

## A Commentary on “Gender Bias in Parental Attitude: An Experimental Approach” by Begum, Grossman, and Islam (2018)

Olle Hammar, Carl Bonander, Gunther Bensch, Niklas Jakobsson, and Abel Brodeur

**ABSTRACT** Begum et al. (2018) examined gender bias in parental attitudes using an experimental approach in rural Bangladesh. Households were reported as randomly assigned to treatment conditions in a lab-in-the-field allocation task. We show that the group assignment was inherited from Islam (2019), a previous, nonrandomized experiment conducted in the same region. The lack of randomization contradicts the design descriptions provided by the authors in Begum et al. (2018) and elsewhere and raises concerns about the validity of comparisons across treatment groups. This also points to serious shortcomings in the reporting and transparency of the study design—issues that mirror those that led to the retraction of Islam (2019) from the *European Economic Review*.

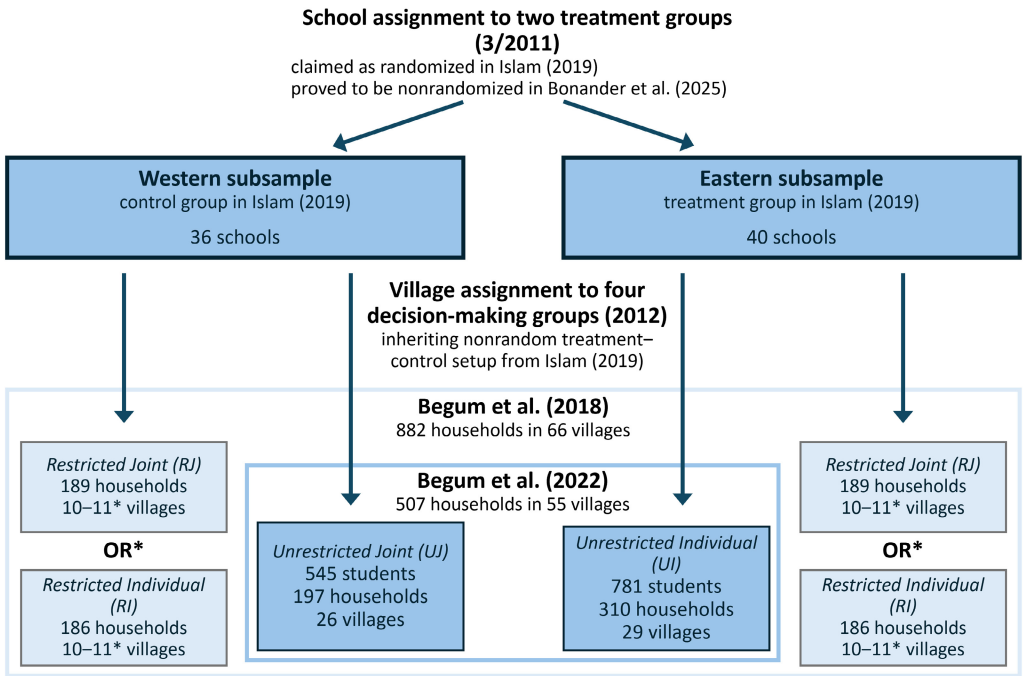
**KEYWORDS** Replication • Reproduction • Parental bias • Gender • Bangladesh

### Introduction

Begum et al. (2018) examined parental attitudes toward male and female children using a high-stakes, lab-in-the-field allocation task with parents of school-age children in rural Bangladesh. Parents were assigned to one of four groups: making either (1) individual or joint decisions about how to allocate (2) restricted or unrestricted endowments, to anonymous girls or boys at a nearby school. The authors concluded that (1) while there is evidence of bias both for and against boys and girls, there is no systematic gender bias by either parent; and (2) decision patterns do not differ significantly between individual and joint decision-making.

We challenge the claim in Begum et al. (2018) that households were assigned randomly to treatment (allocation tasks)—a condition necessary for the validity of their claims about differences across treatment groups and their causal interpretation. Two of the four hypotheses in Begum et al. (2018:1649) relate directly to comparisons of gender bias across allocation tasks and are therefore sensitive to the lack of randomization.<sup>1</sup>

<sup>1</sup> From Begum et al. (2018): “Hypothesis 2 (H2): More bias will be observed in the restricted treatments” and “Hypothesis 3 (H3): Less bias will be observed in our joint treatments than in our individual treatments.”



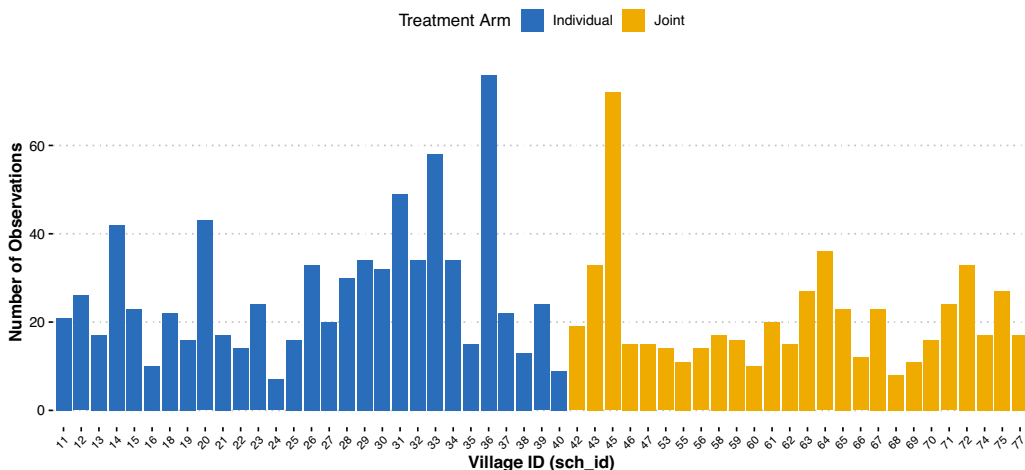
\* Clarification would require replication data.

Data available	No data available
----------------	-------------------

**Fig. 1** Empirical and design-level links between Islam (2019) and Begum et al. (2018). Group assignment in Islam (2019) is inherited by both Begum et al. (2018) and its follow-up paper Begum et al. (2022). While the link for two groups (Unrestricted Individual, UI, and Unrestricted Joint, UJ) can be demonstrated using the replication data available for Begum et al. (2022), the link for Restricted Individual (RI) and Restricted Joint (RJ) can be fully verified only if replication data for Begum et al. (2018) become available (see also the description on replication data in section 1 of the online appendix). Although both papers rely on the same nonrandomized assignment structure, the identification problem is more consequential in Begum et al. (2018), where group comparisons underpin the main causal claims.

Because replication data for Begum et al. (2018) are unavailable, we rely on replication data from a follow-up study (Begum et al. 2022), which uses a subset of the original data (see section 1 of the online appendix for details). These data show that group assignment was inherited from Islam (2019) (see Figure 1). As shown in Bonander et al. (2025), the claimed school-level randomization in Islam (2019) was not randomized, which implies that group assignment in Begum et al. (2018) was also not randomized. In the case of Islam (2019), this finding led to the paper’s retraction from the *European Economic Review* (Islam 2025b).

Bonander et al. (2025) already noted the empirical and design link between Islam (2019) and Begum et al. (2018). Yet, the response by the original author (Islam 2025a) broadly rejected this connection and withheld the claim that treatment was randomly and independently allocated in Begum et al. (2018). Our argument does not rest on any overlap of individual respondents across studies, nor does it depend on whether the original study involved monetary incentives. Rather, the critique concerns the structural inheritance of group assignments and the mischaracterization of the allocation process in the published description.



**Fig. 2** Number of observations in each village, color coded by treatment arm, showing that all observations within a village were assigned to the same group.

## Evidence on Nonrandomized Group Assignment

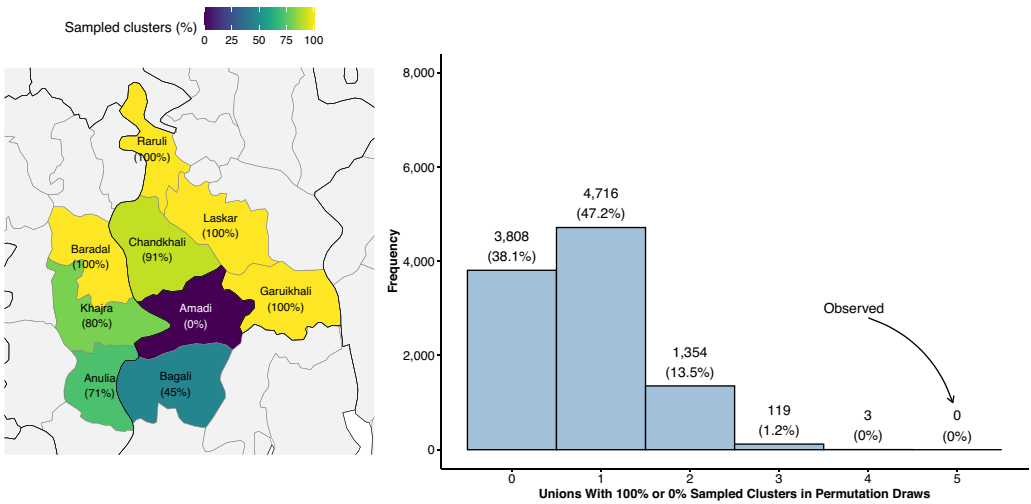
In the replication files from Begum et al. (2022), the variable *Treatment group* (*treat*) indicates assignment to one of two treatment arms: *Individual* ( $treat = 1$ ) or *Joint* ( $treat = 2$ ). Observations represent individuals nested within households. Although Begum et al. (2018, 2022) state that randomization occurred at the household level, the *treat* variable in the replication data is constant within villages (Figure 2), implying village-level rather than household-level allocation, as described in the papers (see section 2 of the online appendix for details). In the following, we demonstrate that allocation is constant not only within villages, but also within larger geographic areas (union councils).<sup>2</sup>

In our reproduction of Islam (2019)—see Bonander et al. (2025)—we found that the same ID variable (*sch\_id*) had been (re)used in Islam (2019) and Begum et al. (2022).<sup>3</sup> The studies took place in the same districts in Bangladesh (Khulna and Satkhira) in the same year (2012). In the response to Bonander et al. (2025), Islam (2025a:17) confirmed that “both studies operated in the same districts, shared village/school IDs were used for administrative tracking of working areas.”

Additionally, Islam (2025a:15) writes, “to minimize logistical challenges, [Dr. Begum] leveraged the broader infrastructure of Dr. Islam’s ongoing work in the

<sup>2</sup> Union councils are administrative areas in Bangladesh. There are approximately 150 unions in the two districts studied by Begum et al. (2018, 2022)—Khulna and Satkhira—of which only eight were included in the study (Begum et al. 2022).

<sup>3</sup> The same ID variable used in both Islam (2019) and Begum et al. (2022) is described as indicating *schools* in the former but *villages* in the latter. The fact that the studies used identical ID variables is further supported by the replication data for another paper coauthored by one of the authors (Islam et al. 2025), whose sample overlaps with Islam (2019), including matching *sch\_id* and school names. The dataset for Islam et al. (2025), however, is no longer available, as the authors have withdrawn that paper from publication.



**Fig. 3** Left plot: Sampled villages in Begum et al. (2022) as a share of schools/villages included in Islam (2019) (with union-level shares in parentheses). Right plot: Results of a permutation test randomly permuting the village sample 10,000 times, counting the number of unions with 0% or 100% sampled clusters in each iteration. Numbers above each bar show the number and percentage of iterations that arrive at a specific value. The observed count in Begum et al. (2022) is 5.

region (Islam 2019), which had established local partnerships and familiarity with the study areas.” Despite this strong connection between papers, neither cited the other.

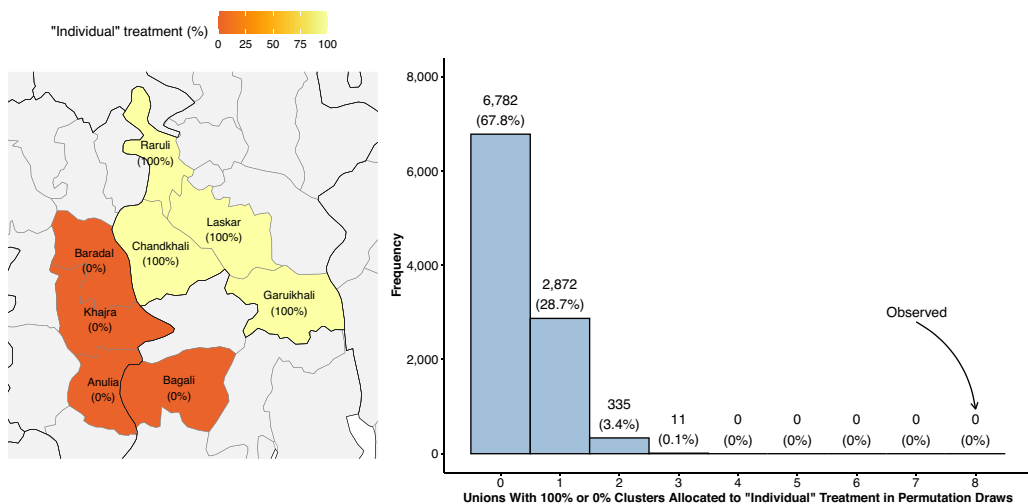
In addition to the shared infrastructure, the replication data from Begum et al. (2022)<sup>4</sup> reveal a perfect overlap between the treatment group assignments in Islam (2019) and Begum et al. (2022) (see Bonander et al. 2025). Specifically, the data show that the same schools/villages that were treated in Islam (2019) were assigned to the *Individual* group in Begum et al. (2018, 2022), while the control schools/villages in Islam (2019) were assigned to the *Joint* group in Begum et al. (2018, 2022).<sup>5</sup>

As shown in Bonander et al. (2025), the treatment allocation in Islam (2019) was not randomized. Contrary to its reported design, neither the selection of schools nor the assignment to treatment was random, meaning the study was not a randomized controlled trial, a fact also confirmed by Islam (2025a). Since Begum et al. (2018, 2022) used the same allocation as Islam (2019), their treatment groups were also not randomly assigned, despite claims to the contrary (see the following commentary section and section 2 in the online appendix for direct quotes suggesting random assignment).

In Figure 3 we repeat the spatial analysis from Bonander et al. (2025) by linking geographic data from Islam (2019) to the replication files from Begum et al. (2022)

<sup>4</sup> We have requested but not obtained access to a replication package for Begum et al. (2018).

<sup>5</sup> While the school/village sample in Begum et al. (2018, 2022) is smaller than that in Islam (2019), with 76 schools/villages included in Islam (2019), 66 in Begum et al. (2018), and 55 in Begum et al. (2022), the Begum et al. (2022) school/village sample constitutes a proper subset of the school sample in Islam (2019), including the exact same school/village IDs and treatment status. See Table A1 in the online appendix for a copy of Table A1 from Bonander et al. (2025).



**Fig. 4** Left plot: Share of individuals allocated to the *Individual* treatment arm in Begum et al. (2022) by union (with union-level shares in parentheses). Right plot: Results of a permutation test randomly permuting the village-level treatment variable 10,000 times, counting the number of unions with 0% or 100% villages assigned to the *Individual* treatment in each iteration. Numbers above each bar show the number and percentage of iterations that arrive at a specific value. The observed count in Begum et al. (2022) is 8. The highest count we obtain in the random permutations is 3.

via the shared `sch_id` variable. In the left panel of Figure 3, we show the union-level share of schools/villages sampled in Islam (2019) ( $n=76$ ) that were also sampled in Begum et al. (2022) ( $n=55$ ). The map shows that the villages sampled by Begum et al. (2022) were not spatially random: northeastern unions sampled in Islam (2019) are fully represented, while southern unions are underrepresented or absent. In the right panel of Figure 3, we show with the use of permutation tests that this sampling is very unlikely to have been randomized.

Figure 4 further shows the union-level share of individuals assigned to the *Individual* treatment arm in Begum et al. (2022). The map in the left panel of Figure 4 echoes the findings of a distinct east–west spatial divide in Bonander et al. (2025). As demonstrated with the use of permutation tests in the right panel of Figure 4, this spatial treatment allocation pattern is inconsistent with both village-level randomization as well as the household-level randomization described in Begum et al. (2018, 2022). This means that the claims about random village selection and random task allocation made in Begum et al. (2018, 2022) (see section 2 of the online appendix) are not true.

## Inconsistent Framing of an Experimental Approach

The assignment to elicitation groups is described as randomized in both Begum et al. (2018) and Begum et al. (2022). This interpretation is reinforced by the experimental framing and language used throughout both papers.

However, the author response to Bonander et al. (2025), as well as an unpublished authors' response on Begum et al. (2022), at times suggests the opposite—that the

study neither involved nor required experimental group assignment. In light of this substantive difference in characterization, we reexamine how the design is presented across the papers and their publication history.

### **Description of Group Assignment in Begum et al. (2018, 2022)**

The following passages in the two published papers describe the assignment to elicitation groups:

Subjects across treatments were randomly assigned. (Begum et al. 2018:1658)

These villages were selected randomly from the region. . . . Participating households were randomly assigned to one of the two mutually exclusive treatment groups. (Begum et al. 2022:1048)

The samples for individual or joint treatment groups were determined randomly. Hence, . . . we compare between two treatment groups that were assigned randomly. (Begum et al. 2022:1050)

The published papers thus unambiguously and repeatedly describe the group assignment as randomized.<sup>6</sup>

### **Use of Causal and Experimental Language in Begum et al. (2018)**

Begum et al. (2018) is subtitled “An Experimental Approach,” which explicitly signals to readers that the study involves randomized treatment conditions. The “field experiment” is not a within-subject or hypothetical allocation task. Instead, the paper refers to “treatments” and compares outcomes across groups, without discussing caveats or identification assumptions. Although these terms are not exclusive to randomized designs, the absence of methodological clarification invites a straightforward interpretation: that the group assignment was randomized. This sense is further strengthened by explicit statements such as “subjects across treatments were randomly assigned. Hence, our results suggest that subjects in UI and UJ behave the same way as those in RI and RJ” (Begum et al. 2018:1658).

### **Changes Between Begum et al. (2014) and Begum et al. (2018)**

The 2014 working paper version of Begum et al. (2018) explicitly states:

The households were randomly assigned to one of the four mutually exclusive treatments, corresponding to different experiment conditions. (Begum et al. 2014:13)

The working paper also describes the study as an “artifactual field experiment” (Begum et al. 2014:33), most likely referring to the Harrison and List (2004) typology,

<sup>6</sup> A full set of sampling and assignment descriptions quoted from Begum et al. (2018) and Begum et al. (2022) is provided in section 2 of the online appendix.

which assumes random assignment to experimental conditions. In the final published version, these statements are softened or removed. For example, the sentence introducing the four mutually exclusive treatments reads as follows in the published paper:

Subjects participated in one of four treatments: . . . (Begum et al. 2018:1647)

The stronger framing of random assignment to groups, however, was maintained in Begum et al. (2022).

## Inconsistencies in Author Responses

The authors’ unpublished response on Begum et al. (2022) offers a new explanation: that enumerators assigned households to elicitation groups on the basis of observable characteristics and “contextual sensitivity”:

Villages were not randomly assigned to conditions, nor was such randomization necessary. (Begum et al. 2025:3)

We note that in Begum et al. (2018), household assignment to elicitation groups did not involve pre-generated randomization codes. Instead, field supervisors—under the direction of the lead researcher, then-PhD student Lutfunnahar Begum—managed group assignments. Enumerators were instructed to assign households in a way that ensured group comparability based on observable socioeconomic and demographic traits. With direct knowledge of the local environment, they were well-positioned to implement this field-based allocation with contextual sensitivity. (Begum et al. 2025:4)

This explanation is inconsistent with the language used in the published papers. Moreover, in the author response to Bonander et al. (2025), the following description of treatment allocation in Begum et al. (2018) was provided:

As for the preference elicitation, participants *within* [emphasis in original] villages were randomly assigned to one of four groups: . . . These groupings of participants were conducted *within* [emphasis in original] villages at the individual level, rather than at the village level. Therefore, they have *no* [emphasis in original] connection to Islam (2019), where treatment and control groups were assigned at the school level, in contrast to the individual-level assignment in the Begum et al. study. (Islam 2025a:14)

There is *absolutely no* [emphasis in original] link between treatment status in Islam (2019) and Begum et al. (2018) study. The allocation task was independent of the treatment status in Islam (2019). (Islam 2025a:16)

These explanations introduce a new inconsistency regarding treatment assignment and directly contradict Table A1 (in the online appendix) and Figure 4, which show that all observations within a given village/union were assigned the same treatment status. Such perfect within-village uniformity is extremely implausible under randomized individual/household-level assignment and strongly suggests that assignment occurred at the village or union level—mirroring the nonrandom assignment structure used in Islam (2019).

## Concluding Remarks

The group assignment in Begum et al. (2018) reflects a nonrandom structure inherited from Islam (2019), contradicting the paper's presentation of its experimental design. This mischaracterization undermines the validity of causal comparisons and highlights serious shortcomings in the framing and reporting of the study design—issues we believe warrant editorial correction.

More broadly, this case underscores the importance of accurate reporting and the availability of replication data. When design details are unclear or misleading, and data are not made available, it becomes difficult—if not impossible—for readers to assess the credibility of published findings. ■

**Acknowledgments** This commentary was prepared for the Institute for Replication (I4R) (Brodeur et al. 2024). Any errors are the authors' own.

**Reproducibility and Data Availability** Our replication package is posted here: <https://osf.io/wp8fr/>.

## References

- Begum, L., Grossman, P. J., & Islam, A. (2014). *Identifying gender bias in parental attitude: An experimental approach* (Discussion Paper, No. 32/14). Monash University, Department of Economics. Retrieved from [https://www.monash.edu/\\_data/assets/pdf\\_file/0006/925692/identifying\\_gender\\_bias\\_in\\_parental\\_attitude\\_an\\_experimental\\_approach.pdf](https://www.monash.edu/_data/assets/pdf_file/0006/925692/identifying_gender_bias_in_parental_attitude_an_experimental_approach.pdf)
- Begum, L., Grossman, P. J., & Islam, A. (2018). Gender bias in parental attitude: An experimental approach. *Demography*, 55, 1641–1662.
- Begum, L., Grossman, P. J., & Islam, A. (2022). Parental gender bias and investment in children's health and education: Evidence from Bangladesh. *Oxford Economic Papers*, 74, 1045–1062.
- Begum, L., Grossman, P. J., & Islam, A. (2025). *Response to Hammar et al. (2025) replication report on Begum et al. (2022)* (Unpublished response submitted to the *Oxford Economic Papers* editors). Centre for Development Economics and Sustainability, Monash Business School, Melbourne, Australia.
- Bonander, C., Hammar, O., Jakobsson, N., Bensch, G., Holzmeister, F., & Brodeur, A. (2025). “Try to balance the baseline”: A comment on “Parent–teacher meetings and student outcomes: Evidence from a developing country” by Islam (2019). *European Economic Review*, 175, 105021. <https://doi.org/10.1016/j.euroecorev.2025.105021>
- Brodeur, A., Mikola, D., Cook, N., Brailey, T., Briggs, R., de Gendre, A., . . . Zhong, Y. (2024). *Mass reproducibility and replicability: A new hope* (I4R Discussion Paper Series, No. 107). Institute for Replication. Retrieved from <https://www.econstor.eu/bitstream/10419/289437/1/I4R-DP107.pdf>
- Harrison, G. W., & List, J. A. (2004). Field experiments. *Journal of Economic Literature*, 42, 1009–1055.
- Islam, A. (2019). RETRACTED: Parent–teacher meetings and student outcomes: Evidence from a developing country. *European Economic Review*, 111, 273–304.
- Islam, A. (2025a). *Response to the comment on “Parent–teacher meetings and student outcomes: Evidence from a developing country”* (Center for Open Science preprint paper). Retrieved from <https://osf.io/2axzm/files/osfstorage/67d05727c515ebd4fbf4b27d>
- Islam, A. (2025b). Retraction notice to “Parent–teacher meetings and student outcomes: Evidence from a developing country” [European Economic Review, 111, 273–304 (2019)]. *European Economic Review*, 175, 105026. <https://doi.org/10.1016/j.euroecorev.2025.105026>
- Islam, A., Vlassopoulos, M., Zenou, Y., & Zhang, X. (2025). *Centrality-based spillover effects* [Dataset]. <https://doi.org/10.5281/zenodo.14790409> (data are no longer available)

Abel Brodeur (corresponding author)  
[abrodeur@uottawa.ca](mailto:abrodeur@uottawa.ca)

*Hammar* • Department of Economics and Statistics, Linnaeus University, Växjö, Sweden; Institute for Futures Studies, Stockholm, Sweden; <https://orcid.org/0000-0002-2341-3751>

*Bonander* • School of Public Health and Community Medicine, University of Gothenburg, Gothenburg, Sweden; <https://orcid.org/0000-0002-1189-9950>

*Bensch* • Department of Climate and Development Policy, RWI–Leibniz Institute for Economic Research, Essen, Germany; <https://orcid.org/0000-0001-7964-7533>

*Jakobsson* • Karlstad Business School, Karlstad University, Karlstad, Sweden; <https://orcid.org/0000-0002-7143-8793>

*Brodeur* • Department of Economics and Institute for Replication, University of Ottawa, Ottawa, Ontario, Canada; <https://orcid.org/0000-0003-3980-4324>