Maria Pía Basurto
Ramiro Burga
José Luis Flor Toro
César Huaroto

Walking on solid ground

A replication study on Piso Firme's impact

August 2015

Replication Paper 7

Health

International Initiative for Impact Evaluation

About 3ie

3ie is an international grant-making NGO promoting evidence-informed development policies and programmes. We are the global leader in funding and producing high-quality evidence of what works, how, why and at what cost. We believe that better and policy-relevant evidence will make development more effective and improve people's lives.

3ie Replication Paper Series

The 3ie Replication Paper Series is designed to be a publication and dissemination outlet for internal replication studies of development impact evaluations. Internal replication studies are those that reanalyse the data from an original paper in order to validate the results. The series seeks to publish replication studies with findings that reinforce an original paper, as well as those that challenge the results of an original paper. To be eligible for submission, a replication study needs to be of a paper in 3ie's online Impact Evaluation Repository and needs to include a pure replication. 3ie invites formal replies from the original authors. These are published on the 3ie website together with the replication study.

The **3ie Replication Programme** also includes grant-making windows to fund replication studies of papers identified on the candidate studies list. Requests for proposals are issued one to two times a year. The candidate studies list includes published studies that are considered influential, innovative or counterintuitive. The list is periodically updated based on 3ie staff input and outside suggestions. The aim of the 3ie Replication Programme is to improve the quality of evidence from development impact evaluations for use in policymaking and programme design.

About this report

This paper was funded through 3ie's Replication Window with generous funding from an anonymous donor. All content is the sole responsibility of the authors and does not represent the opinions of 3ie, its donors or the 3ie Board of Commissioners. Any errors and omissions are the sole responsibility of the authors. Comments and queries should be directed to the corresponding author, Maria Pia Basurto at mbasurto@ucsc.edu.

Suggested citation: Basurto, MP, Burga, R, Flor Toro, JL and Huaroto, C, 2015. *Walking on solid ground: a replication study on Piso Firme's impact,* 3ie Replication Paper 7. Washington, DC: International Initiative for Impact Evaluation (3ie).

3ie Replication Paper Series executive editor: Annette N Brown
Managing editor: Benjamin DK Wood
Assistant managing editor: Jennifer Ludwig
Production manager: Brigid Monaghan
Layout assistant: Jane Burke
Copy editor: Jaime L Jarvis
Proof reader: Yvette Charboneau
Cover design: John F McGill
Printer: VIA Interactive

Walking on solid ground: a replication study on *Piso Firme's* impact

María Pía Basurto
UC Santa Cruz

Ramiro Burga
Barcelona Graduate School of Economics

José Luis Flor Toro
Pontificia Universidad Católica del Perú

César Huaroto
Pontificia Universidad Católica del Perú

3ie Replication Paper 7
August 2015

International Initiative for Impact Evaluation

Acknowledgments

This replication study was funded and facilitated by the International Initiative for Impact Evaluation (3ie) as part of their replication series. The funders had no role in writing the analysis plan or the draft or final reports. We are thankful to the original authors for kindly sharing their codes, raw dataset and methodological documents, and for their support throughout the replication process.

Abstract

There is scarce literature outside the field of medical science that credibly identifies the effects of housing quality on health and socioeconomic outcomes, including the health of children and mothers, and what exists has originated primarily in the field of medical research. One such study, 'Housing, Health and Happiness', by Cattaneo *et al.* (2009), is a quasi-experimental impact evaluation of a government intervention (*Piso Firme*) that replaced dirt floors with cement in Mexican households. Cattaneo *et al*.'s identification strategy and the magnitude of treatment effects found in HHH2009 make it one of the most influential studies on the subject.

In our replication of 'Housing, Health and Happiness', we first perform a pure replication by i) using raw data from authors and other public sources, ii) replicating the authors' sampling strategy, iii) reconstructing all variables in the analysis, iv) replicating figures and tables using the same methodologies proposed by the authors and v) comparing results. We then perform a measurement and estimation analysis to check the robustness of results through i) analysis on the parallel-trends hypothesis using socioeconomic variables different to those considered in the original study, ii) different imputation strategies, iii) alternative definition of mothers' satisfaction measures and iv) using the intervention as an instrument to retrieve intention-to-treat effects. Finally, our theory of change analysis explores the programme's heterogeneous effects.

In our pure replication, we do not find any major discrepancies with the results described by Cattaneo *et al*. Our measurement and estimation analysis generally finds the results to be robust to different types of analysis. In the theory of change analysis, we find that households with high initial levels of cement-floor coverage benefitted significantly less from *Piso Firme*.

Keywords: replication study, housing, health and happiness, housing upgrade programmes

Contents

List of figures and tables

Abbreviations and acronyms

ENIGH	*Encuesta Nacional de Ingresos y Gastos de los Hogares* (National Household Income and Expenditure Survey)
HHH2009	Housing, Health and Happiness
MAR	Missing at Random
MNAR	Missing Not at Random
MCAR	Missing Completely at Random

1. Introduction

The pure replication of Cattaneo *et al.*'s 'Housing, Health and Happiness' study (2009), which we refer to as 'HHH2009', consists of three main parts: sample replication using the Mexican census for the year 2000, replication of figures using the Mexican National Household Income and Expenditure Survey (*Encuesta Nacional de Ingresos y Gastos de los Hogares*, or ENIGH) and replication of tables using the authors' raw data. The authors kindly provided their raw materials and Stata .do files to construct the tables. We downloaded ENIGH datasets from the Mexican National Institute of Statistics and Geography, where we also obtained the census data. We were able to reproduce all figures except those containing variables that were modified to be in real values by using a base year and index to deflate nominal values. Nonetheless, given similarity of the conclusions and trends, we do not see these discrepancies as threatening to the validity of the replication study results.

Our measurement and estimation analysis builds on the pure replication results to examine the robustness of Cattaneo *et al.*'s results to an alternative method for imputation of missing values and an alternative estimation methodology. We find that the study results are robust to these changes. In the theory of change analysis, we broaden the analysis by including heterogeneous effects by initial condition of share cement floor in the house. Because the variable we use to find the heterogeneous effect is endogenous – selection into the initial share of cement flooring – this section should be interpreted as suggestive.

2. Motivation

Inadequate housing is a multidimensional problem that affects a significant portion of people all over the developing world. A house is considered a slum if it lacks access to improved water, access to improved sanitation facilities, sufficient living area, durability (quality of the building) or secure tenure (UN-Habitat 2010). Between 2003 and 2010, the urban population of developing regions that was classified as living in slums (having at least one feature of inadequate housing) dropped from 43 per cent to 33 per cent. In Latin America and the Caribbean, roughly 24 per cent of the urban population lived in slums in 2010 (UN-Habitat 2003; 2010).

The literature that studies how upgrades to housing and slums affect health and socioeconomic outcomes has come primarily from the medical field, using cross-sectional relationships studies.[1] Despite the increasing use of experiments in empirical economic literature, only a few papers use experimental designs to study the causal effects of housing improvements (Galiani *et al.* 2005; Galiani *et al.* 2009; Devoto *et al.* 2012; Galiani *et al.* 2015). Cattaneo *et al.*'s HHH2009 study is an important and extensively cited paper.

Part of the motivation to replicate HHH2009 comes from the Cattaneo *et al.*'s extensive use of dummy variable adjustment imputations in their regressions and other imputation methods in their dataset. In our replication we use alternative imputation methods and find that the results are robust. Additional data available publicly from ENIGH allows us to test whether the parallel-trends assumption between treatment and control groups holds. Finally, we wanted to replicate sample selection, as the HHH2009 methodology was different from typical studies.

[1] See Turley *et al.* (2013) for a literature review.

3. Pure replication

To perform a pure replication of results in HHH2009, we start by replicating the sample selection. We then replicate the main results from the paper.

3.1 Replication of the sampling strategy

Beneficiary selection for the *Piso Firme* programme was not random but was implemented at the state level. Cattaneo *et al.* identify two neighbouring states, Coahuila and Durango, where only the former had implemented *Piso Firme* by 2005, the time of data collection for the programme evaluation. To select a comparable treatment and control group, the authors implemented the following three-stage sampling strategy. First, they geographically restricted the sample to households residing in cities along the border between Durango and Coahuila: the 'twin cities' of Gómez Palacio and Lerdo (control group) and Torreón (treatment). Second, they selected a random sample of treated households using administrative records from *Piso Firme*. Third, in order to select a comparable sample of control households, the authors first identified comparable census blocks in Gómez Palacio and Lerdo that were geographically close to the border with Torreón[2] and then randomly chose households from that sample, conditional on having some dirt floor in 2000.

To define comparable blocks, the authors constructed four variables at the census-block level: (i) proportion of blocks in each census block with dirt floors, (ii) proportion of households in each census block with dirt floors, (iii) number of children between birth and 5 years and (iv) number of households. Finally, for each pair of randomly selected treatment and potential control census blocks, the authors calculated the L-infinite distance[3] between them and 'selected as control areas those census blocks that were closest to the treated blocks in terms of this distance measure' (HHH2009, pp.81). Finally, within the selected control census blocks, households were randomly drawn – again, conditional on having some dirt floor in 2000.

Because random selection cannot be replicated, we only attempt to replicate the selection of the control census blocks. By constructing a cluster ID code,[4] we were able to merge the year 2000 Mexican census with the authors' raw dataset. We were able to match 73 of the 84 treatment census blocks used in HHH2009; the remaining 11 blocks were not traceable back to the 2000 census.

Even though the authors' raw dataset includes variables from the 2000 census for their study sample, we found that the observations corresponding to the 11 untraceable blocks were missing values. In addition, one of these 11 untraceable blocks did not have a census block ID in the authors' dataset. We do not know why these blocks were not included in the

[2] Unfortunately, HHH2009 does not give details on how close to the border households needed to be.

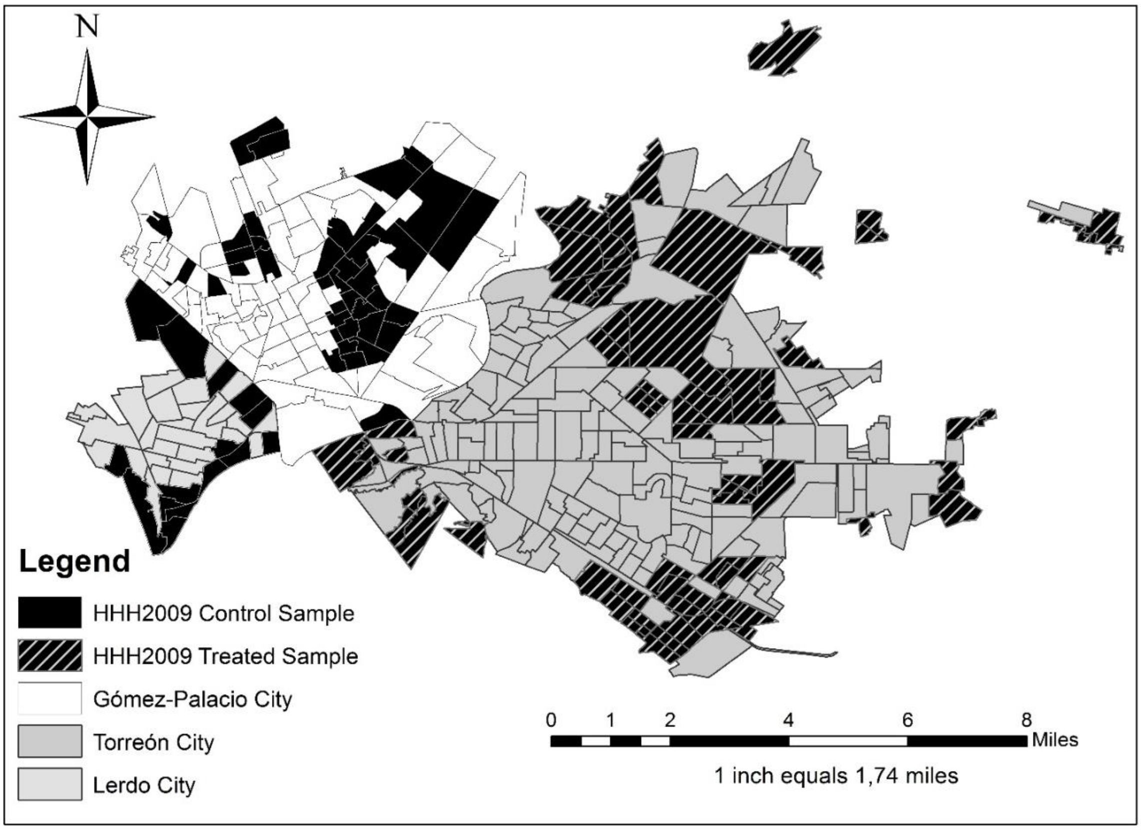
[3] To construct the L-infinite distance variable, the authors calculated the absolute value of differences for each of the four variables and kept the highest absolute distance of the four.

[4] We use the same structure for census block ID as the one in HHH2009; for example, a census block with code 1569 in a municipality with code 12 has an ID code of 120001569 (municipality code * 10,000,000 + code for census block). This ID variable is named clusterid in HHH2009 and in our coding.

2000 census and it is not clear in the original study why these blocks were included in the analysis since they cannot be matched with any control block by its characteristics.[5]

Therefore, although the replication of main results uses the 84 treatment blocks, because census information is not used in any of the study's main results, our sample replication analysis is on the 73 treated census blocks. Figure 3.1 maps the census blocks sampled by HHH2009 in the treatment area (Coahuila)[6] and the control area (Durango). Although it is clear that the census blocks lie within the twin-cities area – Gómez Palacio and Lerdo – there is no evidence that the sample was restricted to areas that straddle the border of the two states.[7] Figure 3.1 does show that treated blocks were mainly on the outskirts of Torreón and no obvious (at least to us) geographic restriction in the cities of Gómez Palacio and Lerdo for the control sample. Thus, we use all 207 census blocks in the twin-city area to replicate the selection of 53 census blocks in the control group.

Figure 3.1 Sampled census blocks in the Coahuila treatment area and the Durango control area



Source: 2000 Mexican Census, National Institute of Statistics and Geography (Elaboration: Own)

[5] A possible explanation is that they were created after the census, but we cannot know for sure.
[6] We obtained geocoded data from the year 2000 Mexican census. The map includes only the 73 treated census blocks we could trace back to it.
[7] However, HHH2009 suggest that selected blocks were close to the border: 'First, we geographically restricted the sample to families residing in the twin cities of Gómez Palacio and Lerdo (control) and Torreón (treatment) that straddle the border of the States of Durango and Coahuila, respectively'; and 'Similarly, we also identified a sample of census blocks in the cities of Lerdo and Gómez Palacio that were geographically close to the border with Torreón' (HHH2009, pp.80–81).

We now construct the L-infinite distance measure and select the closest control census blocks. For each pair of potential control and treatment census blocks, we create the L-infinite distance as the maximum of absolute value difference between the four pre-treatment variables of interest. Lastly, for each treatment block, we select the closest control census block in terms of this L-infinite distance. Without duplicates, this results in 49 census blocks in the control group. We then compare this control-group sample with the one used in HHH2009.

Table 3.1 presents summary statistics for the four variables used to construct the L-infinite distance. Column (1) presents mean values for the 73 traceable census blocks in the treatment group, and columns (2) and (4) present the mean values (standard errors) for all 207 potential control census blocks and the 49 selected control census blocks, respectively. Columns 3 and 5 show the mean difference (standard error), between treatment and control groups. As can be seen, whereas the average potential control census block is different from the treatment group in three of the 4 variables used to construct the L-infinite distance, our average selected control census block is not statistically different from the treatment group in any variable.[8] This is also the case in HHH2009, as can be seen in the first four rows of Cattaneo *et al.*'s Table 2, albeit at the household level and not the census-block level.

Table 3.1: Comparison of the four 2000 census variables used to select control blocks

Variable	(1) Treatment group in HHH2009	All potential controls		Our selection for control group 1/	
		(2) Mean	(3) Difference (1) – (2)	(4) Mean	(5) Difference (1) – (4)
Percentage of blocks with at least 1 private inhabited house with dirt floors	49.134	33.523	15.611***	43.508	5.626
	(2.857)	(2.048)	(3.844)	(3.243)	(4.385)
Percentage of private inhabited houses with dirt floors	14.352	12.909	1.443	16.275	-1.923
	(1.992)	(1.392)	(2.627)	(2.690)	(3.280)
Population 0–5 years old	291.699	197.971	93.728***	238.367	53.331
	(26.038)	(11.415)	(24.655)	(30.733)	(40.550)
Number of households	450.904	339.758	111.146***	359.000	91.904
	(43.433)	(19.246)	(41.396)	(44.904)	(64.545)
Number of census blocks	73	207		49	

Source: 2000 Mexican census. 1/ Keeping each treatment census block's closest control census block, in terms of the L-infinite distance.

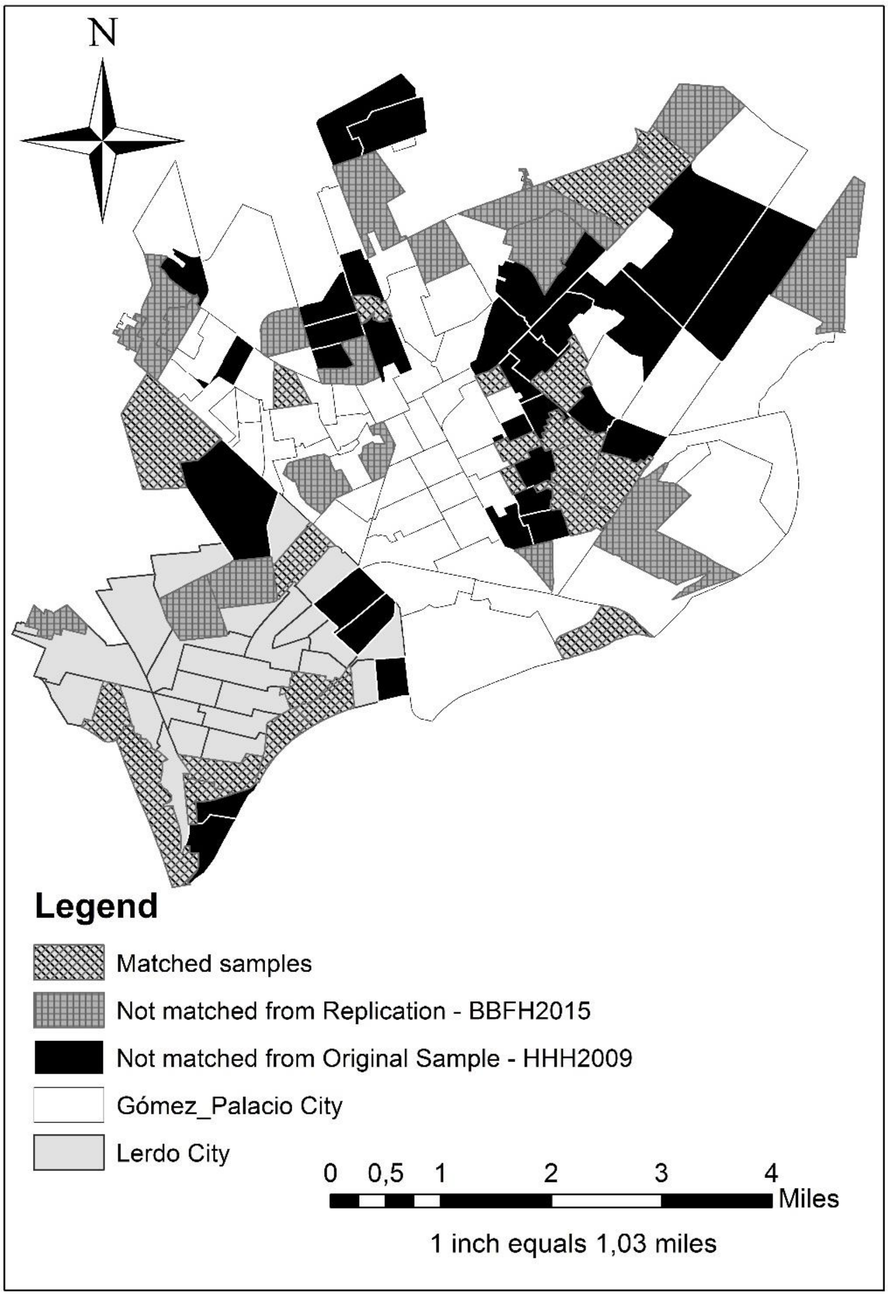
Figure 3.2 shows to what extent our sample replication for the control census blocks matches HHH2009. Of the 53 census blocks in the HHH2009 control group, we are able to match only 22 blocks using the methods described above. We are able to match 36 and 42 control census blocks when we consider each treatment group census block's two and three nearest neighbours, respectively, instead of the single nearest. However, even when considering the five nearest neighbours, only 48 of 53 control census blocks are matched; even with 10 nearest neighbours, two control census blocks in HHH2009 are still not

[8] It is difficult to draw a conclusion, because the sample size is significantly smaller for column 5 than column 3 and, as expected, the standard errors are much larger.

perfectly matched.[9] We believe the mismatch on the selected control group is partially explained by the 11 treated blocks we could not trace to the 2000 census.

In summary, we are unable to fully replicate the selected control group used in HHH2009. First, we show there is no evidence that these were selected from a pool of close-to-the-border census blocks. Second, our replication of the minimum-distance algorithm used in HHH2009 is able to replicate less than half of the actual control sample blocks. We suspect that the difference is explained by the 11 treatment blocks that we could not trace to the year 2000 Mexican census.

Figure 3.2 Control sample replication



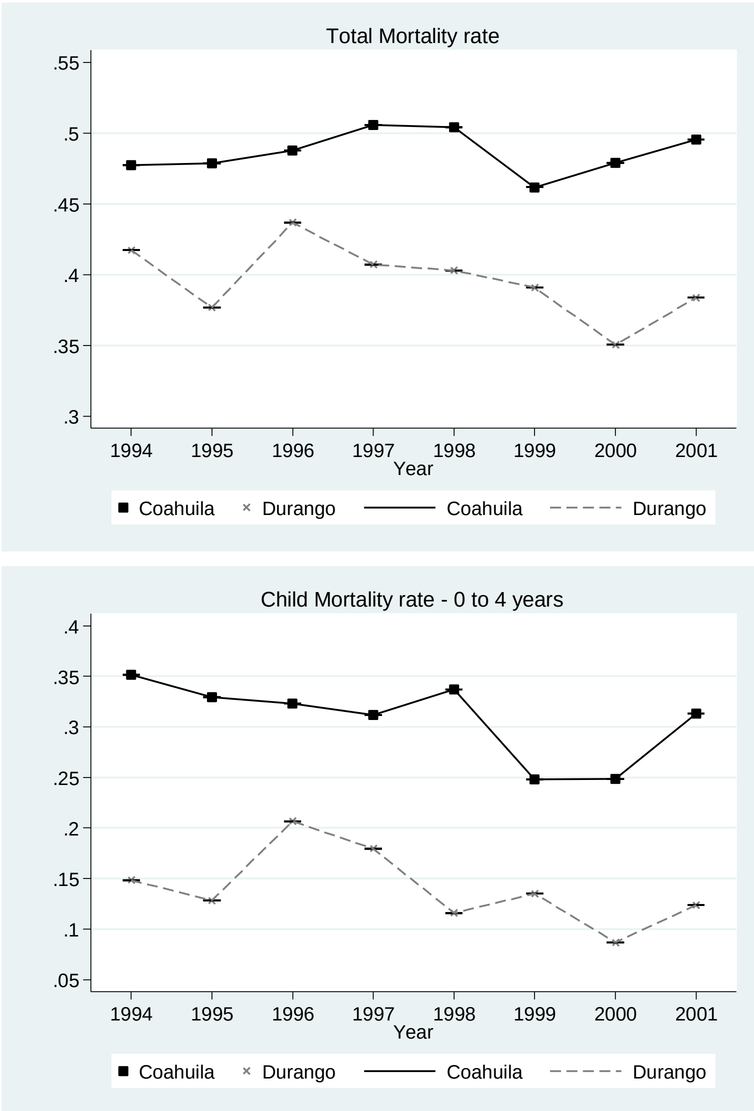
Source: 2000 Mexican Census, National Institute of Statistics and Geography (Elaboration: Own)

[9] Corresponding versions of Figure 3.2 are available upon request.

3.2 Replication of stylised facts: figures

To replicate HHH2009's Figure 1, we use publicly available information from the Mexican health secretary. We divide the number of deaths in the relevant municipalities (Torreón in Durango and the 'twin cities' of Gómez Palacio and Lerdo in Coahuila) by the projected population in the same areas and multiply the resulting number by 100. By comparing our replication Figure 3.3 with HHH2009 Figure 1, we observe very similar tendencies in both states; however, in our graph the mortality rate line for Durango has a lower mean than in HHH2009 Figure 1, so the lines do not intersect, as they do in HHH2009.

Figure 3.3 Mortality rate for Durango and Coahuila states (1994–2001)



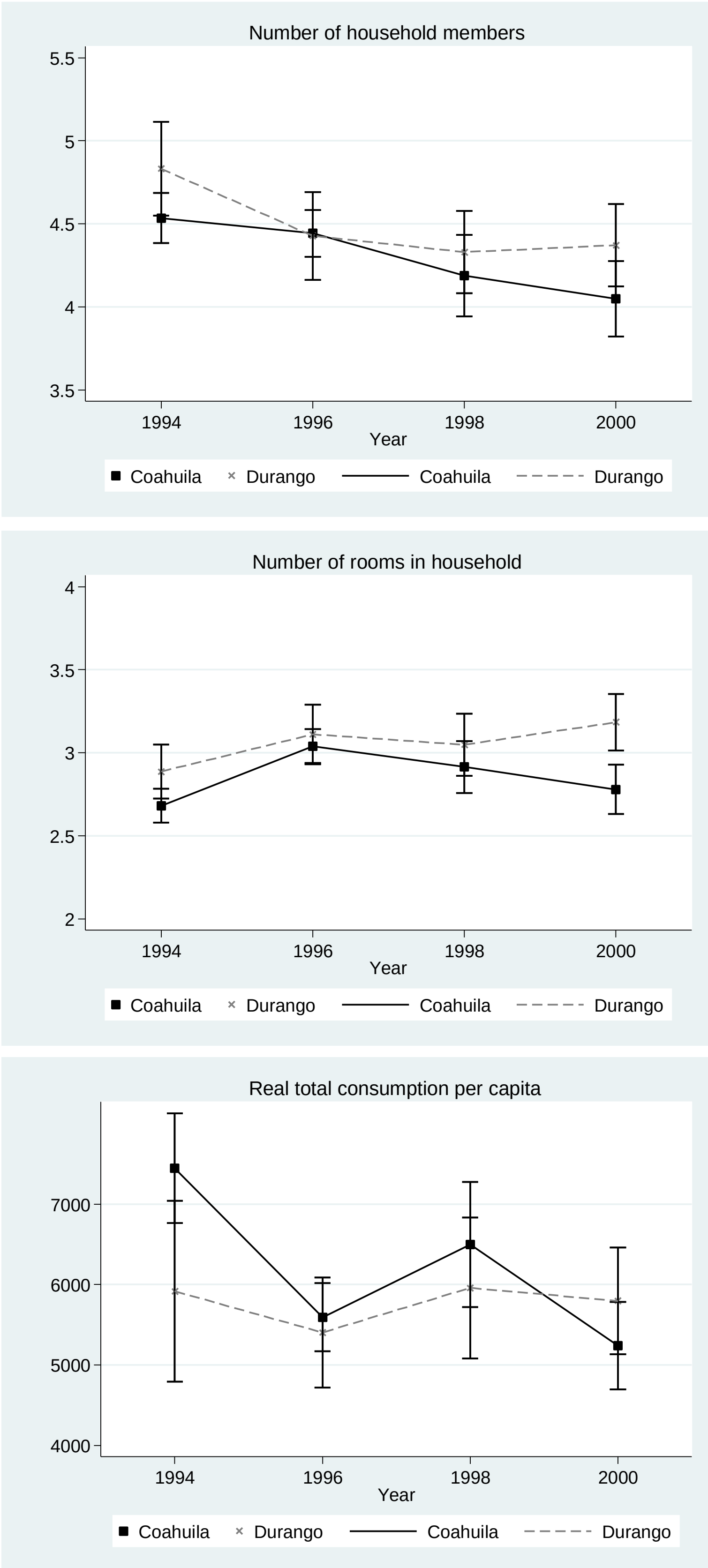
Source: ENIGH 1994-2000 (Elaboration: Own)

As to 'Child mortality rate', the main challenge we encountered was to find data for the relevant population. The Mexican health secretary has publicly available information only on deaths for children under 5 years, whereas HHH2009 include mortality rate data for children between 0 and 5 years (all inclusive). In consequence, our graph includes mortality rates only for children from 0 and 4 years (inclusive). We divide total deaths of children 0 to 4 years by the population aged 4 years or younger. Similar to the total mortality graph described in the previous paragraph, we find some minor visual differences between our graph and the authors' that do not affect the validity of the authors' results. We find that child mortality rate follows the same trend in the two states but, once again, the mean child mortality rate for Durango is lower in our graph.

To replicate the authors' Figures 2A, 2B and 3, we use data from ENIGH. We are able to replicate graphs presenting non-monetary information, and we get results similar to those in HHH2009. The replication is presented in our Figure 3.4, Figure 3.5 and Figure 3.6. The figures show that the treatment group (Coahuila) and the control group (Durango) have similar social and economic outcomes until the year 2000, when the number of households with cement floors jumps in the treatment group, while remaining about the same in Durango, the control group.

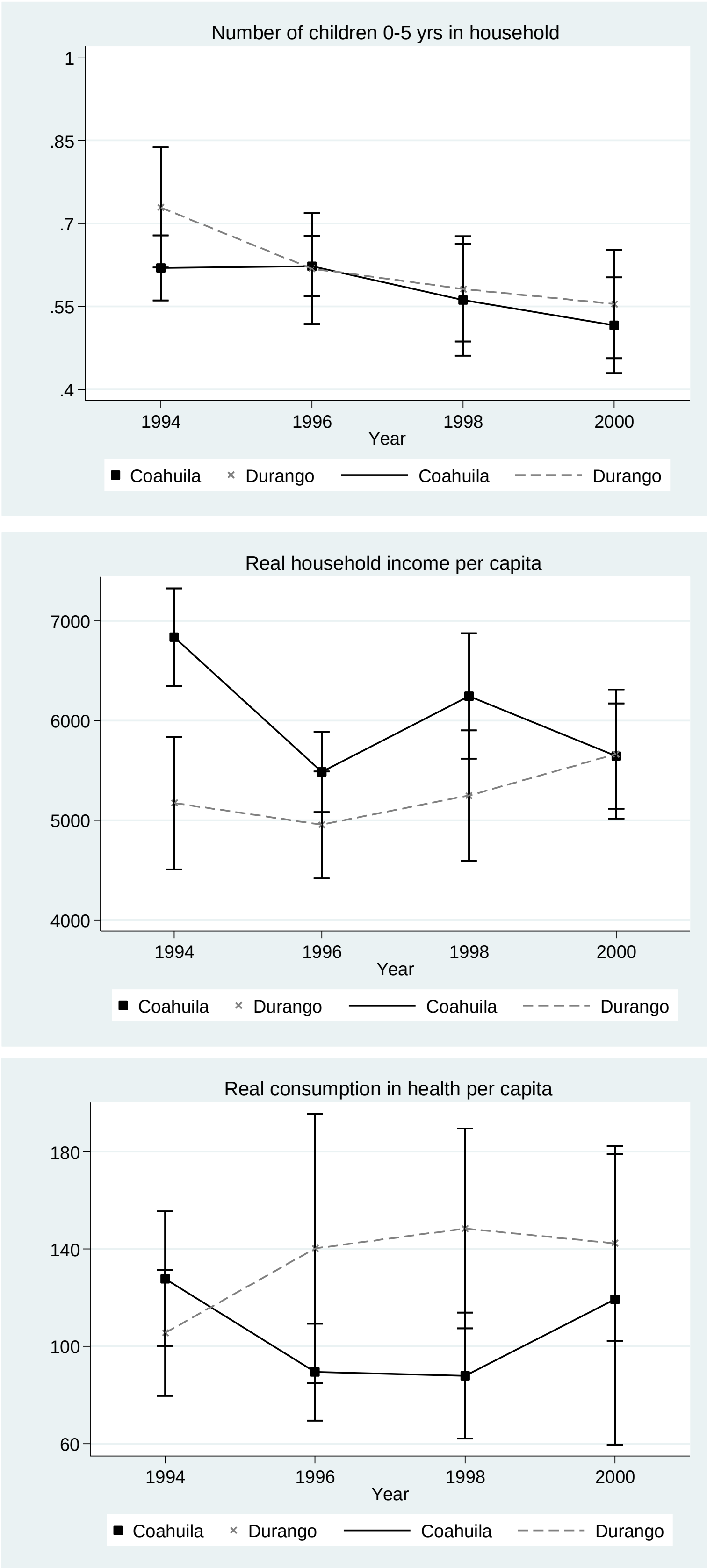
We find small differences with the original author's figures in all graphs that use real values, which we believe stems from the choice of base year for converting nominal values to real values. HHH2009 does not explicitly state which base year is used nor whether a national or a regional price index is used. The original authors kindly shared with us that they used a national deflator, based on July 2002, for each year. We still find differences in the graphs with real values. However, these differences are only in the scale; they do not threaten the validity of the parallel-trends assumptions across treatment and control groups in the data.

Figure 3.4 Number of household members, rooms and real consumption in Durango and Coahuila states (1994–2000)



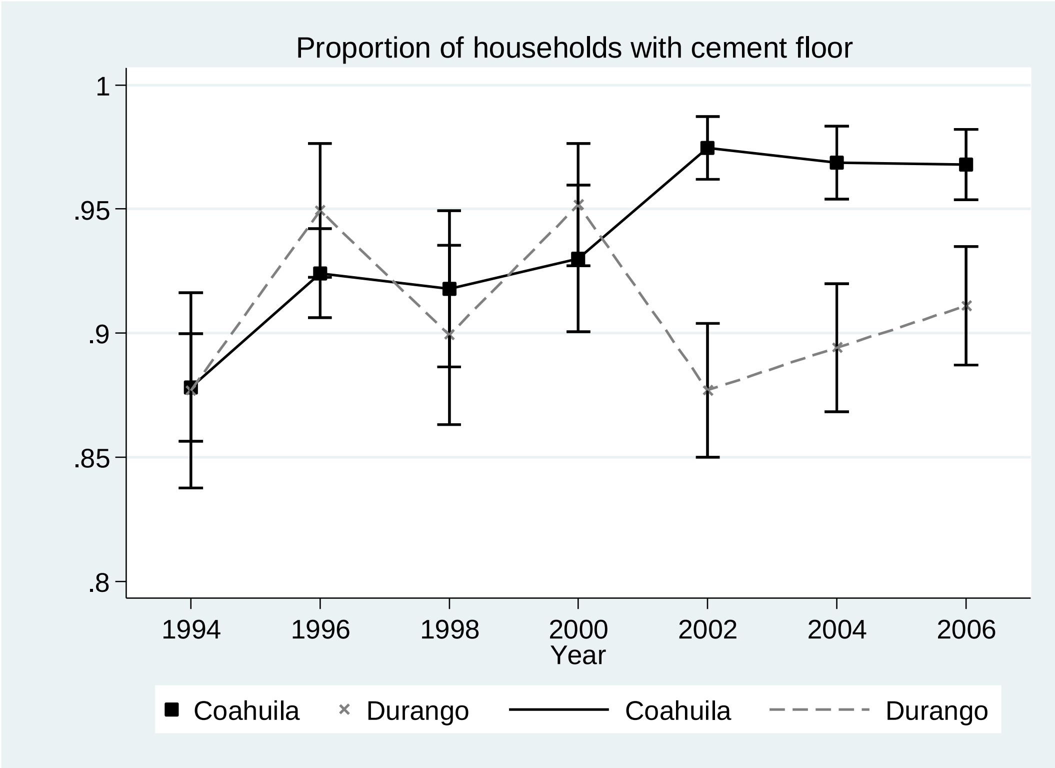
Source: ENIGH 1994–2000 (Elaboration: Own)

Figure 3.5 Number of children, real per capita income and real per capita consumption in health in Durango and Coahuila states (1994–2000)



Source: ENIGH 1994–2000 (Elaboration: Own)

Figure 3.6. Proportion of households with cement floors in Durango and Coahuila states (1994–2006)



Source: ENIGH 1994–2000 (Elaboration: Own)

3.3 Construction of variables

The codes and raw data provided by the authors allow us to replicate variable construction and to search for coding mistakes. After thorough inspection of the code used for HHH2009, we find only three mistakes, which do not alter the authors' main results.

First, the monetary value of each kind of household asset was imputed with the sample mean whenever it was missing, regardless of whether the household had reported having the asset. This artificially increased total household asset value and reduced its dispersion.[10] Second, we found a mistake concerning the imputation of microenterprise earnings. In particular, the binary variable indicating whether a household reported any microenterprise earning was imputed with the sample mean – despite being coded as 1 for 'yes' and 2 for 'no' (instead of being a 0-and-1 dummy) – and then multiplied by the household mean-imputed microenterprise earning. By contrast, we impute the sample median for the missing values on the variable indicating any microenterprise earning; naturally, multiplying this variable with the mean-imputed microenterprise earning variables gives a smaller mean. Finally, the variable recording any housing improvement since 2000, a relevant outcome variable, excluded construction and extension of rooms that were not a bathroom, which we included in our replication. This increased the share of households reporting any housing

[10] The same mistake was found for the value of any housing improvement, although this variable is not really used in HHH2009.

improvement by a small margin, from roughly 30 per cent to 34 per cent.[11] Table 3.2 shows how these coding mistakes affect the means of some variables.

Table 3.2: Imputation mistakes: comparison of summary statistics for before-correction and after-correction values

Variable/Statistics	HHH2009	Replication
A. Total value of household assets per capita		
Observations	2782	2782
Mean	22222.84	1830.528
Standard Deviation	7396.571	3087.91
Range	[5573.627; 64594.2]	[0; 58442]
B. Total household income per capita		
Observations	2780	2780
Mean	1036.676	1036.603
Standard Deviation	3538.019	3538.015
Range	[0; 127266.7]	[0; 127266.7]
C. Any house expansion (excluding installation of cement floors)		
Observations	2783	2783
Mean	.2996766	.3427955
Standard Deviation	.4581986	.4747291
Range	[0; 1]	[0; 1]

Source: Datasets used by Cattaneo *et al.* (2009).

Importantly, even after correcting for coding mistakes in the imputation procedure, the sample size for all regressions shown in the authors' Tables 4 through 7 is unchanged and remains balanced, in terms of means, between treatment and control groups (even when coefficients or standard errors do not match). Finally, after thorough inspection, we do not find any other mistake in the coding routines provided by the authors or any other major differences in our versions of HHH2009 Tables 1, 2 and 3. To show this, we present two tables mirroring the authors' Table 3. In Table 3.3, we show all variables for which results are exactly the same as HHH2009 Table 3, whereas in Table 3.4 we show variables for which results differ (income and assets).[12]

[11] See Appendix A for a more detailed description of these coding mistakes.

[12] Regarding results from HHH2009 in our Table 3.4, there is a slight difference between the mean of 'Total value of household assets per capita for the treated group', which can only be explained as a result of the rounding procedure employed by the software.

Table 3.3: Common results for Table 3 of HHH2009

Variable	Observations treatment	Mean treated	Observations control	Mean control	Mean difference
Household demographics					
Number of household members	1362	5.32 -0.0705	1393	5.374 -0.0714	-0.0539 -0.1
Age of head of household	1362	37.54 -0.413	1393	37.12 -0.49	0.418 -0.641
Head of household's years of schooling	1360	6.128 -0.134	1391	6.408 -0.115	-0.28 -0.177
Age of household spouse	1362	29.65 -0.475	1393	28.77 -0.406	0.874 -0.625
Spouse's years of schooling	1207	6.338 -0.15	1211	6.479 -0.108	-0.141 -0.185
Characteristics of children aged 0–5					
Age	1940	2.643 -0.0321	2112	2.579 -0.0323	0.0642 -0.0456
Male (=1)	1940	0.492 -0.0111	2112	0.517 -0.00738	-0.0243* -0.0133
Mother of at least one child in household present (=1)	1940	0.968 -0.00505	2112	0.964 -0.00543	0.00351 -0.00741
Mother's age (if present)	1861	27.38 -0.187	1992	27.46 -0.169	-0.0823 -0.252
Mother's years of schooling (if present)	1859	7.059 -0.135	1992	6.91 -0.133	0.149 -0.189
Father of at least one child in household present (=1)	1940	0.797 -0.0112	2112	0.763 -0.0133	0.0342* -0.0174
Father's age (if present)	1480	30.37 -0.303	1525	30.63 -0.271	-0.265 -0.407
Father's years of schooling (if present)	1476	6.839 -0.155	1519	7.153 -0.117	-0.313 -0.194
Housing characteristics					
Number of rooms	1362	2.08 -0.054	1393	1.981 -0.0531	0.0994 -0.0757
Water connection (=1)	1362	0.97 -0.00535	1393	0.977 -0.0045	-0.00713 -0.00699
Water connection inside the house (=1)	1362	0.511 -0.0286	1393	0.546 -0.022	-0.0346 -0.0361
Electricity (=1)	1362	0.985 -0.00472	1393	0.993 -0.00239	-0.00751 -0.00529
Share of rooms with cement floors in 2000	1362	0.33 -0.0202	1393	0.327 -0.0208	0.00334 -0.029
Hygienic environment					
Household has animals on land (=1)	1362	0.517 -0.0143	1393	0.48 -0.0178	0.0366 -0.0228
Animals allowed to enter the house (=1)	1362	0.192 -0.0145	1393	0.19 -0.0131	0.00211 -0.0195
Uses garbage collection service (=1)	1362	0.799 -0.0304	1393	0.845 -0.0333	-0.0461 -0.0451
Number of times respondent washed hands the day before	1362	3.754 -0.0572	1393	3.716 -0.0598	0.0383 -0.0827

Note: Model 1: no controls; Model 2: age, demographic and health-habits controls; Model 3: age, demographic, health habits and public social programmes controls. Reported results: estimated coefficient, clustered standard error at census-block level in brackets (136 clusters) and 100 × coefficient/control mean.

***Significantly different from 0 at 1 per cent level; **significantly different from 0 at 5 per cent level; *significantly different from 0 at 10 per cent level.

Source: Datasets used by Cattaneo *et al.* (2009)

Table 3.4: Different results for Table 3 of HHH2009

Variable	Observations Treatment	Mean Treated	Observations Control	Mean Control	Mean Difference
Economic characteristics [HHH2009]					
Total household income per capita	1361	1024.703	1391	1051.676	-26.973
		(71.168)		(102.976)	(125.176)
Total value of household assets per capita	1361	22393.732	1393	22032.320	361.414
		(254.334)		(308.994)	(400.204)
Economic characteristics [Replication]					
Total household income per capita	1361	1024.593	1391	1051.638	-27.045
		(71.168)		(102.979)	(125.178)
Total value of household assets per capita	1361	2000.966	1393	1664.938	336.028
		(113.167)		(75.563)	(136.076)

Note: Model 1: no controls; Model 2: age, demographic and health-habits controls; Model 3: age, demographic, health habits and public social programmes controls. Reported results: estimated coefficient, clustered standard error at census-block level in brackets (136 clusters) and 100 × coefficient/control mean.
*** Significantly different from 0 at 1 per cent level; ** significantly different from 0 at 5 per cent level; *significantly different from 0 at 10 per cent level.
Source: Datasets used by Cattaneo *et al.* (2009)

3.4 Regressions and presentation of results

The results of our replication effort for all of the regressions presented in HHH2009 can be divided into two groups: (i) what we find when trying to replicate Tables 4 through 7 in the original paper using the final datasets provided by the authors and (ii) what we find after fixing the coding mistakes described in Section 3.3. Interestingly, despite the discrepancies, we are able to fully replicate Tables 4, 5 and 6, and most results of Table 7. The three variables for which we find differences with HHH2009 are not part of Tables 4, 5 or 6.[13] All common results for Table 7 are shown in our Table 3.5, and we refer the reader to Tables 4, 5 and 6 in HHH2009 for those results. We find no differences.

[13] In HHH2009, Table 4 presented the results of the program on cement floor coverage measures, such as the share of rooms with cement floor, etc. Table 5 presented the results on children health measures, and Table 6 presented the results on happiness and maternal mental health measures.

Table 3.5: Common results for Table 7 in HHH2009.

Dependent variable	Control group mean (standard deviation)	Model 1	Model 2	Model 3
Respiratory diseases	0.355 (0.479)	0.021	0.019	0.017
		[0.019]	[0.018]	[0.019]
		5.819	5.286	4.762
Skin diseases	0.101 (0.302)	0.001	0.003	0.002
		[0.012]	[0.012]	[0.012]
		1.132	2.762	2.470
Other diseases	0.041 (0.198)	0.006	0.007	0.007
		[0.009]	[0.009]	[0.009]
		14.194	16.554	16.074
Installation of cement floor	0.530 (0.499)	0.375	0.373	0.376
		[0.028]***	[0.028]***	[0.028]***
		70.753	70.374	70.860
Construction/expansion of sanitation facilities	0.101 (0.302)	-0.016	-0.016	-0.015
		[0.015]	[0.015]	[0.015]
		-15.315	-16.094	-15.071
Restoration of sanitation facilities	0.045 (0.206)	-0.001	-0.001	-0.002
		[0.013]	[0.013]	[0.012]
		-2.813	-2.109	-3.811
Construction of ceiling	0.159 (0.366)	0.026	0.019	0.016
		[0.024]	[0.024]	[0.023]
		16.099	11.659	10.287
Restoration of walls	0.111 (0.314)	0.012	0.012	0.014
		[0.017]	[0.016]	[0.016]
		10.830	10.802	12.953
Log of self-reported sale value of house	10.491 (1.168)	-0.044	-0.015	-0.014
		[0.100]	[0.081]	[0.078]
		-0.418	-0.147	-0.132
Log total income of mothers of children 0–5 years	7.791 (0.665)	-0.037	-0.034	-0.029
		[0.064]	[0.065]	[0.066]
		-0.480	-0.436	-0.374
Log total income of fathers of children 0–5 years	8.121 (0.592)	-0.016	-0.005	0.001
		[0.028]	[0.027]	[0.026]
		-0.194	-0.064	0.016

Note: Model 1: no controls; Model 2: age, demographic and health-habits controls; Model 3: age, demographic, health habits and public social programmes controls. Reported results: estimated coefficient, clustered standard error at census-block level in brackets (136 clusters) and 100 × coefficient/control mean.
***Significantly different from 0 at 1 per cent level; **significantly different from 0 at 5 per cent level; *significantly different from 0 at 10 per cent level.
Source: Datasets used by Cattaneo *et al.* (2009)

Using the authors' dataset for the first group of results, we find that the treatment effect on the 'Log of self-reported rental value of house' is significant at a 10 per cent level, in contrast with the result in the authors' Table 7, in which a higher p-value is implied for that treatment effect. We must stress that we find the same coefficient and standard error, as shown in the first row of Table 3.6, so we presume the lack of significance shown for this result in Table 7 in HHH2009 may have been a typo. For the second group, we find two differences in Table 7. First, after correcting the construction of the indicator variable, 'Any house expansion (excluding installation of cement floors)', we find a different treatment effect for all three models. However, there are no substantial changes in sign, significance or magnitude (relative to the control group mean). Also, somewhat puzzlingly, we find a slightly different treatment effect on 'Total household consumption per capita', albeit with the same standard errors and magnitude relative to the control group mean. In all three models, the estimator is only 0.001 larger. A close inspection of our do-files did not reveal any differences in the construction of the consumption variables, so we attribute the differences to the rounding process we use.

Table 3.6: Different results for Table 7 in HHH2009

Dependent variable	Control group mean (standard deviation)	Model 1	Model 2	Model 3
Log of self-reported rental value of house [HHH2009]	5.918 (0.740)	0.033 [0.040] 0.555	0.050 [0.032] 0.853	0.053 [0.031]* 0.899
Log of self-reported rental value of house [Replication]	5.918 (0.740)	0.033 [0.040] 0.555	0.050 [0.032] 0.853	0.053 [0.031]* 0.899
Any house expansion (excluding installation of cement floors) [HHH2009]	0.277 (0.448)	0.043 [0.031] 15.524	0.035 [0.031] 12.787	0.037 [0.030] 13.272
Any house expansion (excluding installation of cement floors) [Replication]	0.326 (0.469)	0.030 [0.033] 9.260	0.022 [0.034] 6.832	0.024 [0.032] 7.363
Total household consumption per capita [HHH2009]	753.733 (1219.488)	5.217 [44.368] 0.692	10.682 [43.686] 1.417	14.012 [43.099] 1.859
Total household consumption per capita [Replication]	753.732 (1219.488)	5.218 [44.368] 0.692	10.683 [43.686] 1.417	14.013 [43.099] 1.859

Note: Model 1: no controls; Model 2: age, demographic and health-habits controls; Model 3: age, demographic, health habits and public social programmes controls. Reported results: estimated coefficient, clustered standard error at census-block level in brackets (136 clusters) and 100 × coefficient/control mean.
***Significantly different from 0 at 1 per cent level; **significantly different from 0 at 5 per cent level; *significantly different from 0 at 10 per cent level.
Source: Datasets used by Cattaneo *et al.* (2009)

In sum, despite having fixed the creation of three important variables in terms of household's health and socioeconomic outcomes, we find only minor changes and only, as expected, in the three regressions specifically related to the modified variables. This is because HHH2009 only shows the results for Models 1 through 3, whereas results for Model 4 (only available in the authors' do-files), where income and assets are used as controls, is not part of the published paper.

4. Measurement and estimation analysis

In Section 3 we show that, despite minor coding mistakes, results in HHH2009 can be replicated using the data provided by the authors. In this section, we extend the analysis to examine the robustness of the study results to an alternative method for missing-values imputation and an alternative estimation strategy.

We pursue this task by tackling four issues. First, we extend stylised facts presented in Figures 1, 2A, 2B and 3 of HHH2009 to check whether the parallel-trends hypothesis (visually) holds for additional relevant variables. Second, we use multiple imputations to assess the sensitivity of HHH2009 results to the treatment of missing values. Third, we check whether the effects of *Piso Firme* on maternal mental health outcomes hold when using an ordered multinomial model. Fourth, we use the *Piso Firme* intervention as an instrument to assess the health effects of in-house cement flooring.

For ease of presentation, we focus on estimated effects reported in Tables 5 and 6 of HHH2009, as those tables contain the intervention's most important health impacts. Results analogous to other tables in HHH2009 are available upon request.

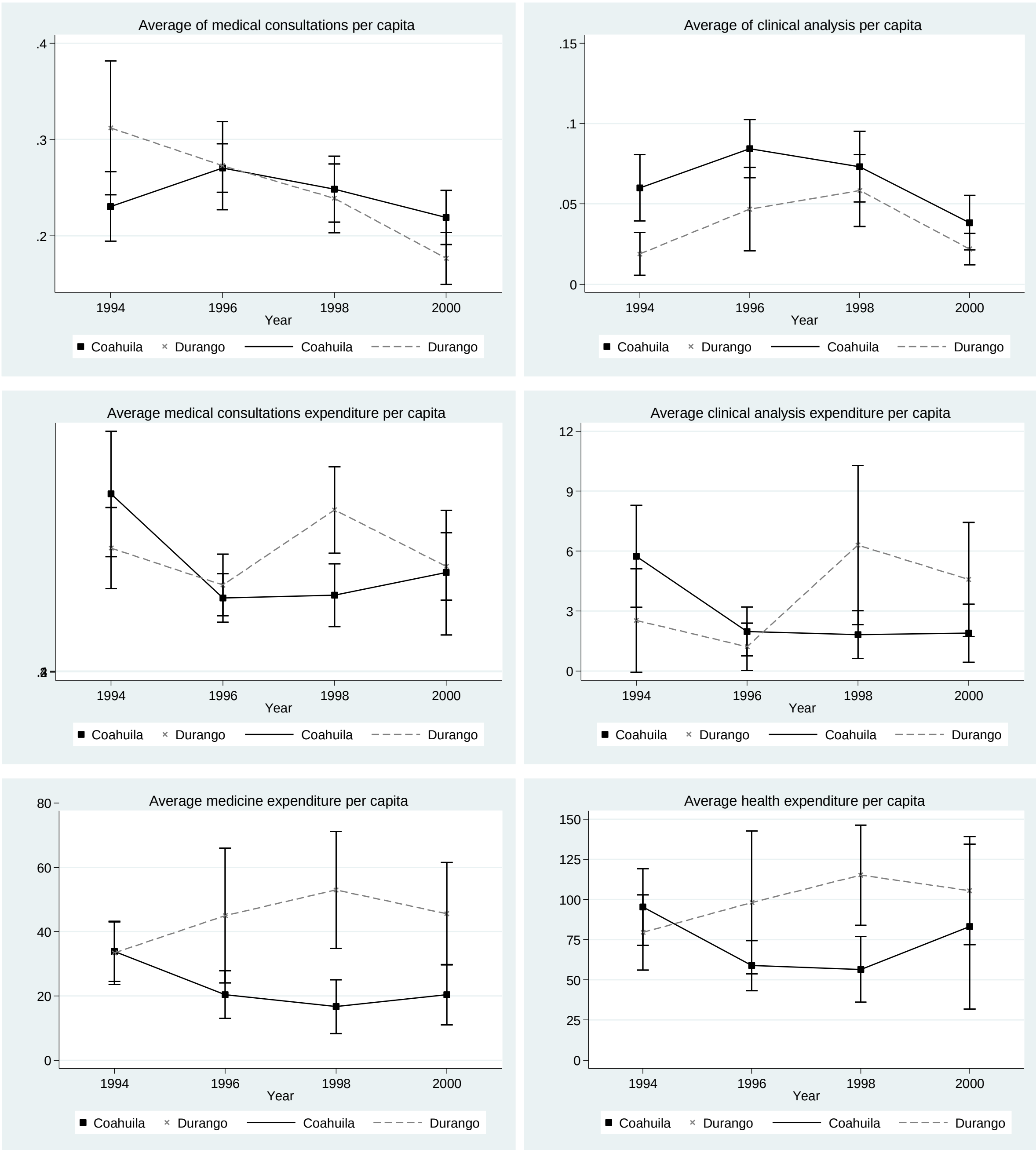
4.1 Further stylised facts using data from ENIGH

As pointed out in HHH2009 and replicated in Section 3.2 above, the control and treatment groups were part of the same urban area and, at the time of the intervention, were different only in that each group was affected by different state-level policies — Coahuila's for the treatment group and Durango's for the control. Further, HHH2009 uses state-level mean statistics to show that before the *Piso Firme* intervention in Coahuila, household characteristics in Coahuila and Durango had similar trends over time. This supports the parallel-trends hypothesis, which cannot be tested exclusively for treatment and control groups in a specific urban area due to the official ENIGH sampling frame.

We extend this analysis in two ways. First, our Figure 4.1 visually supports the authors' statement about treatment and control groups having parallel trends in terms of health-seeking behaviour before *Piso Firme*. This further strengthens the case for the identification strategy of *Piso Firme*'s effects. In other words, our analysis supports HHH2009 findings by expanding the graphical analysis to health-related variables, which are important outcomes in the study.

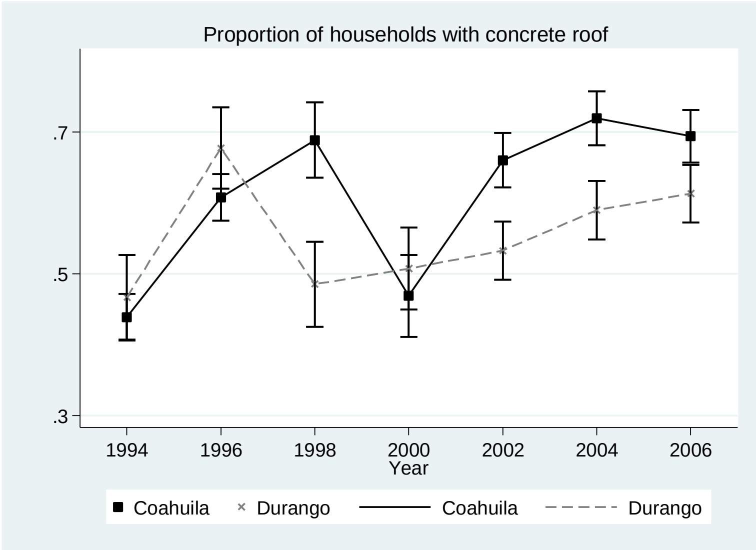
Secondly, in order to provide further evidence that *Piso Firme* affected households only through its direct impact on in-house cement flooring, we show in Figure 4.2 that Coahuila (treatment) and Durango (control) show similar regional evolution in two variables describing important housing conditions: 'Proportion of households with concrete roof[s]' and 'Proportion of households with concrete walls'. We consider this expansion to be relevant, given Cattaneo *et al.*'s remark that there is an important difference in the trends between Coahuila and Durango after the *Piso Firme* intervention. By analysing other variables related to housing quality, we can shed more lights on the potential impact of the programme.

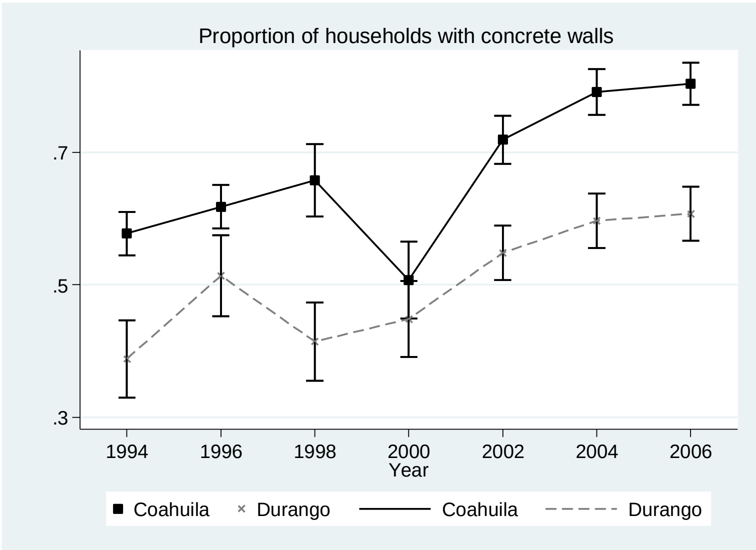
Figure 4.1 Additional health-related outcome variables in HHH2009



Source: ENIGH 1994–2000 (Elaboration: Own)

Figure 4.2 Additional outcome variables related to housing quality in HHH2009





Source: ENIGH 1994–2000 (Elaboration: Own)

However, we advise careful interpretation of all figures presented in this section. Although data from ENIGH was thoroughly prepared, the figures and the patterns they depict rely on the ENIGH sampling strategy, which changed in the late 1990s and early 2000s in ways that may partially explain the observed patterns. In particular, in 2002 the survey's sampling frame changed to include more rural and 'marginalized' households (Damián 2007). Since the control group's location, Durango, has long been more rural,[14] this change is likely to alter average characteristics and make Durango look 'poorer' in some dimensions, such as those presented in the figures above, and particularly cement flooring.

[14] The share of rural households in Durango was 40.82 per cent in 1994 and 48.98 per cent in 2000; corresponding numbers for Coahuila were 25.48 per cent and 11.08 per cent. Following INEGI's current definition of rurality, we define as rural households in 'localities' (*localidades*) with fewer than 15,000 inhabitants. Using the contemporary definition of rurality, in which a household is considered rural if it is located in a *localidad* with fewer than 2,500 inhabitants, Durango is still much more rural (approximately 30 per cent) than Coahuila (approximately 10 per cent).

This may explain why, in Figure 3.6 (available in the previous section), Durango's share of cement-floored households not only lagged behind that of Coahuila after the start of *Piso Firme*'s intervention, but even fell. Nevertheless, the major change in survey sampling occurred after *Piso Firme*'s start date – after the year 2000. Thus, the visual inspection of parallel trends is still informative and it can be seen that there is no remarkable difference in the proportion of houses with cement roofs and walls, compared with what Figure 3.6 shows about cement floors. Therefore, we find evidence that shows that the impacts reported in HHH2009 are more likely attributed to the programme and not to other non-observables.

4.2 Sensitivity of results to imputation of missing values

A common problem in social science research is that primary information is not uniformly obtained, which often leads to imputation of data in practice. In most cases, observations containing missing information are discarded from the analysis, reducing statistical power. In addition, not including them in the analysis might generate selection problems. Similarly, the use of imputation procedures has strengths and shortcomings which are important to take into consideration within the general analysis. In this sense, HHH2009 uses an imputation strategy known as 'dummy variable adjustment' or 'missing indicator' imputation. This method has been criticised because it can introduce bias in estimates and tends to increase the standard deviations of regression parameters, even in a 'missing completely at random' (MCAR) scenario (Jones 1996; Enders 2010).[15]

Puma (2009) shows that biases are important for the dummy variable adjustment imputation method when there is imbalance on the imputed variable between treated and control groups. Such a situation is not likely in this scenario, since there is not *a priori* a reason to assume that missingness is related to the treatment. Puma also claims that biases tend to be irrelevant in the context of experiments; one can therefore argue that this is a minor problem for correctly identified quasi-experiments. Because dummy variable adjustment imputation tends to overestimate estimators' standard errors, results reported in HHH2009 can be considered conservative.

Imputations used in HHH2009 can be classified in two ways: those the authors explicitly declare and those they do not (see Section 3.3). In the following analysis, we use both types of imputation and check whether different imputation methods would yield notably different results. HHH2009 states[16] that missing values in the covariates of the regression models are dealt with using a dummy adjustment imputation method (Cattaneo *et al.* 2009 pp. 96–98; 100). In particular, for any variable $\boldsymbol{x}$ the authors impute missing values with a zero and mark the observation with a specific missing-indicator dummy variable. Table 1 in Appendix A presents variables for which the authors used dummy variable adjustment imputation, the number of observations and the number of missing values. Information is presented separately for subsamples of interest – namely, children 5 years or younger and fathers and mothers with positive monetary income.[17]

[15] See Appendix B for more information about imputation methods and MCAR scenarios.
[16] See the notes below Tables 4 through 7 in HHH2009, pages 96–98 and 100 where is stated: "Missing values in covariates were imputed with zero, and a corresponding dummy variable was then added to the regressions."
[17] Although the authors state that they use logs of total income of mothers and fathers, what they actually use is adults who earn income, whether or not they are parents of children in the other sample.

As for undeclared imputation methods, after thorough inspection of the coding used for HHH2009, we detect three variables that are imputed in the household-level dataset: (i) per capita cash transfers from government programmes, (ii) total per capita value of household assets and (iii) total per capita consumption. Construction of these variables involved the sum of other intermediate variables (transfers from federal, state or municipal governments), which were mean-imputed when missing. This method, known in the literature as 'arithmetic mean imputation', introduces a bias in estimates and overestimates the estimates' variance, even in an MCAR scenario. In our replication, if any intermediate variable is missing, we do not impute it; the same applies for any variable whose construction depends on the missing intermediate variable.

In the remainder of this section, we use multiple imputation instead of the authors' dummy variable adjustment imputation and arithmetic mean imputation methods. As explained in Appendix B, multiple imputation is related to stochastic regression imputation, in that any imputed value comes from a prediction model; nevertheless, its main difference is that by using different samples for imputation, it is possible to use the error from the imputation estimation itself in the model estimation. By doing that, multiple imputation helps avoid overestimating standard errors. Multiple imputation's greatest strength is that it allows for correction of the estimates' standard errors; therefore, we expect the major effect of using multiple imputation analysis to be the re-estimation of standard errors, possibly resulting in changes in significance.

Table 4.1 compares the summary statistics for the mean-imputed variables with the same variables without imputation. We do not find major differences in summary statistics, substantially reducing concern about the undeclared imputations. Although there is no major difference between statistics for the first two variables, the third variable, 'Total value of household assets per capita', shows important drops in mean, standard deviation and maximum. The statistics changed because the imputation method affected the variable distribution.

Table 4.1: Undeclared missing values in HHH2009

	N	Mean	Standard deviation	Minimum	Maximum
A. Variable with mean-imputed values					
Cash transfers per capita from government programmes	2781	14.45	39.18	0	950
Total consumption per capita	2779	751	1148	0	35460
Total value of household assets per capita	2782	1766	3033	0	58442
B. Variables without imputation					
Cash transfers per capita from government programmes	2770	14.43	39.13	0	950
Total consumption per capita	2705	749	1128	0	35460
Total value of household assets per capita	2474	1666	2676	0	37764

Source: Datasets used by Cattaneo *et al.* (2009).

Multiple imputation consists of three main stages. The imputation phase requires defining the number of imputations to be iteratively used later on, the standard being 20. This phase also involves defining the predictive model for variables being imputed. In our case, we used a predictive mean-matching method, which imputes values by using a regression as the predictive model and then choosing a random observation whose observed values are closest to those predicted by the model.[18] Therefore, predictive mean-matching is a combination of a parametrical method (regression) and a non-parametrical method (nearest-neighbour matching). Selection of variables in the predictive model is very important. In this case, we have chosen all the covariates of the models proposed by HHH2009 and the census information that the authors include in their dataset.[19]

The second stage in multiple imputation is analysis, in which, for any variable being imputed, the software estimates the standard error resulting from the (usually 20) different imputations. The third stage is pooling, which consists of using the mean of the different imputations for estimation of the 'final model' – in our case, any of the reduced-form equations to estimate *Piso Firme*'s effect. In the third stage, the standard deviation of a given observation's many imputations is used to pool all observations together and correct standard errors of estimates of interest.

[18] We employ other methods to test for robustness in the results. The results remain statistically similar.

[19] At this step, there is a trade-off from including all variables from the data or including only those with more predictive power and those that are important for the study. Our alternative is some point in the middle, since we restrict ourselves to the information the authors considered important to the study.

Tables 4.2 and 4.3 present our multiple imputation results, analogous to Tables 5 and 6 in HHH2009. Multiple imputation results analogous to Tables 4 and 7 in HHH2009 are presented in Tables 2 and 3 in Appendix A. In each table, we report the same three models discussed in HHH2009. The first one considers only a few covariates, the second includes demographic characteristics and the third includes social characteristics.[20] The lack of sizable differences between the results in HHH2009 and those presented here strongly supports robustness of the original study.

Table 4.2: Multiple imputation regressions of children's health measures on programme dummy imputing control variables

Dependent variable	Control group mean (standard deviation)	Model 1	Model 2	Model 3
Parasite count	0.333 (0.673)	-0.065	-0.066	-0.065
		[0.032]**	[0.032]**	[0.032]**
		-19.545	-19.750	-19.526
Diarrhoea	0.142 (0.349)	-0.018	-0.018	-0.017
		[0.009]*	[0.009]**	[0.009]*
		-12.819	-13.007	-11.948
Anaemia	0.426 (0.495)	-0.085	-0.080	-0.082
		[0.028]***	[0.027]***	[0.027]***
		-20.059	-18.687	-19.188
McArthur Communication Development Test score	13.354 (18.952)	4.031	5.477	5.390
		[1.650]**	[1.614]***	[1.613]***
		30.182	41.014	40.358
Picture Peabody Vocabulary Test percentile score	30.656 (24.864)	2.668	3.074	2.938
		[1.689]	[1.472]**	[1.445]**
		8.702	10.028	9.584
Height-for-age z-score	-0.605 (1.104)	0.007	-0.004	-0.000
		[0.043]	[0.038]	[0.039]
		-1.161	0.613	0.072
Weight-for-height z-score	0.125 (1.133)	0.002	-0.003	-0.009
		[0.034]	[0.036]	[0.036]
		1.790	-2.684	-7.147

Note: Multiple imputation regressions are computed using survey information and replication code from the authors to replicate HHH2009 results. Missing values in covariates are imputed using multiple imputation with predictive mean matching. Model 1: no controls; Model 2: age, demographic and health-habits controls; Model 3: age, demographic, health habits and public social programmes controls. Reported results: estimated coefficient, clustered standard error at census-block level in brackets (136 clusters) and 100 × coefficient/control mean.
***Significantly different from 0 at 1 per cent level; **significantly different from 0 at 5 per cent level; *significantly different from 0 at 10 per cent level.
Source: Datasets used by Cattaneo *et al.* (2009)

[20] As the authors did (see footnote 17 in HHH2009), we estimate a fourth model including economic characteristics of the household, such as income per capita and total value of assets per capita, as control variables. Results are robust to the inclusion of these variables.

Table 4.3: Multiple imputation regressions of satisfaction and maternal health measures

Dependent variable	Control group mean (standard deviation)	Model 1	Model 2	Model 3
Satisfaction with floor quality	0.511 (0.500)	0.219	0.223	0.223
		[0.023]***	[0.024]***	[0.025]***
		42.784	43.636	43.570
Satisfaction with house quality	0.605 (0.489)	0.092	0.087	0.085
		[0.021]***	[0.021]***	[0.022]***
		15.136	14.365	14.075
Satisfaction with quality of life	0.601 (0.490)	0.112	0.110	0.111
		[0.022]***	[0.021]***	[0.022]***
		18.650	18.314	18.470
Depression Scale (CES-D Scale)	18.532 (9.402)	-2.315	-2.431	-2.373
		[0.616]***	[0.582]***	[0.578]***
		-12.493	-13.117	-12.805
Perceived Stress Scale (PSS)	16.514 (6.914)	-1.751	-1.787	-1.769
		[0.428]***	[0.404]***	[0.408]***
		-10.603	-10.823	-10.712

Note: Multiple imputation regressions are computed using survey information and replication code from the authors to replicate HHH2009 results. Missing values in covariates tr imputed using multiple imputation with predictive mean matching. Model 1: no controls; Model 2: age, demographic and health-habits controls; Model 3: age, demographic, health habits and public social programmes controls. Reported results: estimated coefficient, clustered standard error at census-block level in brackets (136 clusters) and 100 × coefficient/control mean.
***Significantly different from 0 at 1 per cent level; **significantly different from 0 at 5 per cent level; *significantly different from 0 at 10 per cent level.
Source: Datasets used by Cattaneo *et al.* (2009)

4.3 Using finer categories for questions on satisfaction with housing and life

The 2005 survey data collected by Cattaneo *et al.* includes categorical questions on mothers' satisfaction with housing characteristics and overall quality of life. The survey recorded answers in four categories, always ranging from 1 (best) to 4 (worst). The authors recode these values into two categories by combining the best two categories (1 and 2) in a single category ('satisfied') and the two worst in the complement. They present estimated programme effects in their Table 4. In order to check how sensitive the results are to that aggregation, we estimate the effect of *Piso Firme* on the original categorical variables using the following ordered probit model[21]:

Equation 1

$$Y_i^* = \beta X_i + u_i$$

Where Y* is a latent variable that can take four values (1, 2, 3 or 4), where 1 is the best and 4 is the worst. The error term (*u*) distributes standard normal and *F(.)* is the standard normal cumulative distribution function.

[21] We follow Cameron and Trivedi (2005, pp.519–520).

The marginal effects in the probabilities are obtained by

Equation 2

$$\frac{\partial \Pr[y_i = j]}{\partial x_i} = \{F'(\alpha_{j-1} - x_i'\beta) - F'(\alpha_j - x_i'\beta)\}\beta$$

Table 4.4 presents the results. Besides the effect of *Piso Firme* on the implied latent variable, there are four possible (average) marginal treatment effects. For ease of presentation, we present effects only on the two best outcomes for each variable and discuss the effect of *Piso Firme* in deeper detail. As improving housing quality should improve living standards, we expect *Piso Firme* to positively affect the probability of being satisfied with different housing characteristics, as HHH2009 shows.

Table 4.4: Marginal treatment effect on the probability of most (outcome 1) and second-most satisfaction (outcome 2) with housing characteristics

Dependent variable	Outcome 1 (most satisfaction)			Outcome 2 (second-most satisfaction)		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Satisfaction with floor quality	0.072 [0.008]***	0.073 [0.008]***	0.073 [0.008]***	0.132 [0.012]***	0.134 [0.012]***	0.133 [0.013]***
Satisfaction with house quality	0.040 [0.009]***	0.039 [0.009]***	0.039 [0.009]***	0.042 [0.009]***	0.041 [0.009]***	0.041 [0.009]***
Satisfaction with quality of life	0.035 [0.006]***	0.035 [0.006]***	0.036 [0.006]***	0.068 [0.012]***	0.068 [0.012]***	0.068 [0.012]***

Note: Model 1: no controls; Model 2: age, demographic and health-habits controls; Model 3: age, demographic, health habits and public social programmes controls. Reported results: estimated coefficient, clustered standard error at census-block level in brackets (136 clusters) and 100 × coefficient/control mean.
***Significantly different from 0 at 1 per cent level; **significantly different from 0 at 5 per cent level; *significantly different from 0 at 10 per cent level.
Source: Datasets used by Cattaneo *et al.* (2009)

Overall, the three satisfaction variables (floor quality, house quality and quality of life) are all positively and significantly affected by the *Piso Firme* treatment, to the extent that there is increased probability of reporting the most satisfaction (outcome 1) and second-most satisfaction. The extent to which these probabilities increase is variable: holding other variables constant, *Piso Firme* increases most satisfaction with floor quality by 7 percentage points, whereas second-most satisfaction increases 13 percentage points. Something similar happens for satisfaction with quality of life. Satisfaction with house quality increases uniformly (by 4 percentage points) in the two categories considered here.

Importantly, adding the marginal effect across the two outcomes in the table yields a result nearly identical to that reported in the authors' Table 6. On average across HHH2009's three models, the authors report a treatment effect of 22 percentage points for satisfaction with floor quality, 9 percentage points for satisfaction with house quality and 11 percentage points for satisfaction with quality of life. These numbers are very similar to our added-up treatment effects (and on average across the three models), 20.6 percentage points for satisfaction with floor quality, 8 percentage points for satisfaction with house quality and 10.3 percentage points for quality of life.[22] This clearly confirms the robustness of the results in HHH2009. Finally, we find that all of these results are the same when using an ordered logit instead of an ordered probit, though we do not present the ordered logit results here for brevity's sake.

4.4 Using *Piso Firme*'s treatment as instrument for the proportion of cement floor in the house

Cattaneo *et al.* use an instrumental variable framework to assess the effect of cement flooring on children's health, with *Piso Firme* treatment as an instrument for the proportion of cement-floored rooms in the house. The authors find that completely replacing dirt floors with cement would reduce the parasite count by 78 per cent, diarrhoea by 49 per cent, and anaemia by 81 per cent and improve cognitive development (as measured by the MacArthur and Peabody Picture Vocabulary Test) among young children by 36 per cent to 96 per cent. In our replication, our best estimation[23] reports point estimates that are very similar in magnitude.

Table 4.5 reports our replication results for the three models discussed in HHH2009. Notably, these results are robust to the inclusion of the share of cement-floored rooms in 2000 either in the first-stage or second-stage equation, which supports Cattaneo *et al.*'s decision in HHH2009 not to include initial share of cement-floored rooms in individual-level regressions such as those for children's health. Thus, overall these results highlight the idea that better housing improves health.

[22] Even for variables not reported here, all treatment marginal effects are positive and significant at a 99 per cent confidence level. This is interesting, as other variables reporting satisfaction with other housing features not directly related to flooring are positively affected by *Piso Firme*, which points to important complementarities in housing features. Results for other variables describing satisfaction with other housing features are very similar in magnitude, significance and robustness to those for overall house quality; treatment increases the satisfaction measures by around 10 percentage points in all models.

[23] In particular, as described by HHH2009, we use *Piso Firme*'s basic eligibility criterion (being a household in Coahuila) as an instrument for the share of cement-floored rooms in the household. We used a two-stage least-squares estimator due to its increased precision and better performance with small sample sizes. Admittedly, the bias in two-stage least-squares can be problematic, especially in the presence of weak instruments. Following Angrist and Pischke (2009 p.157), we used a limited-information maximum-likelihood estimator as a robustness check to our two-stage least-squares results, and found either virtually no differences or very small ones, suggesting that weak instruments are not a cause for concern in this instrumental variable framework.

Table 4.5: IV regressions of children's health measures on share of cement-floored rooms in household

Dependent variable	Control group mean (standard deviation)	Model 1	Model 2	Model 3
Parasite count	0.333 (0.673)	-0.309	-0.311	-0.306
		[0.146]	[0.146]**	[0.149]**
		-92.832	-93.431	-91.957
Diarrhoea	0.142 (0.349)	-0.083	-0.091	-0.084
		[0.043]	[0.041]**	[0.042]*
		-58.612	-64.313	-58.974
Anaemia	0.426 (0.495)	-0.400	-0.381	-0.388
		[0.123]	[0.116]***	[0.116]***
		-94.021	-89.461	-91.056
McArthur Communication Development Test score	13.354 (18.952)	17.933	24.713	23.760
		[6.887]	[6.761]***	[6.554]***
		134.283	185.054	177.920
Picture Peabody Vocabulary Test percentile score	30.656 (24.864)	14.775	17.324	16.364
		[9.108]	[7.829]**	[7.604]**
		48.197	56.510	53.381
Height-for-age z-score	-0.605 (1.104)	0.032	-0.008	0.009
		[0.195]	[0.175]	[0.176]
		-5.326	1.296	-1.483
Weight-for-height z-score	0.125 (1.133)	0.010	-0.024	-0.050
		[0.157]	[0.166]	[0.166]
		8.168	-19.075	-40.034

Note: Model 1: no controls; Model 2: age, demographic and health-habits controls; Model 3: age, demographic, health habits and public social programmes controls. Reported results: estimated coefficient, clustered standard error at census-block level in brackets (136 clusters) and 100 × coefficient/control mean.

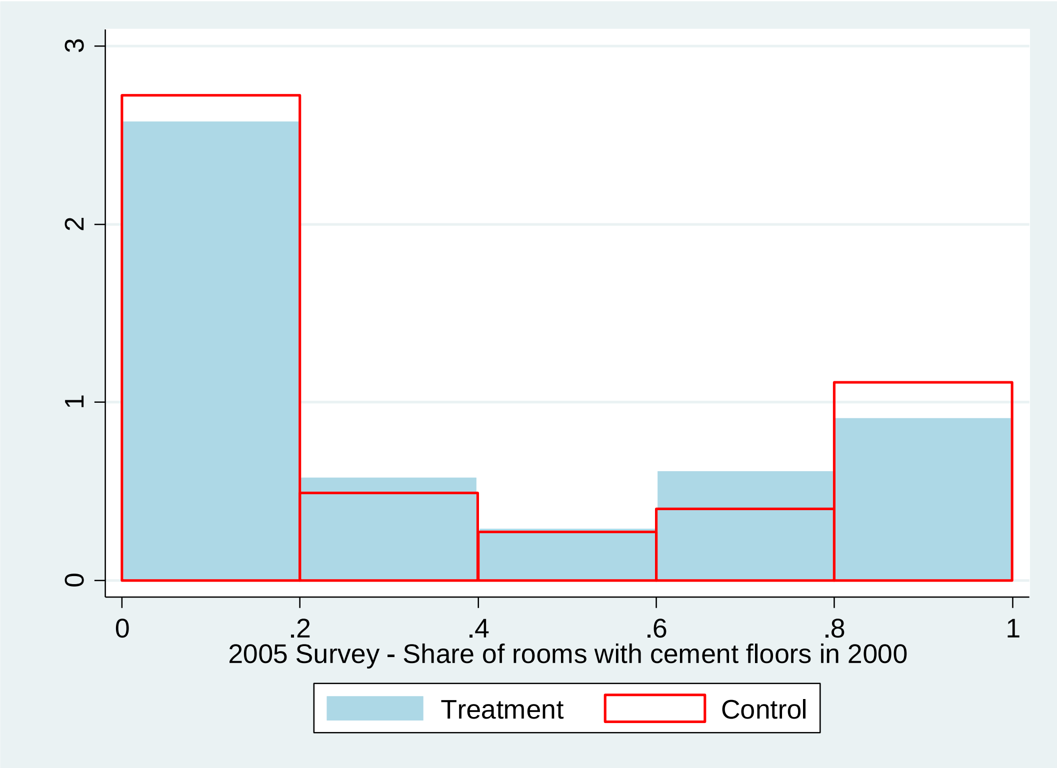
***Significantly different from 0 at 1 per cent level; **significantly different from 0 at 5 per cent level; *significantly different from 0 at 10 per cent level

Source: Datasets used by Cattaneo *et al.* (2009)

5. Theory of change analysis

Thus far, it is clear that results in HHH2009 are very robust to alternative measurement and estimation decisions, confirming that *Piso Firme* had positive effects on children's health and on maternal mental health. In this section, we explore further how in-house cement flooring affects health outcomes. We argue that different initial conditions create heterogeneous treatment effects from *Piso Firme*'s intervention. Thus, we attempt to answer whether *Piso Firme*'s impact on health outcomes is smaller for households which were initially better off. Figure 5.1 shows that there is baseline heterogeneity, balanced across treatment and control groups, in the sample in terms of the proportion of rooms with cement flooring.

Figure 5.1 Distribution of share of rooms with cement floor in 2000, treatment and control of the sample in HHH2009



Source: Datasets used by Cattaneo *et al.* (Elaboration: Own)

5.1 Exploring heterogeneity in the programme's effect

Piso Firme's treatment was assigned conditionally on the household having dirt floors in the year 2000, and households received up to 50 square metres of cement. Therefore, based on the initial share of cement flooring and the size of the house,[24] households received different levels of the treatment. Some households might have gone from having 0 to 100 per cent cement-floored rooms, and others might have gotten less – presumably those that had a larger proportion of cement floors at the baseline, year 2000 (i.e. less than 50 square metres of dirt flooring). Furthermore, if households had already cement-floored their more important rooms, then treatment may be considered effectively different.[25]

The literature on housing interventions offers scant evidence (or arguments) regarding heterogeneity in impacts, but it suggests that those initially worse-off do benefit more from interventions. In a study about slum upgrading in Brazil, Soares and Soares (2005) find that households in the national top quartile (but living in slums benefitted by the programme) benefitted to a lesser extent than those in the bottom quartile, especially in terms of access to sewerage, water and rubbish collection. Devoto *et al.* (2014) explore heterogeneous effects of improved water connection as part of their main specification for studying the

[24] Unfortunately, we do not have measurements for house sizes in square metres; we have information only on the number of rooms in the houses.
[25] Indeed, Table 4 in HHH2009 shows that *Piso Firme* had lower effects on bathroom flooring (13.1 per cent more households had cement floors in the bathroom in the treatment group than in the control group) than in dining room flooring (29.6 per cent), bedrooms (35.6 per cent) or kitchens (37.9 per cent).

impacts of improved water connections in poor households in Morocco. In particular, they interact the treatment dummy with a dummy indicating the household had an initial illegal connection to piped water – in other words, initially better off.

In HHH2009, only a few obvious interactions are found to be statistically different from zero across different specifications; initially better-off households do not free up time by an improved water connection, nor do they see an increase in the quantity of water they can use (Tables 2, 3 and 5). Conversely, we expect initially better-off households to receive fewer benefits from *Piso Firme*'s intervention.[26] We pursue these ideas in the heterogeneity analysis that follows. Table 5.1 summarises treatment and control means in terms of proportion and number of rooms with cement floors in the year 2000 (at baseline).

Table 5.1 Share of rooms with cement floors and number of rooms in 2000 for treated and control groups

Group	Share of rooms with cement floor, 2000	Number of rooms in house, 2000
Control	0.327	1.981
	(0.010)	(0.028)
Treatment	0.329	2.081
	(0.010)	(0.030)
All	0.328	2.031
	(0.007)	(0.020)

Source: Datasets used by Cattaneo *et al.* (2009)
Elaboration: Own

We construct a dummy to identify households above the median[27] share of cement-floored rooms in the house in the baseline year (2000) and interact it with the treatment dummy and check for heterogeneous treatment effects in Tables 5 and 6 of HHH2009. In Table 5.2, we report estimates only for the coefficient on this interaction to explore heterogeneity in results of Table 5 of HHH2009. For all children's health outcomes, we include the full set of covariates in our regressions (we estimate Model 3 in HHH2009). Table 5.2 presents baseline Model 3 estimates, *i.e.* without controlling for the heterogeneity interaction. The next three columns present extensions to Model 3[28]: column (1) includes only the heterogeneity interaction; column (2) also includes the dummy for households above the median 2000 share of cement-floored rooms; and column (3) instead includes the 2000 share itself. Table 5.3 explores heterogeneous effects in Table 6 of HHH2009 in a similar fashion.

As expected, results suggest that households initially better-off benefitted less from *Piso Firme*'s intervention. First, Table 5.2 shows that beneficial health effects on children from better off households were significantly lower for anaemia and the height-for-age z-score. For other measures of children's health, point estimates show smaller effects for the initially better off, but the differences are not statistically different. Second, Table 5.3 shows that the

[26] Also, if worse-off households tend to have less cement-flooring and may benefit marginally more from its increase, then it is likely that the estimated average treatment effect on the treated is larger than what a uniform treatment (e.g. a 10 per cent increase in cement-flooring in the house) would have on a treatable population.
[27] Choosing the median as a cut-off point helps balance the sample to check for heterogeneity.
[28] Results are robust to the consideration of Model 1 and Model 2; these results are available upon request.

programme also had smaller impacts on maternal mental health measures among initially better-off households. Using a back-of-the-envelope calculation, the effect on the initially better off was approximately half[29] of what HHH2009 reports for both satisfaction with floor quality and the Perceived Stress Scale. Using the same calculation, *Piso Firme*'s effects on satisfaction with house quality and quality of life, as reported in HHH2009, is almost cancelled out by the heterogeneity interaction. Finally, there are no significant differences for the initially better off in terms of the programme's effect on the Depression Scale.

Hence, results support the idea that *Piso Firme*'s effect was smaller for children and mothers from initially better-off households, particularly in terms of cement-floor coverage at baseline. Those households received less cement as treatment and might also have used it to cement floors less important to the household.[30]

Table 5.2 Regressions of children's health measures on programme dummy and interactions

Dependent variable	Control group mean (standard deviation)	Programme effect in baseline Model 3	Interaction of dummy for 50% better-off households and programme dummy in Model 3		
			(1)	(2)	(3)
Parasite count	0.332 (0.670)	-0.065	0.026	0.059	0.048
		[0.032]**	[0.032]	[0.042]	[0.040]
		-19.526	7.722	17.881	14.350
Diarrhoea	0.141 (0.348)	-0.017	0.005	0.036	0.034
		[0.009]*	[0.019]	[0.024]	[0.021]
		-11.948	3.778	25.677	23.963
Anaemia	0.425 (0.495)	-0.082	-0.008	0.080	0.055
		[0.027]***	[0.023]	[0.031]**	[0.028]*
		-19.188	-1.810	18.878	12.844
McArthur Communication Development Test score	13.376 (19.057)	5.390	-2.344	-3.176	-2.886
		[1.613]***	[2.386]	[3.062]	[2.676]
		40.358	-17.556	-23.784	-21.614
Picture Peabody Vocabulary Test percentile score	30.622 (24.853)	2.938	3.422	-0.770	-0.413
		[1.445]**	[2.028]	[2.483]	[2.314]
		9.584	11.162	-2.513	-1.347
Height-for-age z-score	-0.604 (1.094)	-0.000	0.083	-0.134	-0.101
		[0.039]	[0.046]	[0.070]*	[0.059]*
		0.072	-13.666	22.172	16.684
Weight-for-height z-score	0.125 (1.133)	-0.009	0.008	-0.019	-0.024
		[0.036]	[0.046]	[0.075]	[0.063]
		-7.147	6.358	-14.992	-19.417

Note: Model 3: age, demographic, health habits and public social programmes controls. Reported results: estimated coefficient of the differential effect on initially better-off households, clustered standard error at census-block level in brackets (136 clusters) and 100 × coefficient/control mean. 'Programme effect in baseline Model 3' reports results for Model 3 in HHH2009 Table 5. Column (1) includes the interaction between the intent-to-treat dummy and a dummy for households above the median year-2000 share of cement-floored rooms. Column (2) includes the latter dummy. Column (3) replaces the dummy with the household's year-2000 share of cement-floored rooms.
***Significantly different from 0 at 1 per cent level; **significantly different from 0 at 5 per cent level; *significantly different from 0 at 10 per cent level
Source: Datasets used by Cattaneo *et al.* (2009)

[29] We divide the point estimate in column (3) by the one in the column showing the baseline result for Model 3. Alternatively, use the proportional effect that results from dividing the point estimate by the control group's mean of the dependent variable.
[30] Households most likely decide to replace dirt floors with cement in terms of their room preference.

Table 5.3 Regressions of satisfaction and maternal mental health measures on programme dummy and interactions

Dependent variable	Control group mean (standard deviation)	Programme effect in baseline Model 3	Interaction of dummy for 50% better-off households and programme dummy in Model 3		
			(1)	(2)	(3)
Satisfaction with floor quality	0.511 (0.500)	0.223 [0.025]*** 43.570	-0.117 [0.032]*** -22.831	-0.116 [0.032]*** -22.604	-0.115 [0.032]*** -22.544
Satisfaction with house quality	0.605 (0.489)	0.085 [0.022]*** 14.075	-0.072 [0.034]** -11.867	-0.067 [0.034]* -11.131	-0.067 [0.034]* -11.096
Satisfaction with quality of life	0.601 (0.490)	0.111 [0.022]*** 18.470	-0.125 [0.038]*** -20.725	-0.127 [0.038]*** -21.215	-0.127 [0.038]*** -21.203
Depression Scale (CES-D Scale)	18.532 (9.402)	-2.373 [0.578]*** -12.805	0.767 [0.681] 4.141	0.986 [0.662] 5.318	0.968 [0.661] 5.225
Perceived Stress Scale (PSS)	16.514 (6.914)	-1.769 [0.408]*** -10.712	0.866 [0.492]* 5.246	0.974 [0.488]** 5.897	0.966 [0.488]** 5.850

Note: Model 3: age, demographic, health habits and public social programmes controls. Reported results: estimated coefficient of the differential effect on initially better-off households, clustered standard error at census-block level in brackets (136 clusters) and 100 × coefficient/control mean. 'Program effect in baseline Model 3' reports results for Model 3 in Table 5 of HHH2009. Column (1) includes the interaction between the intent-to-treat dummy and a dummy for households above the median year-2000 share of cement-floored rooms. Column (2) includes the latter dummy. Column (3) replaces the dummy with the household's year-2000 share of cement-floored rooms.
***Significantly different from 0 at 1 per cent level; **significantly different from 0 at 5 per cent level; *significantly different from 0 at 10 per cent level
Source: Datasets used by Cattaneo *et al.* (2009)

6. Conclusions and remarks

In our pure replication, we do not find any major discrepancies with the results reported in HHH2009. Importantly, the minor coding issues we find are not part of the most important variables used for the original study's results. An exception is the sampling replication, in which we are unable to fully replicate the final sample of control census blocks. However, it is likely that this result stems from a geographic filter used by the authors that we could not access and the presence of 11 treated census blocks that we could not trace to the Mexican census for the year 2000. Thus, the sampling replication deserves more discussion.

The rest of our work seeks to check the robustness of HHH2009 to different specifications and imputation methods. On one hand, we systematically find results in HHH2009 to be robust to different strategies for dealing with missing values, specification issues and estimation procedures. The heterogeneity analysis by baseline proportion of rooms with cement floors shows that initially better-off households benefitted less from *Piso Firme*, indicating a bigger programme impact on initially worse-off households.

We are thankful for the authors' collaboration and support throughout this replication by kindly sharing their codes, datasets and other methodological instruments and answering our emails to provide clarification.

Appendix A: Coding mistakes in HHH2009

We found a coding mistake in the do-file (approximately lines 118 and 303 of the 01 PF_assembly_households.do file) regarding imputation of some monetary variables, value of household assets (variables c324ba–c324b1) and value of any housing improvements (variables c430a2–c430j2). In both cases, the attempted mean imputation aimed to replace missing values with the mean value for all non-missing observations. For instance, in the case of asset values, if a given household reported having asset \`x' (i.e. c324a\`x'!=0) but did not report a resale value for the asset in variable c324b\`x', then the missing value should have been replaced with the mean value for that asset across all non-missing observations. Instead, due to a coding error, the mean value was imputed for all households, whether they possessed the asset or not, since the condition was mistakenly entered as c324a\`x'!=2, which all households satisfied.[31]

To show this, we present the total value of household assets per capita in panel A of Table 1, below (mirroring Table 3 in HHH2009). It is evident that the mistake artificially increased every household's wealth measure; furthermore, it reduces the variable's range and variability for both variables. Finally, although the same mistake occurred for imputation of the value of any housing improvement, HHH2009 does not present summary statistics for the housing improvement variable; thus, for the sake of brevity, we do not present its summary statistics in Table 3.2.

We found another imputation error in the construction of net income of any microenterprise in the household, which leads to a mistake in the construction of the per capita income variable. In particular, to impute the payment to out-of-household labour in the microenterprise (for any of the seven types of activity recorded in the 2005 survey), two variables needed to be imputed: a binary variable indicating whether the household had had any extra-household labour in the microenterprise for a given activity \`x' (mi009\`x') and a continuous variable showing the payment made to those labourers (mi010\`x'). Although mean imputation for the latter made sense, mean imputation for a categorical variable such as the first does not make direct sense. Further, in this case the original binary variable mi009\`x' was coded as 1 for 'yes' and 2 for 'no', and was not recoded to a 0-1 expression before mean imputation. We corrected this by recoding the original binary variable and then imputing the median.[32] Since microenterprise income is used later to construct the per capita income measure — which is in fact a variable used in Table 3 of HHH2009 to check the balance between treated and control groups — we present a comparison of its summary statistics before and after the correction we describe is shown in panel B of Table 1, below.

[31] This mistake may have resulted from a confusion in the variables' values before and after recodification: each variable **c324a\`x'** records whether a household has a given \`x' asset. Initially, that variable was stored as 1 ('yes'), 2 ('no') and other values for missing values. In the same do-file, however, before the imputation mistake is made, these variables are recoded so that 0 represents the 'no' option. Someone may have coded for the early 'no' value and not updated the coding for the recoded variable, in which there were no 2 values.

[32] At this point, once the binary variable was coded as 0 or 1, we could have imputed the mean for all missing values. However, this would have meant assuming that the mean overall payment to extra-household labourers can be broken down as the mean proportion of households that hired out-of-household labour, multiplied by the mean payment to such labour. This would implicitly require no covariance between such two variables, which is not evident by itself.

Finally, there seems to be an error in the construction of the binary variable S_improveany (originally named danyimprov), which indicates if the household made any house improvements (excluding installation of cement floors) between 2000 and 2005. This variable is used as an outcome in a regression in Table 7 of HHH2009, and its name suggests it measures any house improvements other than the installation of cement floors, though in the creation of the variable in 01 PF_assembly_households.do file, the variable S_improveany excludes construction and extension of rooms that are not bathrooms.[33] To fix this omission, we specifically included variable c430f1 in the creation of S_improveany, since it clearly qualifies as a house improvement.[34] Panel C compares summary statistics for this variable before and after the correction. As can be seen, more households reported any household expansion (34.3 per cent) than what was previously measured (30.0 per cent).

In sum, panels A and B in Table 3.2 show the only variables for which results in Table 3 of HHH2009 differ from ours; the difference we find in panel B is only in the second and third decimal places. On the other hand, panel C describes the changes in the only outcome variable for which regressions results in Table 7 of HHH2009 differ from our own.[35]

[33] The other category that is still excluded in our variable is the construction of fences or walls. We were unsure if this category was relevant to include as any house improvements.

[34] We believe this apparent mistake stems from the presentation of Table 7: four variables of house improvements other than cement floors are used as outcome variables and have their regression results presented before S_improveany. This suggests that the authors meant to use S_improveany as a catchall variable for the four house improvement variables. However, this leaves out useful information on, for instance, the construction of new rooms.

[35] There is a final group of ‘mistakes’ whose correction has no impact on any of the paper’s results and may be considered minor ‘judgement calls’, as they were left out of the paper’s four regression models. One such mistake is the exclusion of the value of school breakfasts from the construction of the number of federal programmes the household receives. To be clear, a similar variable is considered for the construction of the variable measuring the value of federal transferences. Another ‘mistake’ of this kind is the exclusion of the presence of scorpions (c4291) from the correction for missing values of the ‘pests in house’ variable (dpest). Since correcting these variables makes no difference on the results, we simply do not consider them in the imputation sensitivity analysis.

Tables regarding imputation methods used in HHH2009 and multiple imputation analysis

Table A 1 Declared imputations in HHH2009

Variables	N	N missing
A. Household-level estimation (N=2783)		
Head of household’s years of schooling	2779	4
Spouse's years of schooling	2444	339
Proportion of males 0–5 years in household	2782	1
Proportion of males 6–17 years in household	2782	1
Proportion of males 18–49 years in household	2782	1
Proportion of males 50+ years in household	2782	1
Proportion of females 0–5 years in household	2782	1
Proportion of females 6–17 years in household	2782	1
Proportion of females 18–49 years in household	2782	1
Proportion of females 50+ years in household	2782	1
Total household income per capita	2780	3
B. Individual-level estimation full sample (N=6693)		
Mother of at least one child in household present (=1)	4092	2601
Mother’s age (if present)	4252	2441
Mother’s years of schooling (if present)	4249	2444
Father of at least one child in household present (=1)	4092	2601
Father’s age (if present)	3290	3403
Father’s years of schooling (if present)	3279	3414
Total household income per capita	6686	7
Total value of household assets per capita	6690	3
C1. Individual-level estimation sub-sample 1 (N=4092)		
Mother of at least one child in household present (=1)	4092	0
Mother’s age (if present)	3890	202
Mother’s years of schooling (if present)	3888	204
Father of at least one child in household present (=1)	4092	0
Father’s age (if present)	3037	1055
Father’s years of schooling (if present)	3027	1065
Total household income per capita	4089	3
Total value of household assets per capita	4090	2
C2. Individual-level estimation sub-samples 2 and 3 (N=2601)		
Mother’s age (if present)	362	2239
Mother’s years of schooling (if present)	361	2240
Father’s age (if present)	253	2348
Father’s years of schooling (if present)	252	2349
Total household income per capita	2597	4
Total value of household assets per capita	2600	1

Source: Datasets used by Cattaneo *et al.* (2009)

Table A 2 Multiple imputation regressions of cement-floor coverage measures on programme dummy

Dependent variable	Control group mean (standard deviation)	Model 1	Model 2	Model 3
Share of rooms with cement floors	0.728 (0.363)	0.202	0.207	0.208
		[0.021]***	[0.019]***	[0.019]***
		27.746	28.431	28.620
Cement floor in kitchen	0.671 (0.470)	0.255	0.259	0.261
		[0.025]***	[0.023]***	[0.023]***
		37.936	38.564	38.935
Cement floor in dining room	0.709 (0.455)	0.210	0.215	0.217
		[0.026]***	[0.025]***	[0.025]***
		29.633	30.345	30.672
Cement floor in bathroom	0.803 (0.398)	0.105	0.115	0.118
		[0.022]***	[0.018]***	[0.018]***
		13.071	14.271	14.689
Cement floor in bedroom	0.668 (0.471)	0.238	0.245	0.244
		[0.020]***	[0.020]***	[0.019]***
		35.598	36.635	36.572

Note: Multiple imputation regressions were computed using survey information and replication code made by authors to replicate HHH2009 results. Missing values in covariates were imputed using multiple imputation with predictive mean matching. Model 1: no controls; Model 2: age, demographic and health-habits controls; Model 3: age, demographic, health habits and public social programmes controls. Reported results: estimated coefficient, clustered standard error at census-block level in brackets (136 clusters) and 100 × coefficient/control mean.
*** Significantly different from 0 at 1 per cent level; ** significantly different from 0 at 5 per cent level; * significantly different from 0 at 10 per cent level
Source: Datasets used by Cattaneo *et al.* (2009)

Table A 3 Multiple imputation regressions of robustness checks imputing control variables

Dependent variable	Control group mean (standard deviation)	Model 1	Model 2	Model 3
Respiratory diseases	0.355 (0.479)	0.021	0.020	0.018
		[0.019]	[0.018]	[0.019]
		5.819	5.597	5.084
Skin diseases	0.101 (0.302)	0.001	0.003	0.003
		[0.012]	[0.012]	[0.012]
		1.132	3.171	3.033
Other diseases	0.041 (0.198)	0.006	0.007	0.006
		[0.009]	[0.009]	[0.009]
		14.194	16.177	15.647
Installation of cement floor	0.530 (0.499)	0.375	0.372	0.375
		[0.028]***	[0.028]***	[0.028]***
		70.753	70.134	70.675
Construction/expansion of sanitation facilities	0.101 (0.302)	-0.016	-0.016	-0.014
		[0.015]	[0.015]	[0.015]
		-15.315	-15.521	-13.632
Restoration of sanitation facilities	0.045 (0.206)	-0.001	-0.000	-0.001
		[0.013]	[0.013]	[0.013]
		-2.813	-1.005	-1.892
Construction of ceiling	0.159 (0.366)	0.026	0.018	0.017
		[0.024]	[0.024]	[0.024]
		16.099	11.222	10.640
Restoration of walls	0.111 (0.314)	0.012	0.013	0.014
		[0.017]	[0.016]	[0.016]
		10.830	11.922	12.789
Any house expansion (excluding installation of cement floors)	0.277 (0.448)	0.043	0.036	0.038
		[0.031]	[0.031]	[0.030]
		15.524	13.073	13.881
Log of self-reported rental value of house	5.918 (0.740)	0.033	0.054	0.053
		[0.040]	[0.032]*	[0.032]*
		0.555	0.912	0.898
Log of self-reported sale value of house	10.491 (1.168)	-0.044	-0.015	-0.021
		[0.100]	[0.080]	[0.080]
		-0.418	-0.141	-0.202
Total household consumption per capita	740.219 (1166.562)	19.021	25.522	25.685
		[43.951]	[44.156]	[43.695]
		2.570	3.448	3.470
Log total income of mothers of children 0–5 years	7.791 (0.665)	-0.037	-0.036	-0.032
		[0.064]	[0.063]	[0.064]
		-0.480	-0.459	-0.407
Log total income of fathers of children 0–5 years	8.121 (0.592)	-0.016	-0.007	-0.001
		[0.028]	[0.026]	[0.026]
		-0.194	-0.088	-0.008

Notes: Multiple Imputation Regressions computed using survey information and replication code made by authors to replicate HHH2009 results. Missing values in covariates were imputed using Multiple Imputation with Predictive Mean Matching procedure. Model 1: no controls; Model 2: age, demographic and health-habits controls; Model 3: age, demographic, health habits and public social programmes controls. Reported results: estimated coefficient, clustered standard error at census-block level in brackets (136 clusters) and 100 × coefficient/control mean.
*** Significantly different from 0 at 1 per cent level; ** significantly different from 0 at 5 per cent level; * significantly different from 0 at 10 per cent level
Source: Datasets used by Cattaneo *et al.* (2009)

Appendix B: 'Missingness' and imputation methods in the literature

Variable imputation outcomes depend mainly on the 'missingness' pattern. For example, if we considered that missing information is completely random, in the sense that the only reason why it is missing is related to exogenous factors (not the subject of analysis), then we are in the 'missing completely at random' (MCAR) scenario. In that case, almost any type of imputation is appropriate.

Nevertheless, this scenario is unlikely to be realistic, because the reasons behind missingness are most often related to the subject of study (for example, migration and dying in panel data) and affect the result variable. Thus, not correcting the bias of this missingness pattern could lead us to biased estimates.

When we know which variables predict the missingness of the imputed variables (and if they are all observables), we are in the 'missing at random' (MAR) scenario. In this case, it is possible to estimate the results using maximum likelihood or multiple imputation methodologies.

Finally, if there are unobservable variables related to the missingness pattern and to result variables of the analysis, then we are in the 'missing not at random' (MNAR) scenario, which will lead to biased estimates even when we control for other variables that predict the missingness of the imputed variable. In this scenario, no imputation method can help solve the problem. It is preferable to use maximum likelihood and multiple imputation methodologies, since they at least control for the other covariates that predict missingness and therefore have less-biased estimates.

The most usual methods of imputation in literature are the following:

- *Listwise deletion:* All observations that contain at least one missing value for any variable are excluded from the analysis. In other words, we only work with the dataset that has no missing variables. The results represent unbiased estimates only if we are in an MCAR scenario.

- *Pairwise deletion:* The more typical method, when we have different models of analysis and for each one we exclude observations that have at least one missing variable used in that model. This leads to different samples for each estimation model. This is the default procedure in many statistical packages such as Stata®. The results represent unbiased estimates only if we are in an MCAR scenario.

- *Arithmetic mean imputation:* Imputation of the mean of a variable for each missing value. Cattaneo *et al.* use this method in the creation of variables for HHH2009. This method gives biased results in MCAR, MAR and MNAR, underestimate the standard deviation of the variable and overestimates the variance of the estimates.

- *Dummy variable adjustment imputation:* Cattaneo *et al.* employed this procedure for HHH2009. It consists of replacing missing values with zeros and creating a dummy variable for each variable that has at least one missing value indicating the imputed values. Given its practicality, the use of dummy variable adjustment imputation is common, but it also gives biased results.

- *Regression imputation:* Uses a regression model based on other covariates that can predict the missing values. Nevertheless, even when it can bring unbiased estimates in an MAR scenario, it underestimates the variance of the estimator.

- *Stochastic regression imputation:* Similar to regression imputation but also includes a normal distributed error term. This can help in correcting the standard error underestimation problem of regression imputation estimator, but it does not fully accomplish the task.

- *Hot-deck imputation:* A collection of methods that replace missing values with the values of 'similar' observations. This method recovers a distribution for each variable but has problems related to keeping the original correlations between variables. Consequently, hot-deck imputation can give biased estimates and bigger standard errors.

- *Maximum likelihood:* Employs all available information to estimate the values that are most likely to occur with the likelihood function. The results under MAR assumption are unbiased with respect to the full information scenario.

- *Multiple imputation:* In essence, similar to stochastic regression imputation, but replicates it n times. By using the n imputations with Rubin's Rules we can estimate, under MAR assumption, the results of a full information scenario. The idea is intuitive – the final estimate is the average of the n imputations – but the standard error of the estimates not only includes the error of the estimation itself, but also the error of the estimation of the imputation. By doing multiple imputation we avoid underestimating the variance as it is done in stochastic regression imputation and regression imputation.

All of these methods have limitations even for the MCAR scenario, which is highly unlikely in social sciences research. According to Enders (2010), the two state-of-the-art methodologies for imputing missing data are maximum likelihood and multiple imputation. Section 4.2 documents our use of multiple imputation.

References

Angrist, J and Pischke, JS, 2008. *Mostly harmless econometrics: an empiricist's companion*. Princeton: Princeton University Press.

Cameron, AC and Trivedi, PK, 2005. *Microeconometrics: methods and applications*. Cambridge: Cambridge University Press.

Cattaneo, MD, Galiani, S, Gertler, PJ, Martinez, S and Titiunik, R, 2009. Housing, health, and happiness. *American Economic Journal: Economic Policy,* pp.75–105.

Damián, A, 2007. Los problemas de comparabilidad de las ENIGH y su efecto en la medición de la pobreza. *Papeles de población*, 51, pp.111–146.

Devoto, F, Duflo, E, Dupas, P, Parienté, W and Pons, V, 2012. Happiness on tap: piped water adoption in urban Morocco*. American Economic Journal: Economic Policy*, 4(4), pp.68–99.

Enders, C, 2010. *Applied missing data analysis*. New York: The Guilford Press.

Galiani, S, Gertler, PJ and Schargrodsky, E, 2005. Water for life: the impact of the privatization of water services on child mortality. *Journal of Political Econom*y, 113(1): pp.83–120.

Galiani, S, Gonzales-Rozada, M and Schardgrodsky, E, 2009. Water expansion in shantytowns: health and saving. *Economica*, 76(304), pp.607–622.

Galiani, S, Gertler, P, Cooper, R, Martinez, S, Ross, A and Undurraga, R, 2015. Shelter from the storm: upgrading housing infrastructure in Latin America slums. 3ie Impact Evaluation Report 21. Washington, DC: International Initiative for Impact Evaluation (3ie).

Jones, M, 1996. Indicator and stratification methods for missing explanatory variables in multiple linear regression. *Journal of the American Statistical Association*, 91(433), pp.222–30.

Puma, M, Olsen, R, Bell, S and Price, C, 2009. What to do when data are missing in group randomized controlled trials. US Department of Education, National Center for Education Evaluation and Research Assistance, [online]. Available at: <http://ies.ed.gov/ncee/pdf/20090049.pdf> [Accessed 31 July 2015].

Soares, F and Soares, Y, 2005. The socio-economic impact of Favela-Bairro: what do the data say? *Working Paper OVE/WP-08*. Washington, DC: Inter-American Development Bank.

Turley, R, Saigh, R, Bhan, N, Rehfuess, E and Carter, B, 2013. Slum upgrading strategies involving physical environment and infrastructure interventions and their effects on health and socio-economic outcomes (review). *Cochrane Database of Systematic Reviews*, 1.

UN-Habitat, 2010. State of the world's cities 2010/2011: Bridging the urban divide.

Publications in the 3ie Replication Paper Series

The following papers are available from http://www.3ieimpact.org/en/publications/3ie-replication-paper-series/:

Quality evidence for policymaking: I'll believe it when I see the replication, 3ie Replication Paper 1. Brown, AN, Cameron, DB and Wood, BDK (2014)

TV, female empowerment and demographic change in rural India, 3ie Replication Paper 2. Iversen, V and Palmer-Jones, R (2014)

Reanalysis of health and educational impacts of a school-based deworming program in western Kenya Part 1: A pure replication, 3ie Replication Paper 3, part 1. Aiken, AM, Davey, C, Hargreaves, JR and Hayes, RJ (2014)

Reanalysis of health and educational impacts of a school-based deworming program in western Kenya Part 2: Alternative analyses, 3ie Replication Paper 3, part 2. Aiken, AM, Davey, C, Hayes, RJ and Hargreaves, JR (2014)

The long and short of returns to public investments in fifteen Ethiopian villages, 3ie Replication Paper 4. Bowser, WH (2015)

Recalling extra data: A replication study of finding missing markets, 3ie Replication Paper 5. Wood, BDK and Dong, M (2015)

The impact of India's JSY conditional cash transfer programme: A replication study, 3ie Replication Paper 6. Carvalho, N and Rokicki, S (2015)

Replication Paper Series

International Initiative for Impact Evaluation
1625 Massachusetts Ave., NW
Suite 450
Washington, DC 20036
USA

replication@3ieimpact.org
Tel: +1 202 629 3939

www.3ieimpact.org