

Replication of Bleakley 2007
Pre Analysis Plan
Author: David Roodman, GiveWell

Goal:

We wish to explore whether the results of Bleakley, 2007[1] are robust to replication and additional testing.

Plan:

We plan to replicate Bleakley, 2007 including re-collection of all data from original sources outlined in the paper and re-creation of as many of the conducted analyses as possible. Some controls introduced in the original for robustness testing may be excluded if collecting them proves especially costly. In order to check robustness, we will also carry out new analysis, most of which will vary original specifications in some way. Some of these are pre-registered below while others may be decided upon after analysis begins. We are pre-registering in order to credibly inform readers about which are which.

Specific Analyses:

We will attempt to match as many published tables and graphs in the original as possible. Exceptions for reasons of data availability may include Panel B of Table III, the second half of Figure III, and the “Full controls” rows of Table VI. If any mismatches arise, which is to be expected, we will seek to work with Hoy Bleakley to narrow or explain them. Additional analyses may include: testing for sensitivity to any data or coding errors exposed in the original; performing two-stage least squares instead of the original’s indirect least squares in order to obtain proper confidence intervals for instrumental variables point estimates; pure time-series versions of the sequential cross-sections (SCS) analysis, in which samples are restricted to areas of above-average baseline prevalence, and the intent-to-treat variable is years of exposure to the deworming campaign (Exp_{ik}) rather than the interaction thereof with baseline prevalence; more-conservative error-clustering choices, such as clustering county-level estimates by state rather than State Economic Area; and re-doing the two-stage assessment of whether the hookworm campaign helps explain the convergence in long-term earnings between low- and high-prevalence areas (equation 5 and Table VI) in a way that factors the uncertainty of the estimates from the first stage into the second, either analytically or by bootstrapping.