No. 5
I4R DISCUSSION PAPER SERIES

# A Successful Replication of "Dust Pollution From the Sahara and African Infant Mortality"

Nikolai M. Cook

October 2022

I4R Discussion Paper Series

I4R DP No. 5

# A Successful Replication of "Dust Pollution From the Sahara and African Infant Mortality"

**Nikolai M. Cook**

*Wilfried Laurier University, Waterloo, ON/Canada*

October 2022



E-Mail: joerg.peters@rwi-essen.de
RWI – Leibniz Institute for Economic Research

Hohenzollernstraße 1-3
45128 Essen/Germany

www.i4replication.org

ISSN: 2752-1931

# A Successful Replication of "Dust Pollution From the Sahara and African Infant Mortality"

Nikolai M Cook*
Wilfrid Laurier University

October 12, 2022

## 1 Summary

This analysis is an independent replication of Heft-Neal et al. (2020).[1] The original authors (HBBVB) provide evidence that particulate matter air pollution increases infant mortality in 30 African nations between 2000 and 2015. They provide three effect estimates. Using ordinary least squares, a 10 $\mu g/m^3$ increase in $PM_{2.5}$ exposure results in an estimated 8.6% increase in infant mortality. Using dust in the Bodélé depression as an instrumental variable, the same exposure increases infant mortality by 23.6%. Using rainfall in the Bodélé depression, the same exposure increases infant mortality by 24.3%. Using similar data and independently developed procedures I find corresponding estimates of 3.4%, 31.0%, and 29.7%.

## 2 Data

For their dependent variable, HBBVB use data provided by The Demographic and Health Surveys Program. They use 990,696 individual birth records from 65 surveys conducted in 30 countries between 2001 and 2015. The dependent variable, infant mortality, is a binary variable which takes a value of one when a child is reported to have died within twelve months of birth. Average infant mortality in their sample is 71 deaths per 1,000 births. Using the same surveys and dependent variable I examine 1,007,474 individual birth records. Average infant mortality in my sample is 72 deaths per 1,000 births.

For their primary independent variables, HBBVB use two datasets provided by Van Donkelaar et al. (2016); V4.GL.02 and V4.GL.02 (Dust and Sea Salt Removed). These provide annual concentrations of ground level $PM_{2.5}$ by combining ground station readings, meteorological models, and remote sensing. The data offer global coverage from 1998 through 2016 at a spatial resolution of 0.01 degrees (approximately 1.11 kilometers). As the data is annual, HBBVB use birth month to construct post-birth $PM_{2.5}$ exposure. For example, if a child is born in the second month of the year, that child's post-birth $PM_{2.5}$ exposure is equal to $\frac{10}{12}$ the current year + $\frac{2}{12}$ following year's local $PM_{2.5}$. While HBBVB do not report average $PM_{2.5}$ exposure, I find average post-birth exposure of 25.23 $\mu g/m^3$ (over five times the 2021 WHO's recommendations). HBBVB also create two instrumental variables. The first uses $PM_{2.5}$ in the Bodélé depression as an instrument for local levels. The second uses rainfall in the Bodélé (which lowers the amount of $PM_{2.5}$ transported from this source). They use precipitation data provided by CHIRPS which consists of average

*This analysis was preregistered prior to data access with the Open Science Foundation on August 15, 2022 and may be accessed here. Data access was granted by the Demographic and Health Surveys Program on August 16, 2022. The original authors' response to this replication is available here. I am grateful to Abel Brodeur for helpful conversations. Errors are my own.

[1]Following the definitions provided by the Institute for Replication, a **computational reproduction** uses same data and procedures as the original authors, a **robustness replication** uses same data and different procedures, a **direct replication** uses different data but same procedures, and a **conceptual replication** uses different data and different procedures. This analysis is a **conceptual replication** as it uses same (although independently collected) data for some parts and different data in others, coupled with different and independently developed procedures. The article is *also likely to be* computationally reproducible. Original (limited due to sharing restrictions) data and procedures can be found here. The original authors' response to this replication is available here.

monthly rainfall over the African continent, and is available from 1981 through 2022 (as of writing) at a 0.05 degree resolution (Funk et al., 2015).

HBBVB include additional independent variables. First, local average annual rainfall using the CHIRPS data, which averages 123 cm (this analysis finds an average of 103.60 cm). Second, local average temperatures. The original article does not provide details on the data used or the sample average temperature. I use annual 0.5 degree resolution temperature data provided by Fan and Van den Dool (2008). My sample average is 24.52 degrees Celsius. Third, variables from the DHS surveys; clean cooking fuel, mother completed primary school, mother's age, mother's age$^2$, child sex, child birth order, and twin child. I do the same. Fourth, HBBVB include satellite measured night-time lights as a proxy for economic activity. I do not include this control, following recent concerns that night-time lights may not reflect economic activity, particularly in low density areas (Gibson et al., 2020, 2021).

# 3 Methods

Following HBBVB, I first estimate the ordinary least squares:

$$Mortality_{i,j,c,m,t} = \beta \times PM_{2.5} + \mu \mathbf{X}_{i,j,c,m,t} + \gamma_j + \delta_t + \nu_{c,m} + \varepsilon_{i,j,c,m,t} \quad (1)$$

Where $Mortality_{i,j,c,m,t}$ takes the value zero unless individual $i$ in cluster $j$ in country $c$ born during month $m$ and year $t$ dies within twelve months of birth. The primary coefficient of interest $\beta$ estimates the effect post-birth $PM_{2.5}$ exposure has on mortality. $\mathbf{X}_{i,j,c,m,t}$ represents a vector of controls. Fixed effects are included for DHS-cluster ($\gamma_j$), birth-year ($\delta_t$) and country-month ($\nu_{c,m}$). In Table 1, I present coefficients with and without $\mathbf{X}_{i,j,c,m,t}$ to demonstrate robustness not offered in the original article.

When applying an instrumental variable approach, I estimate the first stage as:

$$PM_{i,j,c,m,t} = \lambda DV_{i,t} + \mu \mathbf{X}_{i,j,c,m,t} + \gamma_j + \delta_t + \nu_{c,m} + \varepsilon_{i,j,c,m,t} \quad (2)$$

Where $PM_{i,j,c,m,t}$ denotes post-birth exposure to $PM_{2.5}$ for individual $i$ in cluster $j$ in country $c$ born during month $m$ and year $t$. The instrument $DV_{i,t}$ represents the share of $PM_{2.5}$ from natural sources for cluster $j$ multiplied by the dust ($PM_{2.5}$) level in the Bodélé (this is replaced with Bodélé rainfall for the second instrument). The second stage is identical to Equation 1, with predicted levels of local $PM_{2.5}$ in place of directly measured local $PM_{2.5}$.

A summary of the differences between the original article and this analysis is as follows. First, the outcome variable data was independently collected and processed. This includes the DHS surveys and the matching procedures for connecting them to GPS data which likely explains differences in the number of individual birth records. Second, this analysis excludes night-time lights following recent concerns (Gibson et al., 2021). Third, the original article does not make clear its source for temperature data. I use data from Funk et al. (2015) which may have a different spatial resolution and aggregation method than the original. Fourth, this analysis applies both space- and time-clustered standard errors where the original article is silent. Despite these differences, in the next section I present estimates that are of similar sign, magnitude, and statistical significance as HBBVB.

# 4 Results

Table 1 presents results. In all columns, the dependent variable is mortality which takes a value of one if a child dies within twelve months of birth. The primary independent variable is post-birth $PM_{2.5}$ exposure. Columns 1 and 2 present regression coefficients for the ordinary least squares of Equation 1. In Column 1, a 10 $\mu g/m^3$ increase in $PM_{2.5}$ exposure is associated with a 5.5% increase in mortality. The sample average is 72 deaths per 1,000 births, implying that this increase in post-birth $PM_{2.5}$ exposure is associated with an additional $0.055 * 72 = 3.96$ deaths. In Column 2, I now include the same controls as HBBVB with the above-noted exception of night-time lights. The estimated effect of post-birth $PM_{2.5}$ falls to 3.4% but remains statistically significant. Further, the additional controls offer a 'sanity' check. Cleaner cooking fuel is associated with a negative effect on mortality, following the improvements to indoor air quality. Demographic controls also have the expected sign such as better educated mothers. As in HBBVB, I include precipitation and temperature, neither seem to affect mortality. All columns include fixed effects to control for time invariant determinants of mortality by location (DHS-cluster), local seasonality (country-month), and shocks common to the entire sample (birth-year).

Instrumental variables estimates are presented in Columns 3 through 6. In column 3 and 4, I use $PM_{2.5}$ in the Bodélé (which is largely from natural sources i.e. dust) multiplied by the share of local $PM_{2.5}$ from natural sources as an instrument to predict local $PM_{2.5}$. As in the original article, the coefficients of $PM_{2.5}$ on mortality are positive, statistically significant, and larger than the OLS coefficients. HBBVB note at this point that instrumental variable estimates identify 'local average treatment effects', namely the average effect of additional $PM_{2.5}$ on a child for whom additional dust in the Bodélé increases local $PM_{2.5}$. In Column 4, which now includes the suite of controls as in the original article, a 10 $\mu g/m^3$ increase in $PM_{2.5}$ exposure is associated with a 31.0% increase in mortality. The sample average for mortality is 72 deaths per 1,000 births, implying that this increase in post-birth $PM_{2.5}$ exposure is associated with an additional $0.310 * 72 = 22.32$ deaths. In Columns 5 and 6 (which use the rainfall instrument) the results are similar.

# 5 Conclusion

This analysis is an independent conceptual replication of Heft-Neal et al. (2020) and finds effect sizes of similar sign, magnitude, and statistical significance. It uses the same (although independently downloaded) data and different procedures which were developed solely with reference to the text of the published article. While HBBVB provide relatively easy access to replication materials, this analysis opted to replicate as independently as possible. This accomplishes at least three things. First, HBBVB is sufficiently detailed that in the absence of its replication package it can be replicated. Second, it is possible and encouraged to independently replicate published articles regardless of replication packages. Third, future researchers may still complete a less intensive computational reproduction.

Table 1: Impacts of $PM_{2.5}$ on Infant Mortality

| Estimation Method | OLS | | Dust Instrument | | Rainfall Instrument | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Heft-Neal et al. (2020) | N/A | 0.086*** | N/A | 0.236*** | N/A | 0.243*** |
| | | (0.028) | | (0.060) | | (0.061) |
| | | | Dependent Variable: Mortality | | | |
| $PM_{2.5}$ (Per 10 $\mu g/m^3$) | 0.055*** | 0.034* | 0.394*** | 0.310*** | 0.372*** | 0.297*** |
| | (0.016) | (0.017) | (0.074) | (0.065) | (0.082) | (0.074) |
| Clean Cooking Fuel | | -0.158*** | | -0.111*** | | -0.113*** |
| | | (0.032) | | (0.032) | | (0.034) |
| Mother has Primary Educ. | | -0.165*** | | -0.159*** | | -0.159*** |
| | | (0.028) | | (0.027) | | (0.028) |
| Mother's Age | | -0.123*** | | -0.123*** | | -0.123*** |
| | | (0.007) | | (0.007) | | (0.007) |
| Mother's Age$^2$ | | 0.001*** | | 0.001*** | | 0.001*** |
| | | (0.000) | | (0.000) | | (0.000) |
| Child is Female | | -0.154*** | | -0.154*** | | -0.154*** |
| | | (0.010) | | (0.010) | | (0.010) |
| Child Birth Order | | 0.090*** | | 0.089*** | | 0.089*** |
| | | (0.003) | | (0.004) | | (0.004) |
| Child is Twin | | 2.272*** | | 2.274*** | | 2.274*** |
| | | (0.097) | | (0.097) | | (0.097) |
| Local Precipitation (cm) | | 0.002 | | 0.014*** | | 0.013*** |
| | | (0.002) | | (0.004) | | (0.004) |
| Local Temperature (C) | | -0.002 | | -0.002 | | -0.002 |
| | | (0.002) | | (0.002) | | (0.002) |
| DHS-Cluster FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Birth-Year FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Country-Month FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Birth Records | 1,007,320 | 1,007,320 | 1,007,320 | 1,007,320 | 1,007,320 | 1,007,320 |
| First Stage F-Stat. | | | 89.66 | 98.70 | 59.21 | 67.27 |
| Stock-Yogo Crit. Value | | | 16.38 | 16.38 | 16.38 | 16.38 |

All coefficients scaled to represent percentage change in mortality. For example, increasing post-birth $PM_{2.5}$ exposure by 10 $\mu g/m^3$ increases infant mortality by 5.5% in the first column. Sample average mortality is 72 deaths per 1,000 births, suggesting this increase would causes 3.96 additional deaths. Standard errors are clustered at the DHS-cluster and birth-year levels. All estimates do not use sample weights - their introduction does not qualitatively alter results. (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$).

# References

Fan, Y. and Van den Dool, H. (2008). A global monthly land surface air temperature analysis for 1948–present. *Journal of Geophysical Research: Atmospheres*, 113(D1).

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., et al. (2015). The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific data*, 2(1):1–21.

Gibson, J., Olivia, S., and Boe-Gibson, G. (2020). Night lights in economics: Sources and uses 1. *Journal of Economic Surveys*, 34(5):955–980.

Gibson, J., Olivia, S., Boe-Gibson, G., and Li, C. (2021). Which night lights data should we use in economics, and where? *Journal of Development Economics*, 149:102602.

Heft-Neal, S., Burney, J., Bendavid, E., Voss, K. K., and Burke, M. (2020). Dust pollution from the sahara and african infant mortality. *Nature Sustainability*, 3(10):863–871.

Van Donkelaar, A., Martin, R. V., Brauer, M., Hsu, N. C., Kahn, R. A., Levy, R. C., Lyapustin, A., Sayer, A. M., and Winker, D. M. (2016). Global estimates of fine particulate matter using a combined geophysical-statistical method with information from satellites, models, and monitors. *Environmental science & technology*, 50(7):3762–3772.