

Dupraz, Yannick

Working Paper

Recession, Mortality, and Migration Bias: A Comment on Arthi et al. (2022)

I4R Discussion Paper Series, No. 25

Provided in Cooperation with:
The Institute for Replication (I4R)

Suggested Citation: Dupraz, Yannick (2023) : Recession, Mortality, and Migration Bias: A Comment on Arthi et al. (2022), I4R Discussion Paper Series, No. 25, Institute for Replication (I4R), s.l.

This Version is available at:
<https://hdl.handle.net/10419/270605>

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. 25
I4R DISCUSSION PAPER SERIES

Recession, Mortality, and Migration Bias: A Comment on Arthi et al. (2022)

Yannick Dupraz

March 2023

I4R DISCUSSION PAPER SERIES

I4R DP No. 25

Recession, Mortality, and Migration Bias: A Comment on Arthi et al. (2022)

Yannick Dupraz¹

¹CNRS, Paris/France and Aix-Marseille University, Marseille/France

MARCH 2023

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

Recession, Mortality, and Migration Bias: A Comment on Arthi et al. (2022)

Yannick Dupraz*

2023

Abstract Vellore Arthi, Brian Beach and W. Walker Hanlon (2022) investigate the effect of the Lancashire Cotton Famine on mortality, accounting for the migration response to the downturn. They use difference-in-differences to estimate the effect of the cotton famine on mortality. To account for the migration response to the cotton famine, they construct a linked dataset giving mortality rates by district of residence during the cotton famine, rather than by district of residence at the time of death. They find that the cotton famine increased mortality in cotton-textile producing districts, and that accounting for migration matters, in the sense that their estimates would have been markedly different had they not accounted for it. I check that ABH results are fully reproducible using their data and code, and that their claims are robust to (1) decreasing the age window for building the linked dataset, (2) modifying the specification and (3) computing different standard errors. The only significant discrepancy in results is that I find stronger effects of the cotton famine when I decrease the age window for building the linked dataset, likely because this reduces measurement errors.

1 Introduction

Vellore Arthi, Brian Beach and W. Walker Hanlon (2022), henceforth ABH, investigate the effect of the Lancashire Cotton Famine on mortality, accounting for the migration response to the downturn. The Lancashire Cotton Famine was a

*CNRS & Aix-Marseille University

severe recession geographically concentrated in the cotton textile producing districts of England caused by the disruption to cotton imports during the US Civil War. ABH use difference-in-differences to estimate the effect of the shock on mortality rates, comparing cotton textile districts to other districts, before and after the shock. One problem of this approach is that if the cotton famine induced a migration response away from the cotton producing districts, the estimated effect on mortality could be biased. To overcome this problem, ABH construct a dataset linking individuals observed in the 1861 British Census to the records of death occurring anywhere in England and Wales over the period 1861-1865. They estimate the mortality effect of the cotton famine using this dataset and compare it to the effect they would have estimated if they did not have linked data and had to rely instead on district-level mortality figures absent correction for migration.

ABH make three main claims (p. 231, “Our analysis generates three main sets of findings”): (1) the cotton famine increased mortality in cotton-textile producing districts, especially for older people, (2) mortality increased in the households of cotton workers, but also in the households of non-cotton workers living in cotton districts, in particular those providing non-tradable services and those working in sectors linked to the cotton sector, (3) their new approach for taking migration into account matters; their estimates would have been markedly different had they not accounted for the migration response.

The first claim is made, for example, in the introduction, p. 230:

First, we show that the cotton shortage had an adverse impact on mortality for the population initially residing in cotton districts at the time of the shock, especially for the elderly. We estimate that the

shock generated around 24,000 excess death within the cotton textile districts, equal to 9.5 percent of total deaths. Around 10,000 of these occurred among those aged 55 or over, an increase in deaths of 18.8 percent for that age group.

The second claim is made, for example, in the abstract:

Our results indicate that the recession increased mortality among households employed in the cotton textile industry. We also document localized spillover effects on households providing nontradable services in the areas affected by the recession.

It is also made in the introduction, p. 231:

Our results show both that cotton workers, and the family members of cotton workers, experiences substantial mortality increases as a result of the shock. However, we also show substantial effects among noncotton households residing in the cotton textile areas. [...] Digging deeper, we find evidence that the effect of the shock on noncotton households in cotton regions was particularly severe for those providing nontraded local services as well as those working in sectors sharing input-output linkages to the cotton textile sector.

The third claim is made, for example, in the introduction, p. 231:

We find that [the impact of unobserved migration on estimates of the recession-mortality relationship] is substantial in our setting: while our main linked microdata results show that the downturn raised mortality rates, when we intentionally ignore migration, by inferring treatment

status based on the location of death—and thus adopting a data structure similar to what is commonly used in the literature—we fail to recover this effect. Indeed, in some cases, we find the opposite result.

My replication focuses on computational reproducibility (I checked that all the paper’s results could be obtained using the authors’ own code and data) and on robustness replicability. I considered the following robustness tests: (1) Linking robustness: I checked the robustness of results to a slightly different linking procedure between census data and death records, where I restricted links to those with an age difference of less than $-/+ 2$ years (rather than $-/+ 5$ years). (2) Specification robustness: I tested the robustness of results to varying the set of time-varying district controls in the diff-in-diff specification. (3) Standard error robustness: I checked the robustness of results to clustering standard error by county (rather than districts), to using spatial standard errors following Conley (1999), and to computing permutation p-values using a slightly different algorithm. I also checked the robustness of the paper’s third claim (results with and without accounting for migration are significantly different) to a permutation exercise.

This replication was pre-registered: A pre-analysis plan was sent to the Institute for Replication on the 14th of November 2022.¹

2 Reproducibility

I checked that all results of the paper could be reproduced using the authors’ code and data, available at <https://www.openicpsr.org/openicpsr/project/128521/version/V1/view>. The code was overall very clear and easy to under-

¹Email address: instituteforreplication@gmail.com.

stand, and I had no problem reproducing Figures 1 to 4 and Tables 1 to 5. The only error I could spot is in Table 1 (p. 245): the minimum mortality rate in 1851 is 0.569, and not 0.513, as printed in the paper.

I note that ABH obtained census data through a special agreement and could not make the raw data available. I could therefore not reproduce their data from scratch

3 Robustness Replicability

3.1 Linking Robustness

ABH consider a link (between census data and death records) as valid if “the inferred birth year is no more than five years apart” (p. 240). ABH write that “a five-year threshold” is “standard in the linking literature (see, e.g., Abramitzky et al. 2021)”, but in Abramitzky et al. (2021), —and in Abramitzky et al. (2012, 2014) — “matched pairs are allowed to differ by up to 2 years in reported year of birth”, not five years (p. 871).

In Table 1, I test the robustness of results to restricting links to those with an age difference of less than ± 2 years. I find that the effect of the shortage on mortality is about 25% higher. For example, in ABH, Table 1, column (2), the effect of Cotton District \times Cotton Shortage is 2.024, versus 2.572 in this replication, Table 1, column (2). This is likely because reducing the age window reduces the share of wrong links, therefore reducing measurement errors, which are known to bias results towards zero. In light of this result, in Table 2, I restrict the sample to links with an age difference of less than ± 1 year. This should reduce

Table 1: Baseline Effect of the Shortage Restricting Links to Age Difference \leq Two Years

	DV: Deaths per 1,000 individuals (per year)			
	(1)	(2)	(3)	(4)
Cotton District \times Cotton Shortage	2.862*** (0.447)	2.572*** (0.575)	2.775*** (0.697)	
Cotton Emp. Share \times Cotton Shortage				7.305*** (1.938)
Nearby (0-25 km) \times Cotton Shortage			0.668 (0.687)	0.359 (0.663)
Nearby (25-50 km) \times Cotton Shortage			-0.336 (0.596)	-0.468 (0.595)
Nearby (50-75 km) \times Cotton Shortage			0.342 (0.706)	0.246 (0.696)
District Controls		Yes	Yes	Yes
Observations	1,076	1,076	1,076	1,076
R^2	0.027	0.401	0.403	0.402
Permutation test p-values for effect on cotton districts				
p -values	0.108	0.078	0.061	0.182

Notes: Standard errors clustered by district are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

measurement error even further. I find that the effect of the shortage on mortality is about 60% higher than in ABH, Table 1 (3.270 versus 2.534 in column (2)). Interestingly, while ABH find “some marginally statistically significant evidence of adverse spillover effects” in noncotton districts within 25 km of a cotton district (p. 245), I find no evidence of this when focusing on the best links in Tables 1 or 2. It is therefore likely that this, moderate, spillover was due to measurement error (as individuals affected by the cotton shortage were linked to the wrong district).

If geographical spillover might be due to measurement error, could the results decomposing effects by sector of employment be due to measurement error as well? I Table 3, I reproduce ABH results by sector of employment (Table 4, p. 249 in the paper) restricting links to those with an age difference of less than +/- one

Table 2: Baseline Effect of the Shortage Restricting Links to Age Difference \leq One Year

	DV: Deaths per 1,000 individuals (per year)			
	(1)	(2)	(3)	(4)
Cotton District \times Cotton Shortage	3.230*** (0.499)	3.270*** (0.645)	3.137*** (0.776)	
Cotton Emp. Share \times Cotton Shortage				8.098*** (2.172)
Nearby (0-25 km) \times Cotton Shortage			0.081 (0.782)	-0.291 (0.748)
Nearby (25-50 km) \times Cotton Shortage			-0.562 (0.637)	-0.728 (0.636)
Nearby (50-75 km) \times Cotton Shortage			-0.156 (0.795)	-0.280 (0.783)
District Controls		Yes	Yes	Yes
Observations	1,076	1,076	1,076	1,076
R^2	0.029	0.374	0.375	0.374
Permutation test p-values for effect on cotton districts				
p-values	0.091	0.063	0.078	0.929

Notes: Standard errors clustered by district are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

year. The effect of the cotton shortage is larger, but I replicate the findings of ABH (mortality increased in the households of cotton workers, but also in the households of non-cotton workers, notably those providing non-tradeable services and those working in sectors linked to cotton). The effect on households whose head was employed in transport is greatly increased and becomes statistically significant when restricting the age window for linking.

3.2 Specification Robustness

In this Section, I test the robustness of results to varying the set of time-varying district controls in the diff-in-diff specification. ABH present results of an esti-

Table 3: Decomposing Effect by Sector of Employment, Restricting Links to Age Difference \leq One Year

	DV: Deaths per 1,000 individuals (per year)			
	(1)	(2)	(3)	(4)
Cotton District \times Cotton Shortage	1.845** (0.796)		1.676** (0.821)	
Head Employed in Cotton \times Cotton Shortage		1.684*** (0.556)	0.963 (0.578)	
Head Employed in Cotton \times Cotton Dist. \times Shortage				3.102*** (0.953)
Head Employed in Non-Tradeables \times Cotton Dist. \times Shortage				3.674*** (1.041)
Head Employed in Linked IO \times Cotton Dist. \times Shortage				1.784* (0.954)
Head Employed in Other \times Cotton Dist. \times Shortage				0.914 (1.103)
Head Employed in Tradeable Manuf. \times Cotton Dist. \times Shortage				1.048 (1.172)
Head Employed in Transport \times Cotton Dist. \times Shortage				1.822* (1.049)
Head Outside Labor Force \times Cotton Dist. \times Shortage				-2.026 (1.627)
Observations	32,677	32,677	32,677	32,677
R^2	0.023	0.023	0.023	0.024

Notes: Standard errors clustered by district are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

mation without time-varying district controls and with the following time-varying district controls: the log of population density, the share of individuals enumerated in the census with a “linkable” name, the share of population in each of the categories under 15, 15-54, over 54, and region by period F.E. I consider a larger set of time-varying district controls: I add to the previous controls the share of individuals with a “linkable” name in each cohort, and district latitude, longitude, and area interacted with the “Cotton Shortage” time dummy. I also consider population shares in finer cohorts.² I then select an optimal set of controls using the post-double machine learning selection algorithm of Belloni et al. (2014). This algorithm takes all the controls with their squares and cross term interactions and

²Rather than under 15, 15-54, and over 54, I consider under 15, 15-24, 25-34, 35-44, 45-54, 55-64, and over 64.

uses a LASSO procedure to select significant predictors of the treatment and outcomes. It then uses as controls only the significant predictors of both treatment and outcome. Table 4 shows that including this extended set of controls barely changes the effect of the Cotton Shortage on mortality (compare column (2) and (3)).

Table 4: Baseline Effect of the Shortage, Specification Robustness

	DV: Deaths per 1,000 individuals (per year)				
	(1)	(2)	(3)	(4)	(5)
Cotton District \times Cotton Shortage	2.194*** (0.463)	2.024*** (0.519)	2.214*** (0.559)	2.727*** (0.648)	
Cotton Emp. Share \times Cotton Shortage					7.331*** (1.684)
Nearby (0-25 km) \times Cotton Shortage				1.327** (0.647)	1.065* (0.631)
Nearby (25-50 km) \times Cotton Shortage				-0.037 (0.602)	-0.152 (0.602)
Nearby (50-75 km) \times Cotton Shortage				0.273 (0.690)	0.179 (0.682)
District Controls		Baseline	Extended set	Extended set	Extended set
Observations	1,076	1,076	1,076	1,076	1,076

Notes: Standard errors clustered by district are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Baseline district controls are $\ln(\text{population density})$, share of individuals enumerated in the census with a “linkable” name, the share of the population in each of the following age categories (under 15, 15–54, and over 54, with 15–54 as the omitted category), and region-by-period fixed effects. The Extended set of controls is selected using the post-double machine learning selection algorithm of Belloni et al. (2014). The following controls are selected: region by period fixed effects, latitude and longitude times period, linkable share under 15, 25–34, 55–64, and over 64, share of population under 15.

3.3 Standard Error Robustness

ABH are rightly worried about spatial autocorrelation and tackle this problem by using a randomization inference approach: they run their analysis on all possible permutations of treatment having the same spatial correlation as the one present in their data and compute the share of permutation results that are larger than their estimates. In the Online Appendix (p. 58), they write that they have also generated results clustering standard errors by county (a larger geographic unit

than the one they use) and implementing spatial standard errors following Conley (1999), and that this reduced standard errors. Because of lack of space, these results are not included in the paper or in the appendix.

In Table 5, I check the robustness of the results in ABH to clustering standard errors by county, and implementing Conley (1999) standard errors, with distance cutoff of 50 and 100 km. As stated by ABH in their appendix, all these alternative standard errors are smaller than the baseline standard errors (clustered by district).

I also produce alternative permutation p-values for testing the significance of the effect of the cotton shortage. In their paper, ABH simulate as many placebo datasets as there are districts (538). In each dataset, they create a placebo cotton region centered on a different anchor district. Because there are 24 cotton districts in the true data, they create a cotton region consisting of the anchor district plus the 23 closest districts. They then estimate a distribution of 538 placebo effects and compute a p-value by computing the share of placebo effects larger than the true effect. I implement a slightly different permutation procedure where, to create a placebo cotton region, instead of selecting the 23 districts closest to the anchor district, I select the districts within 45 km. I chose 45 km because this distance produces an average of 25 treated districts. As shown in the bottom of Table 5, this alternative permutation test actually produces slightly lower p-values.

ABH produce permutation test p-values for Tables 2 and 3 of their paper, but not for Table 5 (p. 251), the table displaying one of their most important claims (claim 3): the fact that accounting for migration has a meaningful impact on the estimated effect of the cotton shortage on mortality. In this table, ABH test whether their result accounting for migration (panel C) is statistically different from a naive result not accounting for migration (panel A). This test does not

take into account the problem of spatial auto-correlation.

In Table 6, I test the robustness of claim 3 to a permutation test. For each of 538 permutations (each creating a different placebo cotton region), I compute the difference between the coefficient of Panel A (not accounting for migration) and the coefficient of Panel C (accounting for migration). I then obtain p-values by computing the share of these differences that are larger (in absolute value) than the difference obtained on the true data. As shown in the bottom of Table 6, results accounting for migration are statistically different than results not accounting for migration, even when accounting for spatial autocorrelation: the permutation test p-value is lower than 0.05 for the cohort under 15 and the cohort 25–34. It is lower than 0.1 for the cohorts 45-54 and 55-64.

4 Conclusion

The results of ABH are fully reproducible using their own data and code. Their results are also robust to a specification robustness and several standard-error robustness. I find different results only when decreasing the age window for linking census data and death records, but I find a *larger* effect of the cotton downturn on mortality, likely because there is less measurement error. Therefore, I conclude that the main claims of ABH are all justified. It should be noted that (1) ABH obtained census data through a special agreement and could not make the raw data available, I could therefore not reproduce their data from scratch, (2) I did not conduct direct replicability (using the same procedure on different data) of conceptual replicability (using different data and procedures).

References

- Abramitzky, R., Boustan, L. and Eriksson, K. (2012). Europe's tired, poor, and huddled masses: Self selection and economic outcomes in the age of mass migrations, *American Economic Review* **102**(5): 1832–1856.
- Abramitzky, R., Boustan, L. and Eriksson, K. (2014). A nation of immigrants: Assimilation and economic outcomes in the age of mass migrations, *Journal of Political Economy* **122**(3): 467–506.
- Abramitzky, R., Boustan, L., Eriksson, K., Feigenbaum, J. and Perez, S. (2021). Automated linking of historical data, *Journal of Economic Literature* **59**(3): 865–918.
- Arthi, V., Beach, B. and Hanlon, W. W. (2022). Recessions, mortality, and migration bias: Evidence from the lancashire cotton famine, *American Economic Journal: Applied Economics* **14**(2): 228–55.
- Belloni, A., Chernozhukov, V. and Hansen, C. (2014). Inference on treatment effects after selection among high-dimensional controls, *Review of Economic Studies* **81**(2): 608–650.
- Conley, T. G. (1999). GMM estimation with cross sectional dependence, *Journal of Econometrics* **92**(1): 1–45.

Table 5: Baseline Effect of the Shortage, Standard Error Robustness

	DV: Deaths per 1,000 individuals (per year)			
	(1)	(2)	(3)	(4)
Cotton District \times Cotton Shortage	1.974*** (0.537) [0.464] ((0.402)) [[0.458]]	2.160*** (0.702) [0.645] ((0.493)) [[0.461]]	2.715*** (0.769) [0.613] ((0.526)) [[0.490]]	
Cotton Emp. Share \times Cotton Shortage				8.319*** (2.205) [1.954] ((1.355)) [[1.262]]
Nearby (0-25 km) \times Cotton Shortage			1.447 (0.925) (0.970) ((0.599)) [[0.637]]	1.142 (0.915) [0.966] ((0.579)) [[0.608]]
Nearby (25-50 km) \times Cotton Shortage			0.735 (0.686) [0.574] ((0.404)) [[0.379]]	0.673 (0.681) [0.586] ((0.406)) [[0.386]]
Nearby (50-75 km) \times Cotton Shortage			0.143 (0.646) [0.621] ((0.427)) [[0.475]]	0.118 (0.640) [0.617] ((0.425)) [[0.475]]
District Controls		Yes	Yes	Yes
Observations	1,076	1,076	1,076	1,076
Permutation test p-values for effect on cotton districts				
<i>p</i> -values	0.232	0.087	0.072	0.026
Alternative permutation test p-values				
<i>p</i> -values	0.206	0.052	0.046	0.022

Notes: Standard errors clustered by district are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered by county are in []. Conley standard errors with a distance cutoff of 50 km are in (()). Conley standard errors with a distance cutoff of 100 km are in [[]].

Table 6: Aggregate-to-Linked Comparison: Does Migration Have a Meaningful Impact on Results? Robustness to permutation p-values

	DV: Deaths per 1,000 individuals (per year)						
	Under 15 (1)	Age 15-24 (2)	Age 25-34 (3)	Age 35-44 (4)	Age 45-54 (5)	Age 55-64 (6)	Over 64 (7)
<i>Panel A. Actual aggregate data (Drawn from Registrar's Reports)</i>							
Cotton District	-4.660***	-1.325***	-1.511***	-0.499	-0.624	0.267	2.094
× Cotton Shortage	(1.019)	(0.284)	(0.346)	(0.496)	(0.632)	(0.793)	(1.807)
<i>Panel B. Aggregate-like linked data (Links assigned to district of death)</i>							
Cotton District	-4.248	-1.361**	-0.952	0.052	-2.035	2.104	4.989
× Cotton Shortage	(3.035)	(0.654)	(0.975)	(1.399)	(1.918)	(3.285)	(6.991)
Different from panel A?							
p-value	0.889	0.961	0.596	0.702	0.485	0.587	0.674
permutation p-value	.870	0.961	0.708	0.766	0.615	0.504	0.690
<i>Panel C. Preferred migration-corrected linked data (Links assigned to district of enumeration)</i>							
Cotton District	0.224	0.171	0.894	1.512	3.066***	6.740***	13.477***
× Cotton Shortage	(1.078)	(0.551)	(0.678)	(0.939)	(1.086)	(1.861)	(3.899)
Different from panel A?							
p-value	0.000	0.018	0.002	0.049	0.003	0.002	0.009
permutation p-value	0.011	0.230	0.013	0.331	0.074	0.071	0.139

Notes: Standard errors clustered by district are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.