I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17784

Comment on "Food Insecurity and Mental Health of Women During COVID-19: Evidence from a Developing Country" by Rahman et al.

Lenka Fiala
Anders Kjelsrud
Essi Kujansuu
Abel Brodeur

MARCH 2025

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17784

Comment on "Food Insecurity and Mental Health of Women During COVID-19: Evidence from a Developing Country" by Rahman et al.

Lenka Fiala
University of Bergen and Institute for Replication

Anders Kjelsrud
Oslo Metropolitan University

Essi Kujansuu
University of Innsbruck, University of Turku and Institute for Replication

Abel Brodeur
University of Ottawa, Institute for Replication and IZA

MARCH 2025

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Comment on "Food Insecurity and Mental Health of Women During COVID-19: Evidence from a Developing Country" by Rahman et al.*

Rahman et al. (2021) study the correlation between mental health and food insecurity during the COVID-19 pandemic in Bangladesh. They report that food insecurity increases in the sample and that this is associated with increased stress. This result is not reproducible from the author-provided dataset. In fact, the data suggests that higher food insecurity reduces stress. Additionally, we identify undisclosed overlaps of respondents from datasets of related papers, with inconsistencies in sample selection and data collection descriptions. Taken together, we believe these issues undermine the credibility of the paper.

JEL Classification: B41, C12, I12, I18, J16, O12

Keywords: reproduction, replication, food insecurity, mental health, COVID-19

Corresponding author:
Abel Brodeur
University of Ottawa
75 Laurier Ave E
Ottawa, ON K1N 6N5
Canada

E-mail: abrodeur@uottawa.ca

* We are grateful to Carl Bonander, Niklas Jakobsson and Gunther Bensch for comments and suggestions. This comment was prepared for the Institute for Replication (Brodeur et al. 2024). Errors are ours.

Introduction. Rahman et al. (2021) claim to document a positive association between food insecurity and stress level for a sample of rural Bangladeshi women during the COVID-19 pandemic. This claim is not consistent with the supplementary data for the paper. In fact, this data suggests a *negative* association between food insecurity and stress. Furthermore, the authors disclose that their study is "part of a larger study to understand the health and well-being during COVID-19" (Rahman et al. 2021, p.4). Using the authors' supplementary data file, we were able to match the study participants to several other research papers, two of which have been shown to suffer from several data problems (Kjelsrud et al. 2025, Brodeur et al. 2025).[1] We also point out a critical coding inconsistency in the measure of stress, as well as numerous inconsistencies between the written text, figures, and supplementary dataset.

Coding error in the stress measure. The majority of the items in the stress questionnaire (Perceived Stress Scale; PSS, see Cohen and Williamson 1988) are asked such that higher values indicate more stress (for instance, "How often did you get angry?"). Four of the items, however, are asked in a way meant to be reverse-coded (for instance item 4, "How often did you think you can solve your problem?"). Brodeur et al. (2025) document that Vlassopoulos et al. (2024) incorrectly treat the four reverse-code questions just like the other stress questions in the second endline at 10 months, implying that positive outcomes for these items (less stress) are treated as negative (more stress).

We find that the results presented in Rahman et al. (2021) share the same coding inconsistency as in Vlassopoulos et al. (2024). We discovered this using the Excel sheet provided as supplementary data for the paper. Note that the authors did not provide replication codes, nor did they give descriptions of how they constructed

[1] We match individuals based on the variable RECORD_ID. Between Rahman et al. (2021) and Vlassopoulos et al. (2024), we find a perfect match on several outcome variables, including 10 stress-related questions, 16 COVID-19 knowledge questions, and 8 food insecurity questions. Additionally, individuals align on select demographic variables, such as household size and monthly income (excluding a case where income is reported as zero in the PLOS One dataset). Based on this, we think it is probable that RECORD_ID denotes the same individuals across the different datasets.

their key variables. Based on the Excel sheet, however, we were able to reproduce the mean of the stress variable reported in the paper (and reproduce the paper's Table 3), but only when we implemented the same coding inconsistency as in Vlassopoulos et al. (2024). We are therefore confident the coding inconsistency appears also in this study.

Non-reproducible results. Of the remaining key variables, we were able to reproduce the sample mean as reported in the paper for the COVID-19 knowledge and attitude indices (i.e., reproduce the paper's Tables 1 and 2), but we failed to reproduce the food insecurity measure.

The categorization of food insecurity presented in Figure 1 of Rahman et al. (2021) suggests that the authors used a method similar to Ballard et al. (2013).[2] In Table 1, we summarize the values we get when we apply this method to the dataset available in Rahman et al. (2021)'s supplementary files (column: Data). We also compare this to the numbers mentioned on page 6 of the paper (column: Text). As can be seen, the numbers do not match. We further use computer vision (OpenCV in Python) to estimate the length of the bars in Figure 1 in the paper and present these numbers in the last column (column: Figure).[3] Unexpectedly, overlooking small rounding errors in the estimates from the figure, the figure's wave 1 results correspond to wave 2 from the dataset and wave 2 from the text, however, it is necessary to switch the text's percentages around for "Food secure" and "Mildly food insecure". Similarly, wave 2 results in the figure correspond to wave 1 results in the text, but not those in the dataset.

Importantly, the paper concludes that food insecurity worsens between the two data collection waves. This is not the case with the numbers the authors report in

[2]Table 4 of Ballard et al. (2013) describes how a score can be constructed from the FIES questionnaire. Score 0 corresponds to no food insecurity and answering "no" to all eight questions. Score 1 corresponds to mild food insecurity and answering "yes" to at least one of the questions 1-3 and "no" to any subsequent ones. Score 2 corresponds to moderate food insecurity and answering "yes" to at least one of the questions 4-6 and "no" to all subsequent ones. Score 3 indicates severe food insecurity and requires answering "yes" to either question 7 or 8 or both.

[3]As a robustness exercise, we also manually counted the number of pixels in the figure, which gave almost identical numbers.

the text or supply in the Excel sheet – it is only true in the figure, which seems to flip the data collection waves around.

Table 1: Comparison of Food Insecurity Scores Across Estimate Sources

Wave 1	**Source**		
FIES Score	**Text**	**Data**	**Figure**
Food secure	8% (183)	3% (78)	16.2%
Mildly food insecure	17% (418)	10% (232)	21.1%
Moderately food insecure	67% (1604)	51% (1220)	55.3%
Severely food insecure	8% (197)	36% (871)	7.4%
Wave 2			
FIES Score	**Text**	**Data**	**Figure**
Food secure	21% (507)	16% (380)	8.4%
Mild food insecure	16% (380)	21% (507)	17.2%
Moderate food insecure	56% (1336)	56% (1336)	66.0%
Severe food insecure	7% (179)	7% (179)	8.4%
Improvement	Yes	Yes	No

Note: The percentages report the proportion of women with each FIES score. In parenthesis, when available, the number of individuals with the FIES score. The number estimates from the figure were obtained using OpenCV package in Python.

To sum up so far, we were able to reproduce the sample mean of the stress measure as reported in the paper, and we could produce a plausible food security measure for survey wave 2 (as reported in the text of the paper, correcting two numbers that might have flipped by accident). Given this, we expected that we would be able to reproduce the main result of the paper: the positive association between stress and food insecurity reported in Table 4. We were not successful.

As mentioned above, the authors do not provide replication codes and the description in the text is vague. Given the summary statistics provided in the paper, our first thought was to regress the PSS scores (0-40) on the FIES scores (0-3). The result of this is presented in Column 1 of Table 2. As can be seen, we find a negative correlation, meaning that more food insecurity is associated with *less* stress (p-value < 0.001) – the opposite of what is reported in the paper and what is highlighted in the abstract. Next, we tried to use an alternative food insecurity measure, constructed by simply averaging the eight binary FIES items. The results from this exercise are presented in Column 2. We find a coefficient of comparable magnitude as what is reported in the paper, but again, with the *opposite* sign (p-

value < 0.001). Finally, we computed binary variables denoting each of the four levels of the FIES score and regressed these variables on the PSS scores, using "food secure" as the base category. The results are presented in Column 3. We find that respondents with "mild food insecurity" are more stressed on average than respondents being "food secure", but the respondents being "moderately" or "severe food insecure" are the *least* stressed of them all – again contracting the main claim of the Rahman et al. (2021).

Note that the analysis above includes a coding error in the calculation of the PSS scores (as evident from the reported sample mean) in the attempt to reproduce the authors' results. In Columns 4-6 we therefore reproduce the same set of regressions as above using the corrected PSS scores as an outcome.

We find no correlation between stress and the FIES scores (Columns 4 and 5), but some signs of a positive correlation when using the itemized measure (Column 6). Specifically, respondents with "mild" or "moderate food insecurity" are more stressed on average than those being "food secure". However, there is no difference in stress levels between those being "food secure" and those being "severely food insecure".

We also attempted to reproduce the subsequent regression results reported in the paper, but we did not succeed for any of them.

Other inconsistencies in relation to connected papers. The sample of Bangladeshi women in Rahman et al. (2021) overlaps perfectly with the sample of Vlassopoulos et al. (2024), while around two-thirds of the sample also appear in Siddique et al. (2024).[4] All three papers claim to have *randomly* sampled study participants from an earlier database. The large overlap across papers clearly shows that this randomization procedure was flawed.

Moreover, despite Rahman et al. (2021) and Vlassopoulos et al. (2024) ending up with the same sample of women, the two studies outline entirely different sampling

[4]Both in Rahman et al. (2021) and Vlassopoulos et al. (2024), the sample is 2,402 women. Out of these 2,402 women, 1,632 are also in Siddique et al. (2024), meaning 770 are not, and in total, Siddique et al. (2024) has 7,492 individuals in their sample.

Table 2: Stress levels and food insecurity

	PSS-scores			PSS-scores, corrected		
	(1)	(2)	(3)	(4)	(5)	(6)
FIES score (0-3)	-0.562***			-0.014		
	(0.111)			(0.107)		
Food insecurity (0-1)		-1.728***			-0.115	
		(0.385)			(0.371)	
Mildly insecure			1.012***			1.648***
			(0.303)			(0.292)
Moderately insecure			-0.751***			0.657**
			(0.266)			(0.266)
Severely insecure			-0.873**			0.139
			(0.420)			(0.380)
Observations	2402	2402	2402	2402	2402	2402

Note: The outcome variable in Columns 1-3 is the PSS score used in the original paper, while the outcome in Columns 4-6 is the same variable but corrected for the coding error mentioned in the text. Standard errors in parentheses.
*** p<0.01 ** p<0.05 * p<0.10

procedures (see Brodeur et al. (2025) pages 24-25 (February 24, 2025 version) for more details, including links to other papers not mentioned here).

The stated timing of the second wave of the current study roughly matches the stated timing of the baseline survey of Vlassopoulos et al. (2024) ("3-4 weeks after" May 5-June 2 versus May 31-June 15). The fact that both the stress indicators and COVID-19 knowledge items have identical values in the two datasets per RECORD_ID suggests that the second survey wave and the Vlassopoulos et al. (2024) baseline indeed are the same survey. Yet, we find that the food security measures in the baseline of Vlassopoulos et al. (2024) perfectly match those in the *first* survey wave of the current study, supposedly collected May 5-June 2, and that the food security measures from the *second* wave perfectly match those in Siddique et al. (2024), supposedly collected much later, in June 22 to July 7. The different timelines are summarized in Figure 1.

Rahman et al. (2021) present their data as observational, while in fact around two-thirds of the sample was exposed to the COVID-19 informational treatment of

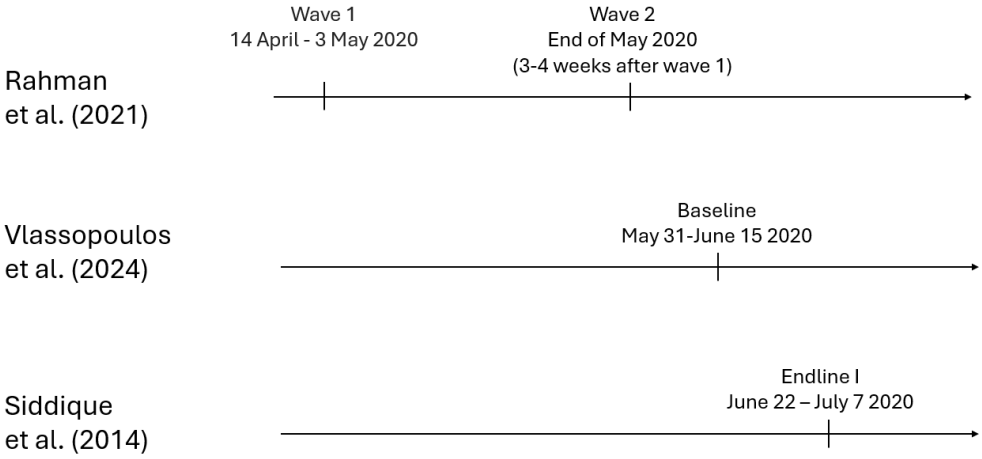

Figure 1: Summary of key data collections in this and related papers.

Siddique et al. (2024) *just* before the second survey wave.[5] This should have been disclosed, especially since Siddique et al. (2024) report massive treatment effects on COVID-19 awareness and knowledge, outcomes closely related to those studied in this paper.

There are also several inconsistencies in the reported demographics across the three datasets. While we get (almost) perfect matches with monthly income and number of family members among the 2,151 individuals for whom demographic data is available in Rahman et al. (2021) — data is missing for 251 individuals, which goes unreported — we do not get matches for the other overlapping demographic variables. There is a known error in how age gets reported in Vlassopoulos et al. (2024) (the problem is explained in detail in Appendix E of Brodeur et al. (2025) and it involves mixing the age of the mother with that of the father). Education matches for 961 (648) individuals but mismatches for 1,190 (1,503) when compared with the Vlassopoulos et al. (2024) dataset (Siddique et al. (2024)) dataset). There are also inconsistencies between the education and literacy variable within the PLOSone sample: many individuals classified as illiterate report having completed eight or more years of education. Finally, primary occupation shows poor alignment across the datasets, primarily because the variable from Vlassopoulos et al. (2024) contains

[5] Siddique et al. (2024) report on a randomized field experiment that compared the effectiveness of text and voice based interventions to raise awareness of and compliance with COVID-19 prevention guidelines. The intervention took place from April 5 to May 20, 2020.

uninterpretable values.

Concluding remarks. In this short note we have documented numerous inconsistencies between the written text, figures, and supplementary dataset; a large number of undisclosed facts of the dataset; a coding error in the main outcome variable; and that the main result of a positive association between food insecurity and stress is not reproducible from the author-provided dataset. In fact, the data suggests a negative association between food insecurity and stress. Taken together, we believe these issues undermine the credibility of the paper.

References

Ballard, T. J., Kepple, A. W. and Cafiero, C.: 2013, The food insecurity experience scale: Development of a global standard for monitoring hunger worldwide. FAO Technical Paper.

Brodeur, A., Fiala, L., Fitzgerald, J., Kujansuu, E., Valenta, D., Rogeberg, O. and Bensch, G.: 2025, A comment on "Improving Women's Mental Health During a Pandemic" by Vlassopoulos et al. (2024). I4R Discussion Paper.

Brodeur, A., Mikola, D., Cook, N. et al.: 2024, Mass reproducibility and replicability: A new hope. I4R Discussion Paper 107 (Preprint).

Cohen, S. and Williamson, G. M.: 1988, *Perceived stress in a probability sample of the United States*, SAGE Publications, p. 3167.

Kjelsrud, A., Kotsadam, A., Rogeberg, O. and Brodeur, A.: 2025, A comment on "Raising health awareness in rural communities: A randomized experiment in Bangladesh and India" by Siddique et al. (2024). I4R Discussion Paper.

Rahman, T., Hasnain, M. G. and Islam, A.: 2021, Food insecurity and mental health of women during COVID-19: Evidence from a developing country, *PLoS One* **16**(7), e0255392.

Siddique, A., Rahman, T., Pakrashi, D., Islam, A. and Ahmed, F.: 2024, Raising health awareness in rural communities: A randomized experiment in Bangladesh and India, *Review of Economics and Statistics* **106**(3), 638–654.

Vlassopoulos, M., Siddique, A., Rahman, T., Pakrashi, D., Islam, A. and Ahmed, F.: 2024, Improving womens mental health during a pandemic, *American Economic Journal: Applied Economics* **16**(2), 422–455.