0.032	Min: 0.007 Mean: 0.15 Max: 0.248	0.034	Min: -0.011 Mean: 0.107 Max: 0.192
Standard Errors	0.011	Min: 0.034 Mean: 0.044 Max: 0.065	0.025	Min: 0.033 Mean: 0.044 Max: 0.058
P value	0.016	88% significant at 5% level	<0.001	64% significant at 5% level
d.f.	3,417	2,997	3,322	2,982
Commercial Neighborhood	0.032	Min: 0.116 Mean: 0.287 Max: 0.401	0.035	Min: 0.095 Mean: 0.238 Max: 0.34
Standard Errors		Min: 0.039 Mean: 0.061 Max: 0.085		Min: 0.04 Mean: 0.057 Max: 0.078
P value	0.009	98% significant at the 5% level	<0.001	96% significant at the 5% level

d.f.	2,901	Min: 1631 Mean: 1689 Max: 1751	2,895	Min: 1624 Mean: 1682 Max: 1744
Residential Neighborhood	0.032	Min: -0.297 Mean: -0.106 Max: 0.071	0.039	Min: -0.332 Mean: -0.138 Max: 0.055
Standard Errors		Min: 0.047 Mean: 0.07 Max: 0.109		Min: 0.045Mean: 0.07 Max: 0.107
P value	0.453	32% significant at 5% level	0.222	50% significant at 5% level
d.f.	514	Min: 1243 Mean: 1305 Max: 1363	424	Min: 1236 Mean: 1298 Max: 1356

Note: Dependent variable in the original study is a relative measure of distance between pedestrians and confederates as a proportion of the sidewalk, derived from video footage from New York City traffic cameras. Dependent variable in the replication is the absolute distance (in inches) between pedestrians and confederates, captured by 3D cameras on Berlin streets. The replication bootstrapped 50 sub-samples of 3,000 observation points (where individuals can contribute multiple observations). Accordingly, point estimates are averages of sub-samples. The original study included a total sample size of 3,419 individuals, each contributing one observation. In both studies, clustered standard errors and standard errors estimated using a wild block bootstrap are clustered by 15-minute experimental conditions. As DS do not report standard errors in their main manuscript, the above standard errors are derived from our calculation based on their reported t-value, confidence intervals and point estimate.

**Table 3 – Average Treatment Effects in Replication Study: Pooled data and areas**

Location	Model	Estimate			Standard Error			Share Significant ( $p \leq 0.05$ )
		Min	Mean	Max	Min	Mean	Max	
All Locations	Clustered SE (ATE)	0.007	0.150	0.248	0.134	0.155	0.171	0.00
	Clustered SE (CTE)	-0.011	0.107	0.192	0.068	0.086	0.102	0.12
	Wild Bootstrap (ATE)	0.007	0.150	0.248	0.034	0.046	0.065	0.88
	Wild Bootstrap (CTE)	-0.011	0.107	0.192	0.033	0.044	0.058	0.64
Residential	Clustered SE (ATE)	-0.297	-0.106	0.071	0.129	0.169	0.231	0.00
	Clustered SE (CTE)	-0.332	-0.138	0.055	0.070	0.110	0.158	0.18
	Wild Bootstrap (ATE)	-0.297	-0.106	0.071	0.047	0.070	0.109	0.32
	Wild Bootstrap (CTE)	-0.332	-0.138	0.055	0.045	0.070	0.107	0.50
Commercial	Clustered SE (ATE)	0.116	0.287	0.401	0.167	0.199	0.233	0.06
	Clustered SE (CTE)	0.095	0.238	0.340	0.048	0.073	0.106	0.96
	Wild Bootstrap (ATE)	0.116	0.287	0.401	0.039	0.061	0.085	0.98
	Wild Bootstrap (CTE)	0.095	0.238	0.340	0.040	0.057	0.078	0.96

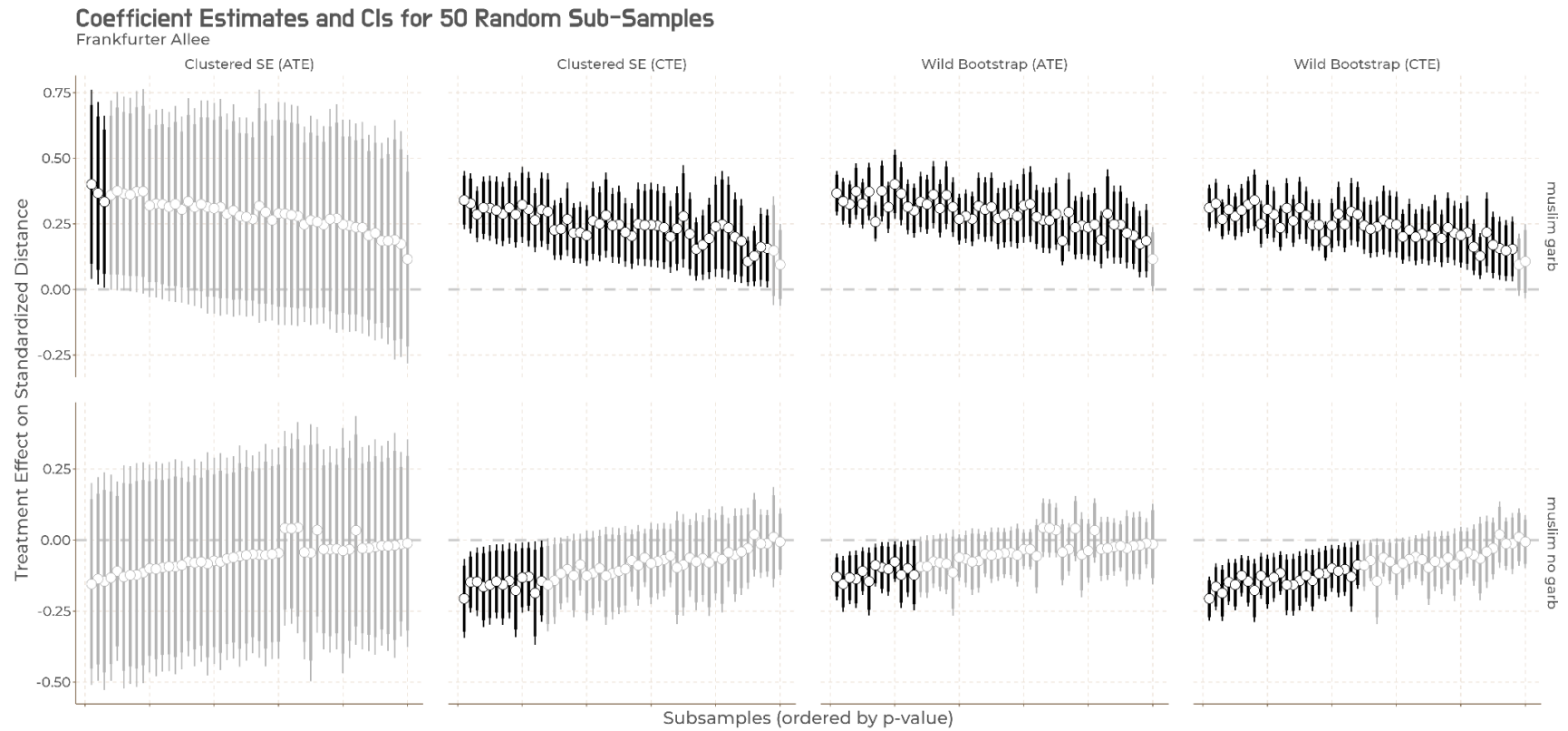
## Appendix I – Experimental Results by Street

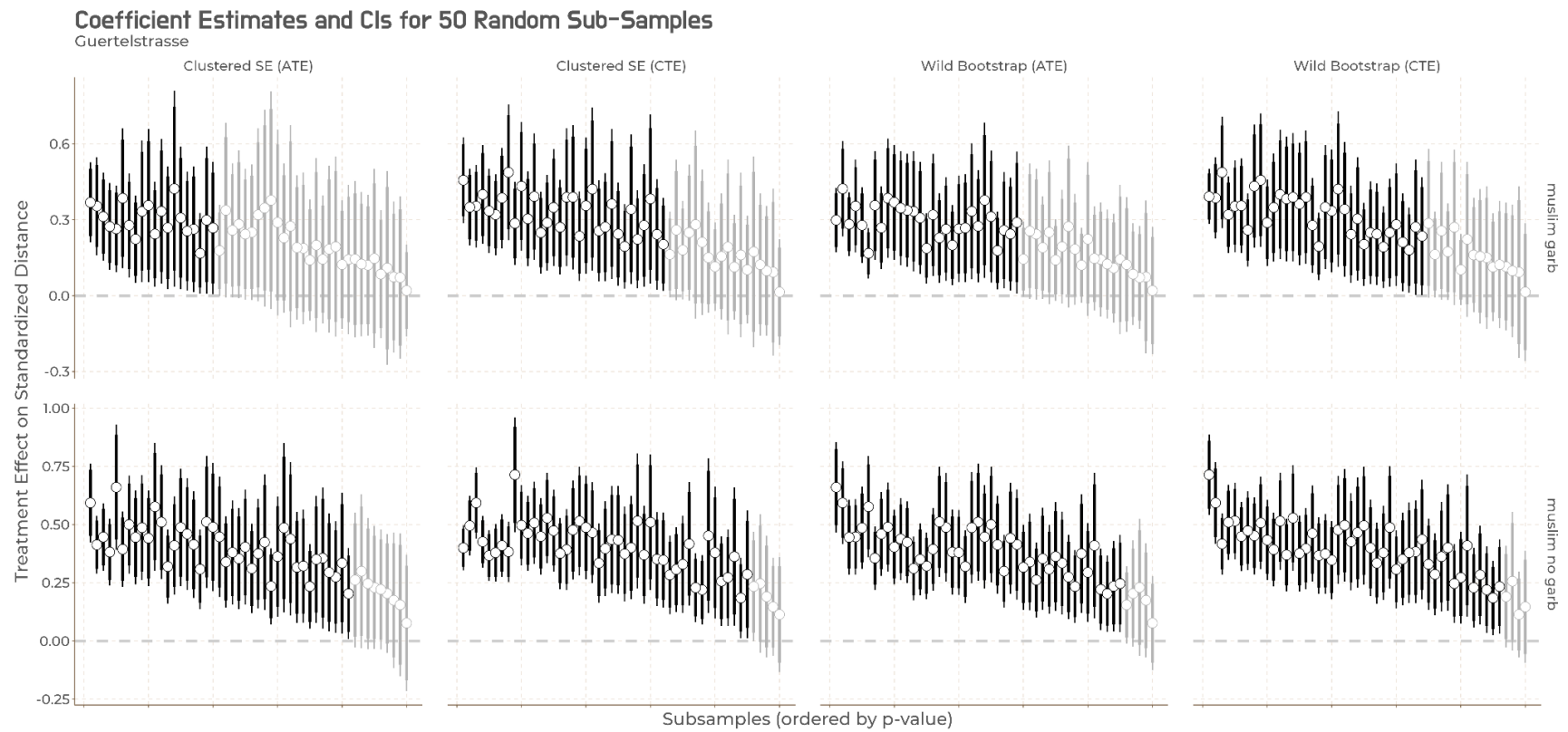
Towards the end of our first day of filming, a resident of the residential neighborhood complained about the filming setup. They insisted that we cease the experiment. Though we had obtained institutional ethical approval, legal advice, and drafted a “Data Protection Impact Assessment (DPIA)” in which we documented all steps that secured we were within legal bounds of the General Data Protection Regulation (GDPR), the negotiation stretched beyond 30 minutes, at which point we determined that we did not have sufficient resources (mostly paid confederate time) for further delay. In addition, it seemed highly likely that the resident would interrupt any effort to continue our data collection. We hence decided to switch to a nearby street section in the same area. Below are our main findings disaggregated by street.

**Table A1 – Experimental Results, Pooled Data and Streets**

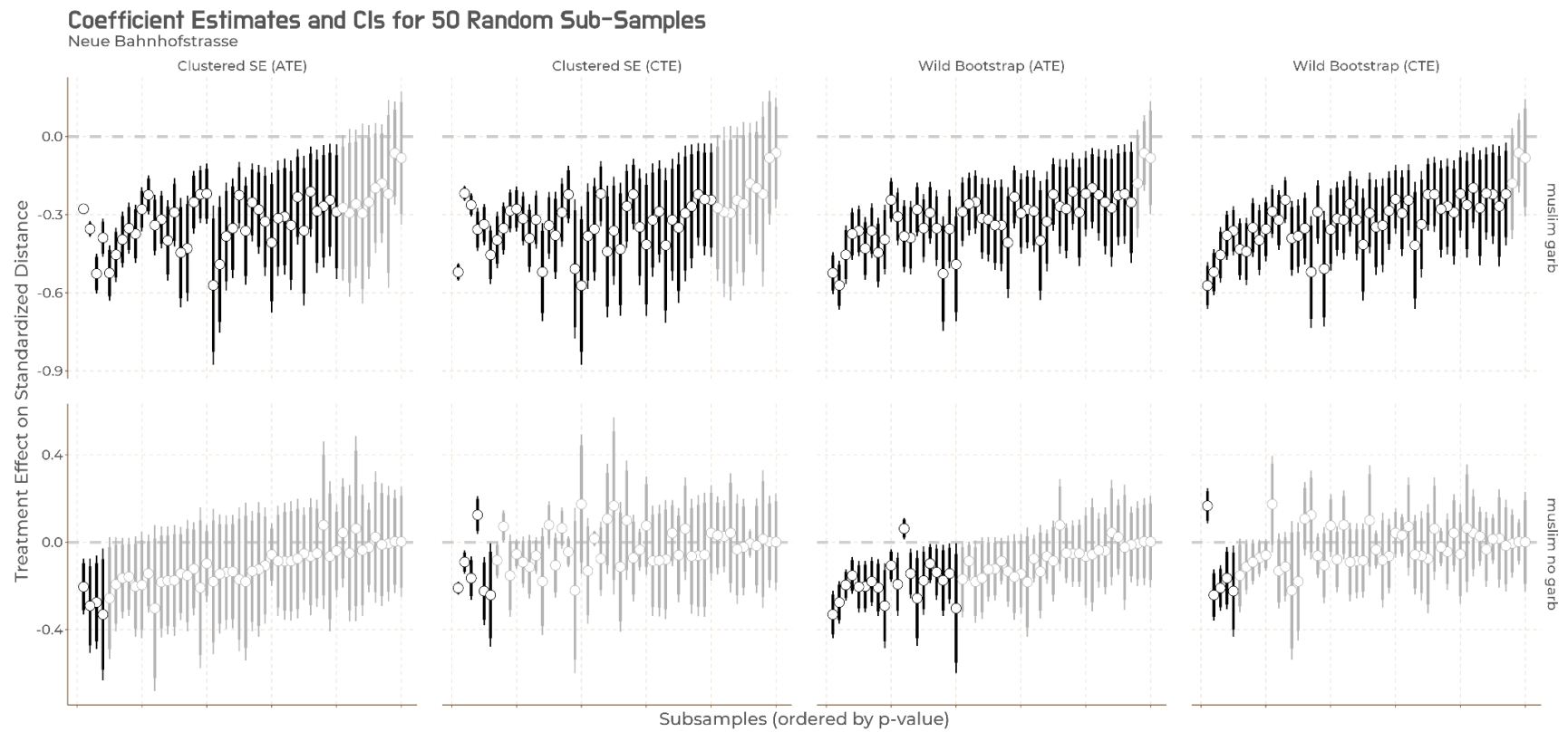
Location	Model	Estimate			Standard Error			Share Significant ( $p \leq 0.05$ )
		Min	Mean	Max	Min	Mean	Max	
All Locations	Clustered SE (ATE)	0.007	0.150	0.248	0.134	0.155	0.171	0.00
	Clustered SE (CTE)	-0.011	0.107	0.192	0.068	0.086	0.102	0.12
	Wild Bootstrap (ATE)	0.007	0.150	0.248	0.034	0.046	0.065	0.88
	Wild Bootstrap (CTE)	-0.011	0.107	0.192	0.033	0.044	0.058	0.64
Neue Bahnhofstraße	Clustered SE (ATE)	-0.571	-0.314	-0.064	0.007	0.095	0.184	0.80
	Clustered SE (CTE)	-0.572	-0.317	-0.063	0.015	0.098	0.181	0.80
	Wild Bootstrap (ATE)	-0.571	-0.314	-0.064	0.041	0.079	0.118	0.94

	Wild Bootstrap (CTE)	-0.572	-0.317	-0.063	0.045	0.080	0.124	0.94
Frankfurter Allee	Clustered SE (ATE)	0.116	0.287	0.401	0.167	0.199	0.233	0.06
	Clustered SE (CTE)	0.095	0.238	0.340	0.048	0.073	0.106	0.96
	Wild Bootstrap (ATE)	0.116	0.287	0.401	0.039	0.061	0.087	0.98
	Wild Bootstrap (CTE)	0.095	0.238	0.340	0.039	0.057	0.078	0.96
Gürtelstraße	Clustered SE (ATE)	0.021	0.236	0.423	0.081	0.141	0.219	0.40
	Clustered SE (CTE)	0.015	0.266	0.488	0.077	0.119	0.192	0.64
	Wild Bootstrap (ATE)	0.021	0.236	0.423	0.051	0.110	0.163	0.58
	Wild Bootstrap (CTE)	0.015	0.266	0.488	0.055	0.114	0.173	0.66

**Figure A.1A – Experimental Results, Frankfurter Allee**

**Figure A.1B – Experimental Results, Gürtelstraße**



**Figure A.1C – Experimental Results, Neue Bahnhofstraße**

**Appendix II – Model Covariates**

Due to data protection restrictions in data collection, we were unable to include the same set of covariates in our models with controls. DS used controls for “pedestrian race and gender, 45 min time block and indicators for pedestrians travelling in groups or pairs on the Upper East Side and for pedestrian race and time block in Midtown.” Our model with controls includes only the 45-min time block as a control.