The data files (in Stata format) that are supplied are simple excerpts from a master file that is a merge of the data at the individual level produced from the survey modules, each corresponding to a questionnaire. The arsenic data are also merged to the individuals in the data set. Linerge-specific arsenic levels were constructed using the familial relationships among individuals as well as the location information.

To protect the confidentiality of the respondents, no person-id's or locality information is provided in the data sets, beside numerical codes for village ('villmany' variable)

To replicate the numbers in the tables, there is a do file that runs on the data sets named in the file, which are also included. The do files are labeled according to the table they produce. So, table1.do delivers the numbers in table 1 etc. Similarly, there are files that deliver the numbers in the figures. These files are named according to the figure number. So, figure3.do delivers the numbers graphed in figure 3.

The table3a.do file (producing Table 3) takes approximately 1.75 hours to complete because of the number of instruments constructed. All the rest of the programs finish quickly.

The do files that create figures data are used to create PowerPoint graphs, which are included.

All programs were run on Stata 14. Two Stata packages are used that are not part of the standard Stata package:

cluster2 from http://fmwww.bc.edu/RePEc/bocode/c permits 2-level clusters.

cmp (st0224 and st0224_1 ) is a mixed process estimator that permits multilevel random effects described in the *Stata journal* (https://www.stata-journal.com/article.html?article=st0358) and in *Stata Journal* 11: 159–206.