Online Appendix for

Long-term and Intergenerational Effects of Education: Evidence from School Construction in Indonesia

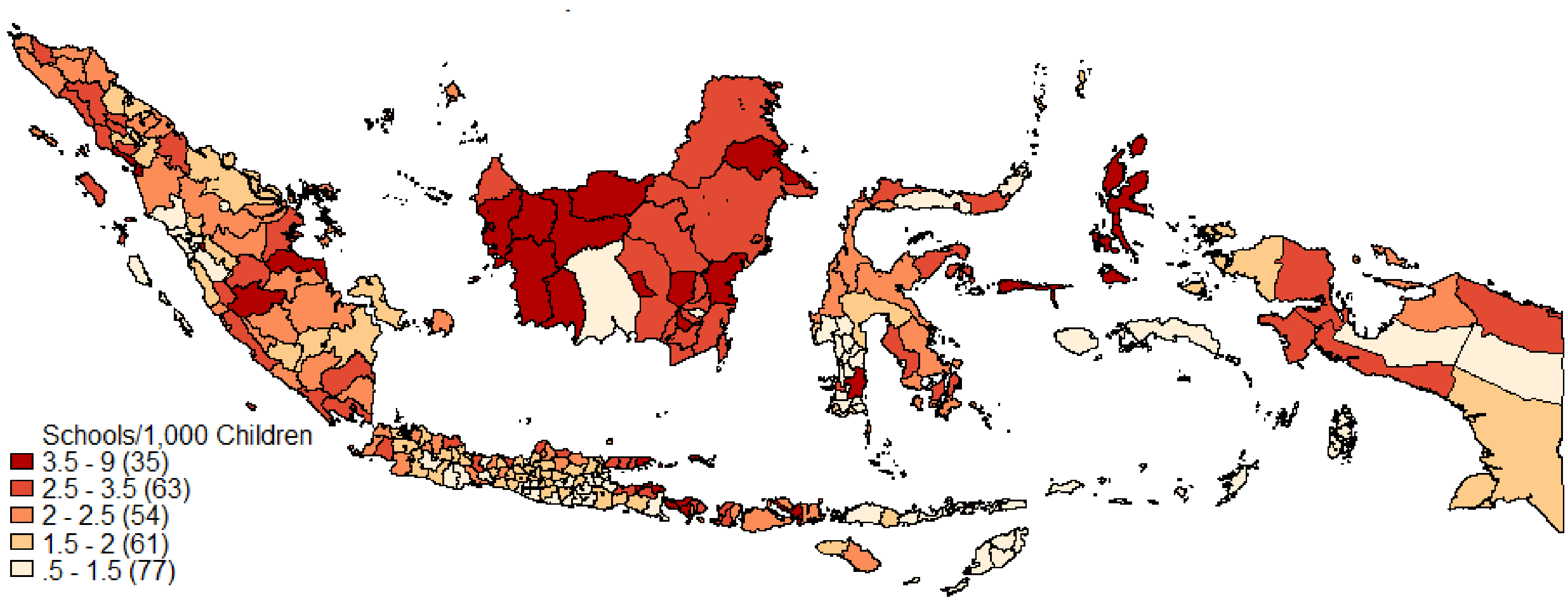
Richard Akresh
University of Illinois at Urbana-Champaign

Daniel Halim
University of Illinois at Urbana-Champaign

Marieke Kleemans
University of Illinois at Urbana-Champaign

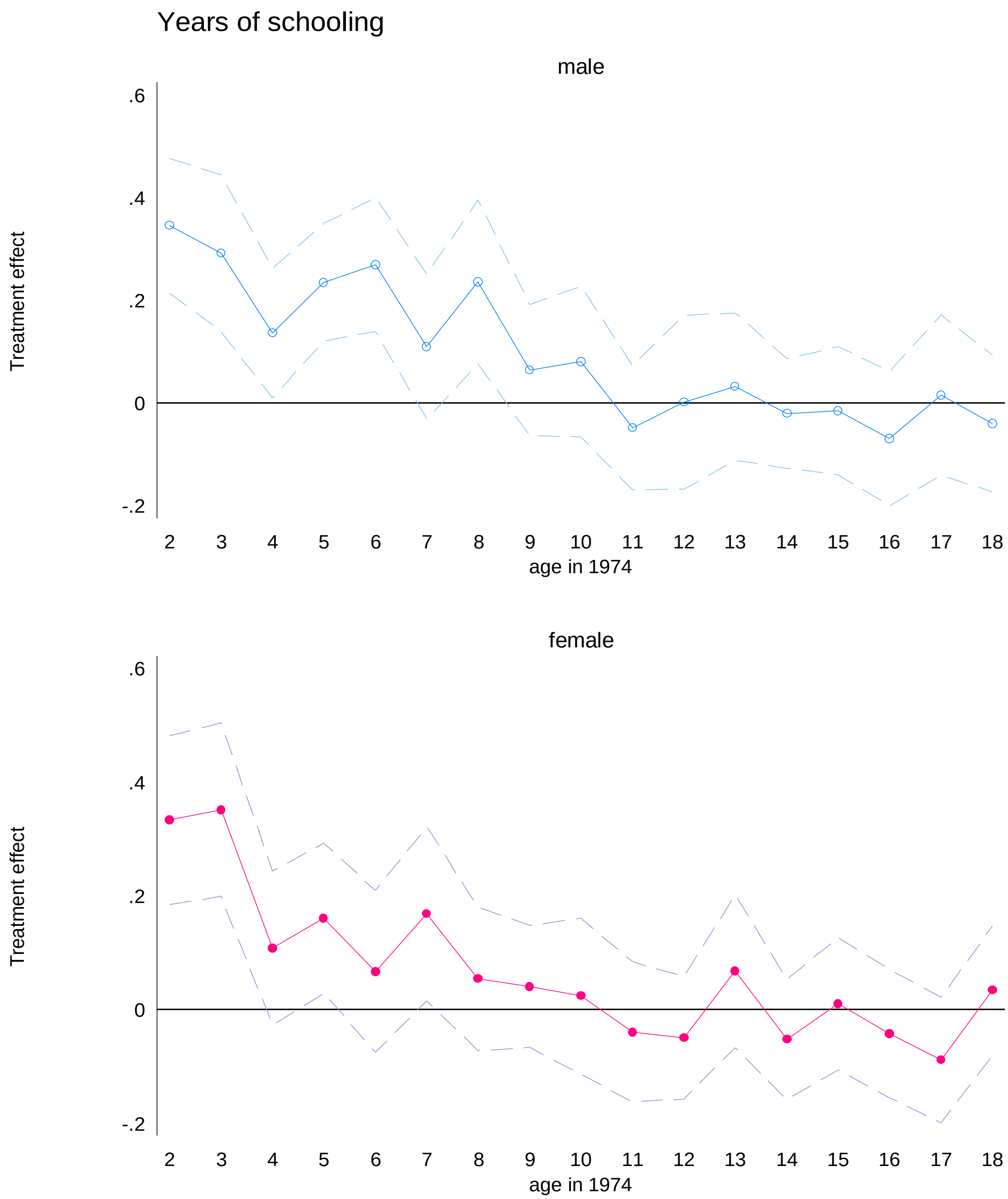
November 20, 2018

A. Online Appendix Figures and Tables



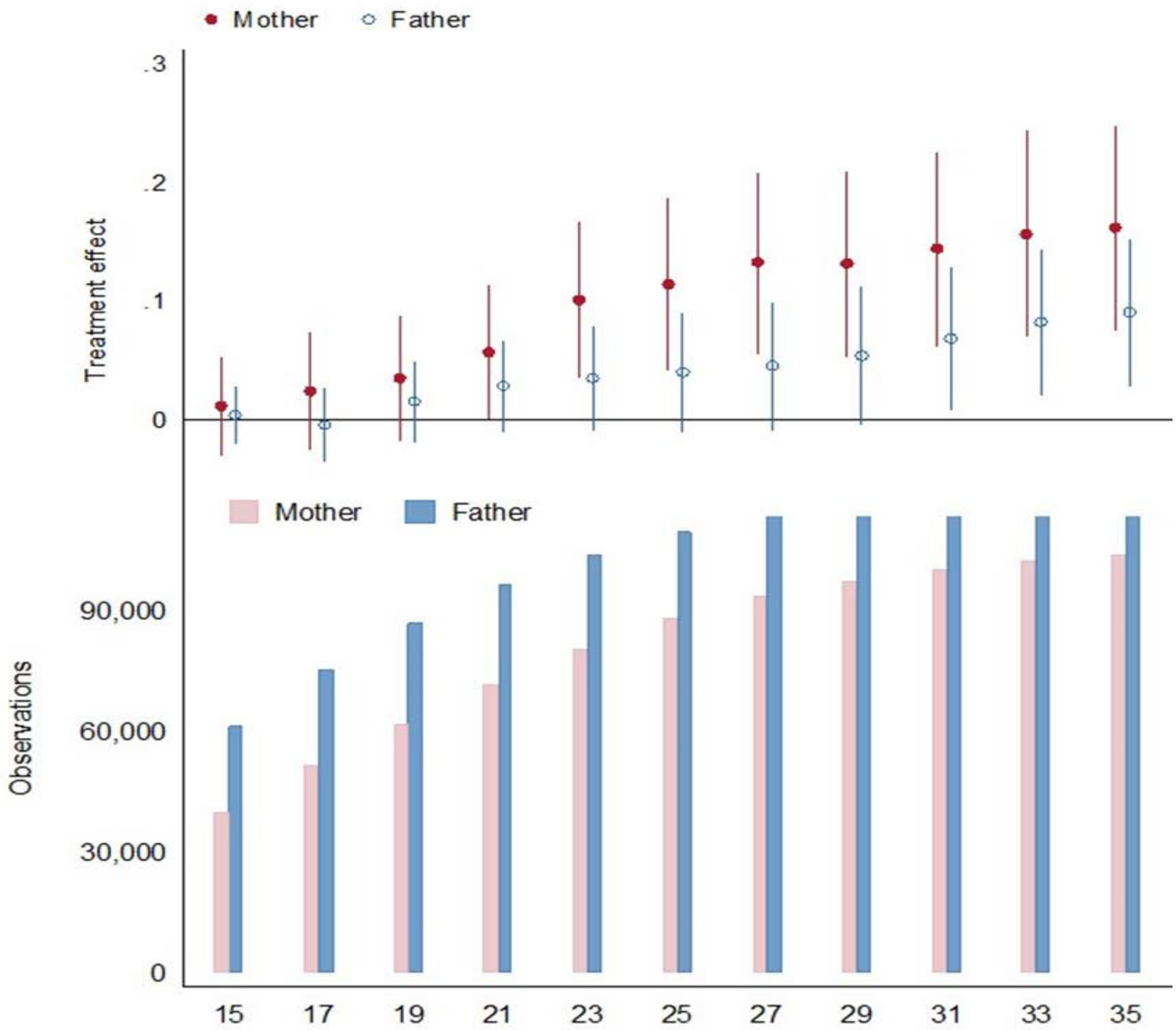
Appendix Figure A.1. Spatial distribution of schools constructed per 1,000 children between 1973 and 1979

Notes: Number of schools constructed between 1973 and 1979 and children's population in 1971 are obtained from Duflo (2001) and the Indonesian 1971 Census. The legend indicates the range and distribution of schools constructed across the Indonesian archipelago. The numbers in parentheses refer to the number of districts that fall in that range. The total number of districts, 290, reflects their existence in 1993. Districts often split over time; by March 2016, there were 511 districts. In our analyses, we maintain the 1993 district boundaries to allow matching with Duflo (2001)'s school construction data.



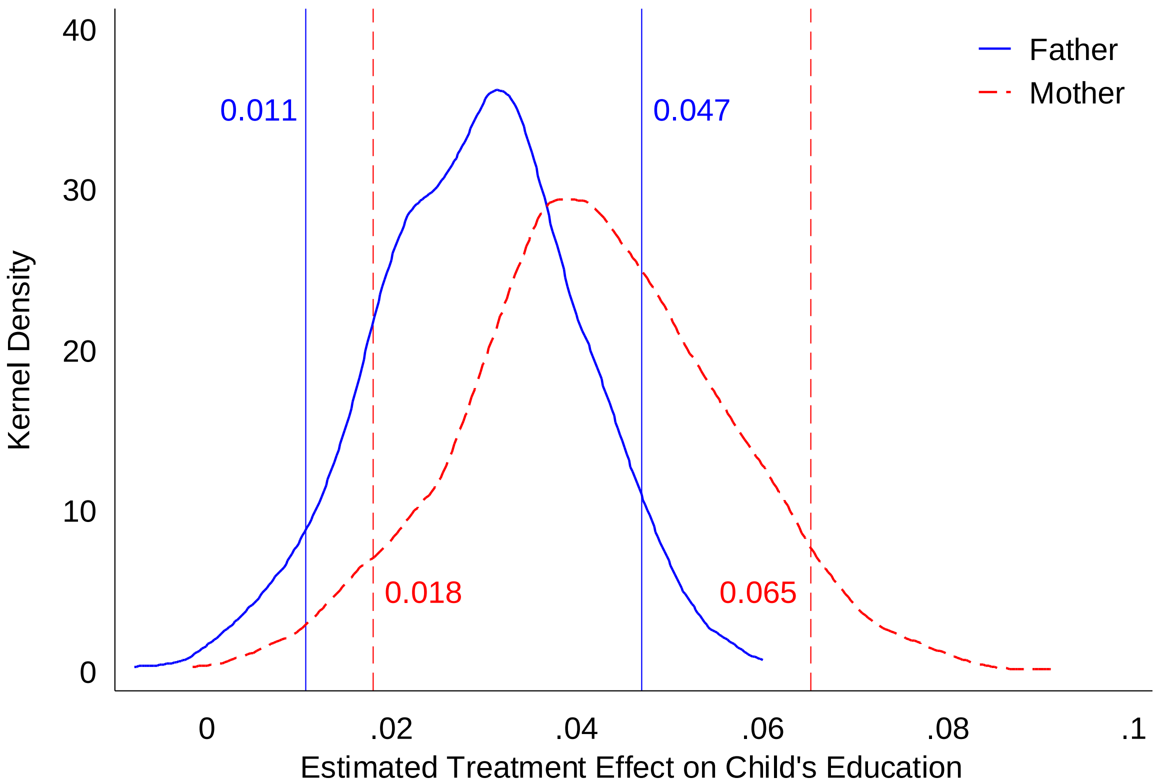
Appendix Figure A.2. Effect of school construction on first generation individual's years of schooling by age in 1974

Notes: Sample is restricted to individuals aged 2-24 in 1974 (born between 1950 and 1972). Each dot represents the interaction coefficient of the number of INPERS primary schools constructed in one's birth district and an age in 1974 dummy. The age group 19-24 is omitted from the regression. The dashed lines represent 95% confidence bands.



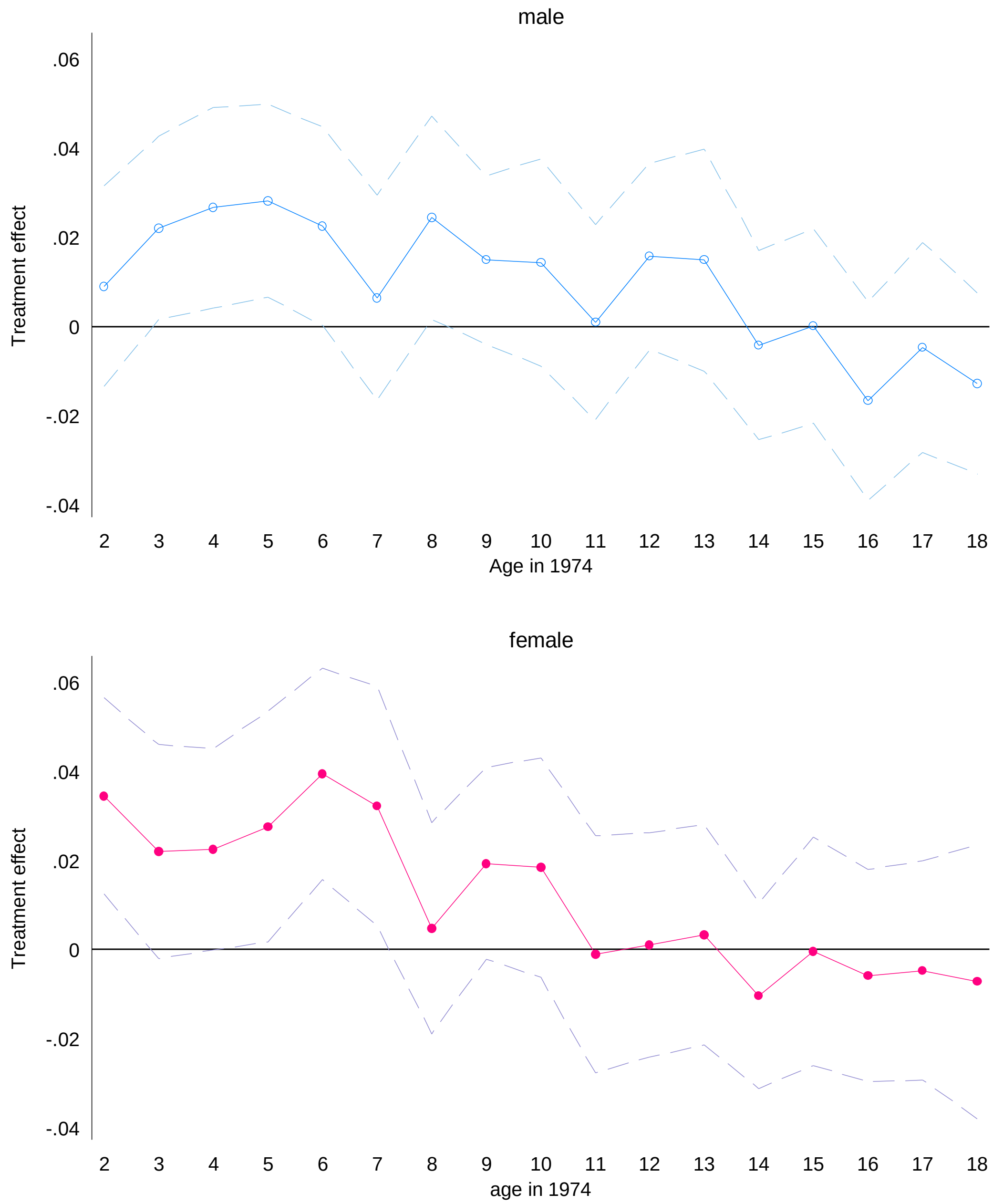
Appendix Figure A.3. Effect of school construction on second generation's years of schooling, using alternative upper-bound age restrictions

Notes: In the top panel, treatment effects indicate the effect of one additional school constructed per 1,000 children in the mother's or father's birth district on the years of schooling for second generation individuals. Each dot represents a coefficient in a separate regression. We show estimated regression coefficients and their respective 95% confidence intervals. Sample is restricted to children from age 5 up to the value on the x-axis. Bottom panel shows the number of additional observations added to each regression when the upper-age limit is increased.



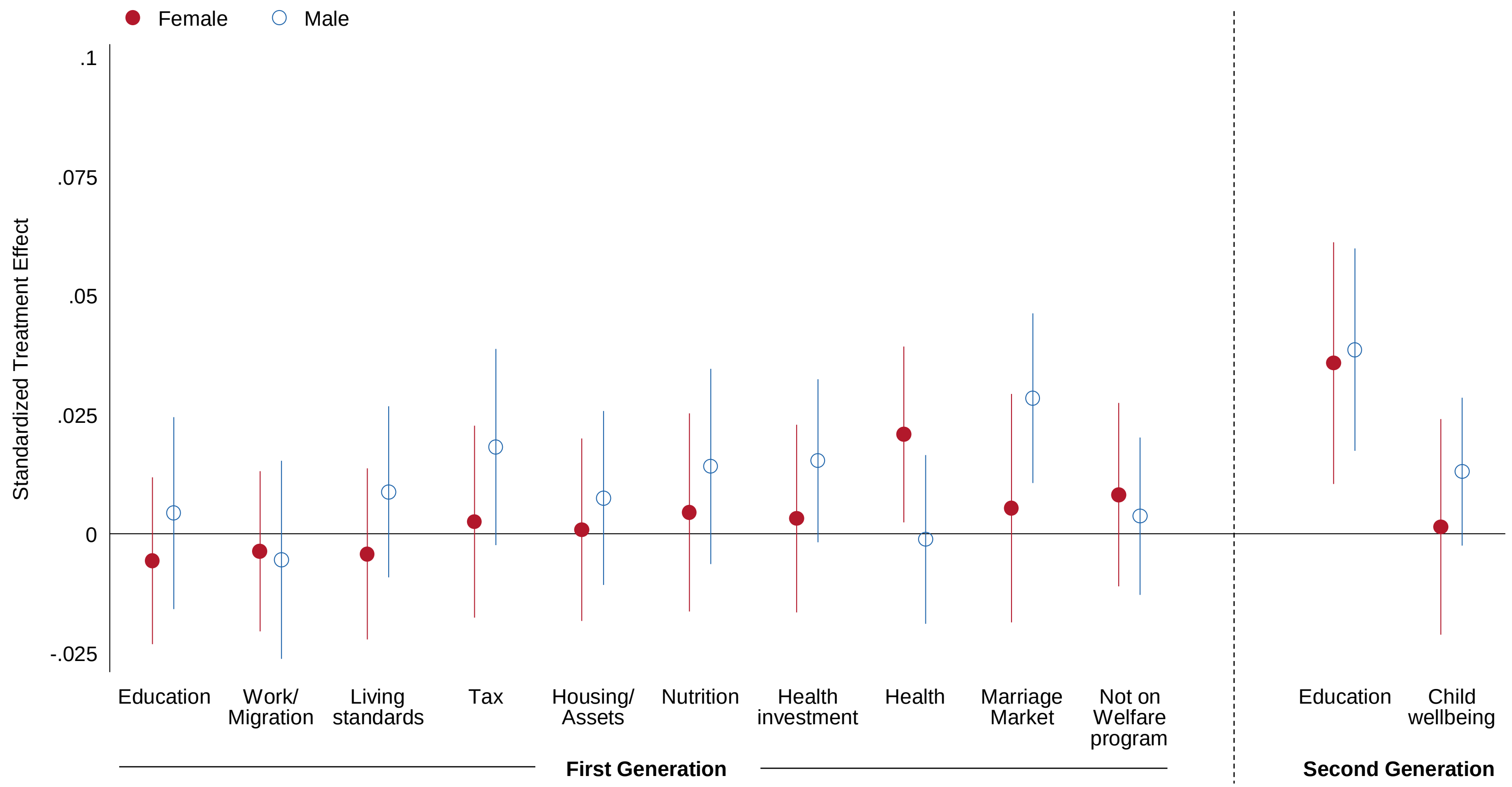
Appendix Figure A.4. Distribution of estimated treatment effects on second generation's years of schooling from simulated exposure assignment

Note: To address the selection issue about co-resident second generation children observed in the Susenas data, we use the IFLS to obtain the fraction of children at each age who are born to old and young cohort parents among all children no longer living with their parents. We then use these IFLS-based fractions to randomly assign non-co-resident children at each age in the Susenas data to either old or young cohort parents and to exclude the others from the regression. We then simulate this randomization assignment procedure 1,000 times and estimate the second generation years of schooling regression. This figure plots the density distribution of estimated coefficients from these 1,000 repetitions for father's and mother's exposure to the school construction. Solid lines indicate the distribution of father's effects and dashed lines indicate the distribution of mother's effects. Vertical lines indicate the 5^{th} and 95^{th} percentiles.



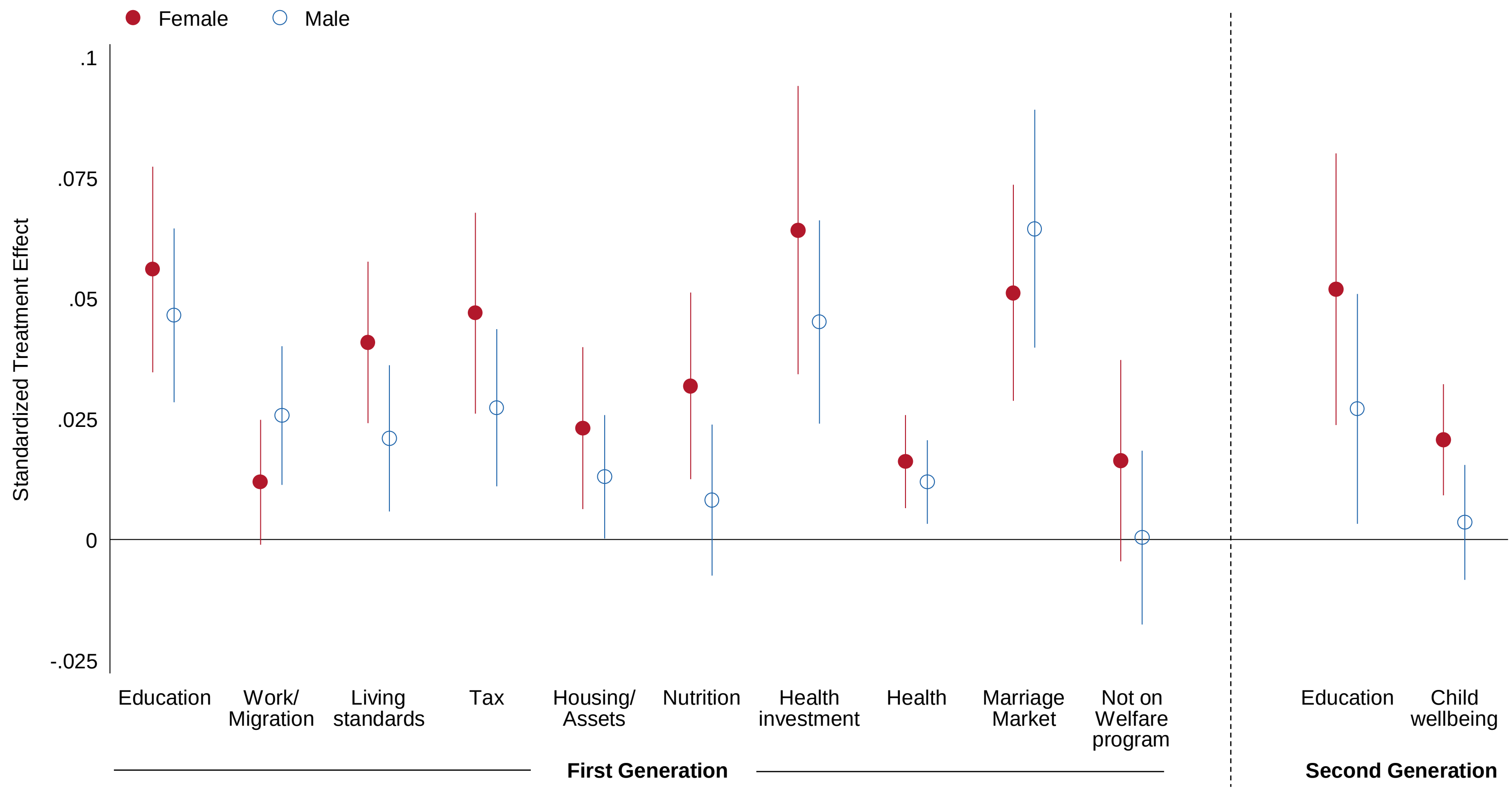
Appendix Figure A.5. Effect of school construction on household expenditures by age in 1974

Notes: Sample is restricted to individuals aged 2-24 in 1974 (born between 1950 and 1972). Each dot represents the interaction coefficient of the number of INPERS primary schools constructed in one's birth district and an age in 1974 dummy. The age group 19-24 is omitted from the regression. The dashed lines represent 95% confidence bands.



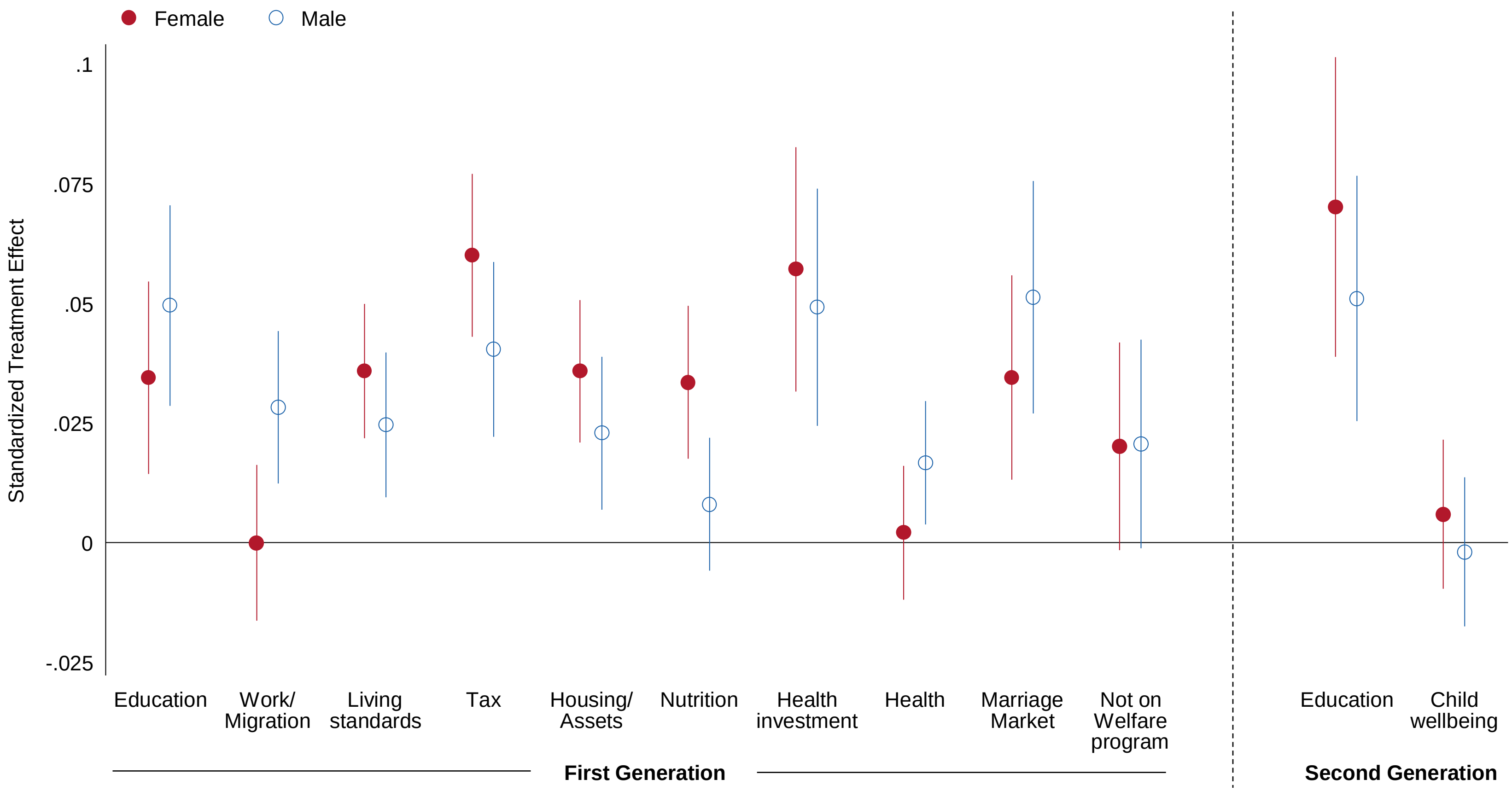
Appendix Figure A.6. Placebo effect of school construction on indexes of long-run outcomes for individuals too old to benefit from primary school construction

Notes: Similar to Figure 1, where we compare individuals born between 1957-1962 (old cohort) and 1968-1972 (young cohort), we now estimate a placebo regression by restricting the sample to individuals born between 1950-1956 (an older cohort) and 1957-1962 (old cohort). Each dot represents the interaction coefficient of the number of INPRES schools built between 1973 and 1979 in one's birth district and a dummy for being born between 1957 and 1962. The solid lines represent 95% confidence bands. This figure serves as a placebo test since the old cohort was too old to be enrolled in primary school when the schools were constructed, and thus could not benefit from the school construction. The individual outcomes making up the index for each family are listed in Tables 1-7 and Appendix Tables A.1 to A.5.



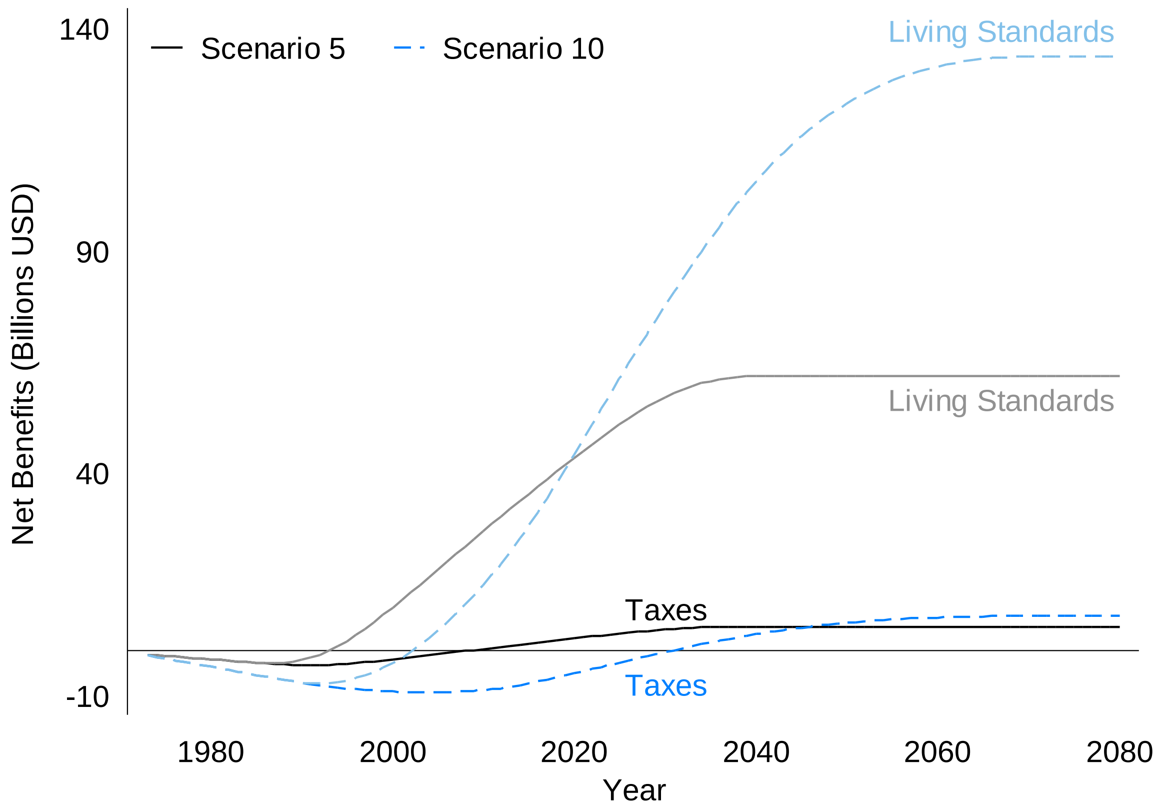
Appendix Figure A.7. Effect of school construction on indexes of long-run outcomes extending the sample to all individuals born between 1950 and 1980

Notes: Similar to Figure 1, but regressions now include all individuals born between 1950 and 1980. Each dot represents the interaction coefficient of the number of INPRES schools built between 1973 and 1979 in one's birth district and a dummy for being born between 1968 and 1980. The solid lines represent 95% confidence intervals. The individual outcomes making up the index for each family are listed in Tables 1-7 and Appendix Tables A.1 to A.5.



Appendix Figure A.8. Effect of school construction on indexes of long-run outcomes using alternative control variables

Notes: Similar to Figure 1, but regressions now exclude the interaction of birth year dummies and water and sanitation programs from the control variables. Each dot represents the interaction coefficient of the number of INPRES schools built between 1973 and 1979 in one's birth district and a dummy for being born between 1968 and 1972. The solid lines represent 95% confidence intervals. The individual outcomes making up the index for each family are listed in Tables 1-7 and Appendix Tables A.1 to A.5.



Appendix Figure A.9. Discounted net benefits of school construction in Indonesia

Note: We plot net benefits (the difference in discounted total benefits and total costs) over time. Benefits are either tax receipts collected by the government or improved living standards of the citizens. Net benefits are reported in billions of 2016 USD. We present two scenarios using the parameters from the cost-benefit model in column (5) and column (10) of Table 11. Solid lines indicate net benefits—in taxes and living standards—under Scenario 5. Dashed lines indicate net benefits under Scenario 10.

Appendix Table A.1. Effect of school construction on first generation's housing and assets

		Mean / SD		Effect of Program Exposure on:	
Outcome	**Description**	**Men**	**Women**	**Men**	**Women**
Urban	Indicator for residing in an urban area	0.425	0.438	-0.001	0.002
		(0.494)	(0.496)	(0.004)	(0.004)
				[0.822]	[0.576]
Rent equivalent (Rp10k)	Actual monthly rent if house is rented, or estimated rent value if house is owned or leased by the employer	42.991	43.085	0.012	0.028***
		(56.342)	(56.573)	(0.008)	(0.008)
				[0.293]	[0.001]
Floor area (m^2)	House's floor area in square meters	79.894	81.355	1.229**	1.480***
		(58.651)	(59.726)	(0.566)	(0.510)
				[0.119]	[0.011]
Utilities (Rp10k)	Expenditure on electricity, water, gas, and kerosene	15.714	15.729	0.051**	0.085***
		(20.983)	(21.796)	(0.022)	(0.024)
				[0.102]	[0.002]
Asset index	PCA index on binary ownerships of motorcycle, car, home phone, computer, TV, jewelry, refrigerator, water heater, LPG gas tube, boat, and air conditioner	-0.035	-0.069	0.030*	0.040**
		(1.868)	(1.882)	(0.017)	(0.015)
				[0.223]	[0.020]
Housing/Assets index	Aggregates all 5 outcomes and standardizes it to the mean of the old cohort in low-program regions. Effects are interpreted as standard deviation changes from the mean.			0.021**	0.035***
				(0.009)	(0.009)

Notes: Effect of program exposure are the regression coefficients of young cohort dummy interacted with the number of schools constructed in region of birth. All regressions control for district of birth and cohort of birth fixed effects, children's population and enrollment in 1971, and water and sanitation program intensities that vary by region of birth interacted with birth year dummies. Robust standard errors clustered at region of birth are shown in parentheses. Stars denote statistical significance at 1, 5, and 10% levels based on unadjusted p-values. FDR q-values are computed over all 5 outcomes and are shown in square brackets. FDR q-values indicate the probability of false positives among *significant* tests. There are 68,687 and 66,249 observations in the men's and women's regressions, respectively. We apply an inverse hyperbolic sine transformation to all monetary values. Estimates can be interpreted as percentage changes.

Appendix Table A.2. Effect of school construction on first generation's nutrition

		Mean / SD		Effect of Program Exposure on:	
Outcome	**Description**	**Men**	**Women**	**Men**	**Women**
Calories	Household's accounts of units of food consumed in the past week (e.g. 5 kg of rice) are converted into nutritional intake by the Central Statistics Agency. Following their procedure, we convert the weekly intake to monthly intake. In the regressions, we apply an inverse hyperbolic transformation for reasons discussed above. The mean of calories intake is reported in 1 kcal increments. The means of protein, fat, and carbohydrate intakes are reported in 1 kg increments.	260.915 (106.001)	249.699 (109.833)	0.005 (0.004) [0.301]	0.018*** (0.005) [0.001]
Protein		7.116 (3.254)	6.831 (3.330)	0.006 (0.005) [0.301]	0.018*** (0.005) [0.001]
Fat		6.074 (3.110)	5.810 (3.150)	0.011** (0.004) [0.061]	0.023*** (0.006) [0.000]
Carbohydrates		40.869 (17.728)	39.040 (18.245)	0.005 (0.004) [0.301]	0.017*** (0.005) [0.001]
Nutrition index	Aggregates all 4 outcomes and standardizes it to the mean of the old cohort in low-program regions. Effects are interpreted as standard deviation changes from the mean.			0.014 (0.009)	0.039*** (0.010)

Notes: Effect of program exposure are the regression coefficients of young cohort dummy interacted with the number of schools constructed in region of birth. All regressions control for district of birth and cohort of birth fixed effects, children's population and enrollment in 1971, and water and sanitation program intensities that vary by region of birth interacted with birth year dummies. Robust standard errors clustered at region of birth are shown in parentheses. Stars denote statistical significance at 1, 5, and 10% levels based on unadjusted p-values. FDR q-values are computed over all 5 outcomes and are shown in square brackets. FDR q-values indicate the probability of false positives among *significant* tests. There are 68,687 and 66,249 observations in the men's and women's regressions, respectively. We apply an inverse hyperbolic sine transformation in the regressions. Estimates can be interpreted as percentage changes.

Appendix Table A.3. Effect of school construction on first generation's health investment

		Mean / SD		Effect of Program Exposure on:	
Outcome	**Description**	**Men**	**Women**	**Men**	**Women**
Total health expenditure (Rp10k)	Total monthly household health expenditures, which aggregates curative, medicine, and preventive health expenditures	7.517 (34.130)	7.961 (35.245)	0.071* (0.038) [0.114]	0.055 (0.041) [0.185]
Preventive measures (Rp10k)	Consist of pregnancy checks, immunizations, medical check-ups, family planning, and other expenditures, e.g., vitamins, massage, gym memberships	0.744 (3.225)	0.671 (3.135)	0.242*** (0.068) [0.002]	0.193*** (0.071) [0.013]
Family planning (Rp10k)	A sub-category under preventive health expenditures, which includes costs of contraceptives and consultations	0.286 (0.872)	0.219 (0.856)	0.321*** (0.061) [0.000]	0.226*** (0.071) [0.008]
Private hospital (Rp10k)	A sub-category under curative health expenditures and is distinct from expenditures on public hospitals, clinics, and traditional healers	2.101 (20.718)	2.200 (22.266)	0.048** (0.023) [0.114]	0.075*** (0.024) [0.008]
Health insurance (Rp10k)	Health insurance is distinct from life, accidental, vehicle, and house insurances	3.821 (16.425)	3.635 (14.047)	0.083 (0.055) [0.134]	0.142*** (0.048) [0.009]
Health investment index	Aggregates all 5 outcomes and standardizes it to the mean of the old cohort in low-program regions. Effects are interpreted as standard deviation changes from the mean.			0.065*** (0.015)	0.063*** (0.016)

Notes: Effect of program exposure are the regression coefficients of young cohort dummy interacted with the number of schools constructed in region of birth. All regressions control for district of birth and cohort of birth fixed effects, children's population and enrollment in 1971, and water and sanitation program intensities that vary by region of birth interacted with birth year dummies. Robust standard errors clustered at region of birth are shown in parentheses. Stars denote statistical significance at 1, 5, and 10% levels based on unadjusted p-values. FDR q-values are computed over all 5 outcomes and are shown in square brackets. FDR q-values indicate the probability of false positives among *significant* tests. There are 68,687 and 66,249 observations in the men's and women's regressions, respectively. We apply an inverse hyperbolic sine transformation to all monetary values. Estimates can be interpreted as percentage changes.

Appendix Table A.4. Effect of school construction on first generation's welfare program participation

		Mean / SD		Effect of Program Exposure on:	
Outcome	**Description**	**Men**	**Women**	**Men**	**Women**
Cash Transfer	Unconditional cash transfer to compensate for the removal of gas price subsidy for poor households	0.041 (0.197)	0.039 (0.194)	-0.002 (0.002) [0.742]	-0.001 (0.002) [0.914]
Rice for Poor	Monthly rice allowance for poor households	0.392 (0.488)	0.406 (0.491)	0.002 (0.004) [0.850]	-0.009* (0.005) [0.200]
Poor Student's Assistance	Cash transfer conditional on school enrollment	0.056 (0.363)	0.127 (0.333)	-0.001 (0.004) [0.850]	0.000 (0.004) [0.914]
Social Protection Card	Card provided to poor households, which entitles them to social welfare programs mentioned above	0.186 (0.389)	0.180 (0.384)	-0.001 (0.004) [0.850]	-0.000 (0.004) [0.914]
Welfare program non-participation index	Aggregates all 4 outcomes and standardizes it to the mean of the old cohort in low-program regions. Effects are interpreted as standard deviation changes from the mean. For the index, we reverse the sign for the 4 welfare programs to indicate a positive outcome.			0.006 (0.011)	0.010 (0.012)

Notes: Means indicate the fraction of program recipients. Effect of program exposure are the regression coefficients of young cohort dummy interacted with the number of schools constructed in region of birth. All regressions control for district of birth and cohort of birth fixed effects, children's population and enrollment in 1971, and water and sanitation program intensities that vary by region of birth interacted with birth year dummies. Robust standard errors clustered at region of birth are shown in parentheses. Stars denote statistical significance at 1, 5, and 10% levels based on unadjusted p-values. FDR q-values are computed over all 4 outcomes and are shown in square brackets. FDR q-values indicate the probability of false positives among *significant* tests. There are 68,687 and 66,249 observations in the men's and women's regressions, respectively.

Appendix Table A.5. Effect of school construction on second generation's child wellbeing

		Mean / SD		Effect of Program Exposure by:	
Outcome	**Description**	**Father**	**Mother**	**Fathers on Children**	**Mothers on Children**
Non-work days	Number of days not worked in the past week by the child unconditional on work, i.e. 7 for non-working individuals	5.317 (2.670)	4.820 (2.865)	0.044** (0.021) [0.136]	0.031 (0.019) [0.463]
Non-work hours	Number of hours not worked in the past week by the child unconditional on work, i.e. 168 for non-working individuals	156.679 (19.704)	153.047 (21.597)	0.299* (0.157) [0.173]	0.215 (0.151) [0.463]
No health complaint	Self-reported indicator defined as 1 if child did not experience a health complaint in the past month	0.797 (0.402)	0.823 (0.382)	-0.008*** (0.003) [0.042]	0.004 (0.003) [0.463]
Non-disrupted days	Self-reported number of days in the past month (maximum of 30 days) that a health complaint did not disrupt child's daily activities	29.492 (2.086)	29.550 (2.067)	-0.026* (0.016) [0.198]	0.007 (0.015) [0.893]
No severe health complaint	Self-reported indicator defined as 1 if child did not experience a severe health complaint in the past month	0.978 (0.147)	0.980 (0.140)	-0.000 (0.001) [0.751]	-0.000 (0.001) [0.893]
Second generation wellbeing index	Aggregates all 5 outcomes and standardizes it to the mean of the old cohort in low-program regions. Effects are interpreted as standard deviation changes from the mean.			-0.004 (0.009)	0.017** (0.008)

Notes: Effect of program exposure are the regression coefficients of father or mother's young cohort dummy interacted with the number of schools constructed in father or mother's region of birth. All regressions control for parent's district of birth and cohort of birth fixed effects, child age fixed effects, children's population and enrollment in 1971, and water and sanitation program intensities that vary by region of birth interacted with birth year dummies. Robust standard errors clustered at parent's region of birth are shown in parentheses. Stars denote statistical significance at 1, 5, and 10% levels based on unadjusted p-values. FDR q-values are computed over all 5 outcomes and are shown in square brackets. FDR q-values indicate the probability of false positives among *significant* tests. The survey restricts questions on labor market outcomes to individuals aged 10 and older; questions on health outcomes are asked to all individuals. There are 100,293 and 94,067 observations in the father's and mother's regressions for labor market outcomes; 129,971 and 108,607 observations in the father's and mother's regressions for health outcomes.

Appendix Table A.6. Spouse's characteristics as mediators of the effect of school construction on first generation's living standards

	Dependent Variable: Living standards									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Mediator:	None	Years of Schooling	Completed Primary	Literate	Work	Formal worker	Non-agriculture sector	Migrant	No health complaint	All
Panel A: Father										
Schools constructed * Young cohort	0.021***	0.009	0.007	0.010	0.020***	0.019**	0.017**	0.018**	0.020***	0.008
	(0.007)	(0.008)	(0.007)	(0.007)	(0.008)	(0.009)	(0.008)	(0.008)	(0.008)	(0.009)
Mediator		0.062***	0.337***	0.374***	0.002	0.371***	0.480***	0.273***	-0.009	
		(0.001)	(0.009)	(0.014)	(0.008)	(0.014)	(0.010)	(0.011)	(0.006)	
Observations	68,687	64,416	64,416	64,416	64,416	39,545	39,545	64,416	64,416	39,545
Mean	8.011	8.068	8.068	8.068	8.068	8.007	8.007	8.068	8.068	8.007
Panel B: Mother										
Schools constructed * Young cohort	0.032***	0.016**	0.015**	0.017**	0.023***	0.018**	0.018**	0.021***	0.023***	0.012
	(0.007)	(0.008)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Mediator		0.063***	0.351***	0.380***	-0.057***	0.319***	0.412***	0.293***	0.030***	
		(0.001)	(0.009)	(0.015)	(0.013)	(0.012)	(0.009)	(0.012)	(0.007)	
Observations	66,249	55,449	55,449	55,449	55,449	50,884	50,884	55,449	55,449	50,884
Mean	7.152	7.313	7.313	7.313	7.313	7.323	7.323	7.313	7.313	7.323

Notes: Each column shows a regression of the first generation's living standards on exposure to the school construction program and includes a potential mediator variable. These mediator variables (as indicated by the column heading) are spouse's characteristics in Table 9. Regressions are as in row 1 of Table 3. Effect of program exposure are the regression coefficients of father or mother's young cohort dummy interacted with the number of schools constructed in father or mother's region of birth. All regressions control for parent's district of birth and cohort of birth fixed effects, children's population and enrollment in 1971, and water and sanitation program intensities that vary by region of birth interacted with birth year dummies. Robust standard errors clustered at parent's region of birth are shown in parentheses. Stars denote statistical significance at 1, 5, and 10% levels based on unadjusted p-values.

Appendix Table A.7. Effect of school construction on second generation's years of schooling on various samples

	Susenas	Susenas with Extreme Assumptions		IFLS		
	(1)	(2)	(3)	(4)	(5)	(6)
		Assume Not Exposed	Assume Exposed	All	Stayers	Movers
Panel A: Father						
Schools constructed * Young cohort	0.097***	0.021	0.000	0.103	0.030	-0.020
	(0.032)	(0.016)	(0.014)	(0.104)	(0.109)	(0.251)
Observations	120,838	644,675	644,675	6,186	4,048	2,138
Mean	7.967	7.731	7.731	7.807	6.434	10.396
Panel B: Mother						
Schools constructed * Young cohort	0.169***	0.052***	0.030*	0.300**	0.539***	0.126
	(0.045)	(0.017)	(0.017)	(0.147)	(0.128)	(0.239)
Observations	105,523	644,675	644,675	7,227	3,756	3,471
Mean	8.854	7.731	7.731	9.038	8.097	10.034

Note: Column (1) is from Table 11. Column (2) and (3) estimate extreme bounds in which all non-co-resident children aged 0-40 are assumed to have parents who are either exposed or not exposed (Manski, 1990). Columns (4)-(6) use the IFLS 2014 Round 5 data. We match parents to their co-resident children ("Stayers") found in the household roster and to their non-co-resident children ("Movers") in the respective module.

Appendix Table A.8. Effect of school construction on first generation's years of schooling (extended cohort definitions)

	(1)	(2)	(3)	(4)	(5)
Cohorts Included:	1957-1962 and 1968-1972	… + 1950-1956	… + 1963-1967	… + 1973-1980	1950-1980
Panel A: Male					
Schools constructed *	0.268***	0.267***	0.221***	0.211***	0.172***
Young cohort	(0.047)	(0.039)	(0.037)	(0.044)	(0.032)
Observations	72,367	98,895	98,781	138,617	197,951
Mean	8.022	7.500	7.938	8.478	8.047
Panel B: Female					
Schools constructed *	0.234***	0.219***	0.209***	0.245***	0.210***
Young cohort	(0.042)	(0.044)	(0.039)	(0.044)	(0.045)
Observations	71,423	97,268	99,843	140,142	200,644
Mean	7.105	6.496	6.901	7.790	7.194

Notes: Robust standard errors clustered at region of birth are shown in parentheses. Stars denote statistical significance at 1, 5, and 10% levels based on unadjusted p-values. Column (1) sample is restricted to individuals born in the sample period 1957-1962 (old cohort) and 1968-1972 (younger cohort) and is the sample used in the analysis in the rest of the paper. Columns (2) to (5) extend the sample as indicated in the column headings. Panel A looks only at males and Panel B only at females. School constructed denotes the number of INPRES schools constructed per 1,000 children in one’s birth district. Young cohort is an indicator defined as 1 for being born after 1967.

Appendix Table A.9. Effect of school construction on first generation's household expenditures (various transformations)

	Total expenditure				Education expenditure			
	IHS Total	Log	Nominal	IHS Per-capita	IHS Total	Log	Nominal	IHS Per-capita
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Father								
Schools constructed *	0.021***	0.021***	9.882***	0.016**	0.160**	0.013	0.309	0.140**
Young cohort	(0.007)	(0.007)	(3.628)	(0.007)	(0.064)	(0.010)	(0.309)	(0.056)
Observations	68,687	68,687	68,687	68,687	68,687	48,123	68,687	68,687
Mean	391.649	391.649	391.649	391.649	13.971	13.971	13.971	13.971
Panel B: Mother								
Schools constructed *	0.032***	0.032***	11.022***	0.018***	0.193**	-0.010	-0.191	0.167**
Young cohort	(0.007)	(0.007)	(2.583)	(0.007)	(0.076)	(0.014)	(0.383)	(0.067)
Observations	66,249	66,249	66,249	66,249	66,249	39,492	66,249	66,249
Mean	375.616	375.616	375.616	375.616	12.202	12.202	12.202	12.202

Notes: Effect of program exposure are the regression coefficients of young cohort dummy interacted with the number of schools constructed in region of birth. All regressions control for district of birth and cohort of birth fixed effects, children's population and enrollment in 1971, and water and sanitation program intensities that vary by region of birth interacted with birth year dummies. Robust standard errors clustered at region of birth are shown in parentheses. Stars denote statistical significance at 1, 5, and 10% levels based on unadjusted p-values. All expenditure values are defined at the household level and refer to the household's average monthly expenditure. Nominal values are reported in 10,000 Indonesian Rupiah (IDR) increments. In 2016, the average daily exchange rates was 1 USD=13,308 IDR. Columns (1)-(4) examine total household expenditure; columns (5)-(8) examine education expenditure. Inverse hyperbolic sine (IHS) transformations are applied to total and per capita household expenditures (columns 1 and 4) and to total and per-capita education expenditures (columns 5 and 8). Log transformations are applied in columns (2) and (6). Column (1) and (5) are the preferred specification and are the same as Table 3, rows 1 and 5.

B. Data Appendix

Two critical data issues about the Susenas 2016 survey are relevant for our analysis. First, to estimate the difference-in-differences specification described in Section 3.1, it is necessary to have information about an individual's residence at birth. Current residence could be endogenous to the school construction program as households might move to provide access to schools to their children. Location of birth and location where the individual obtains their education are highly correlated.[1] However, birth location is not endogenous with respect to the school construction since all of the individuals in the analysis were born before the program started. Given the importance of knowing where the individual was born, it is unfortunate that most household surveys in Indonesia only provide information about the individual's current location of residence. This lack of information about an individual's birth location is the case for the Indonesian Labor Force Survey (Sakernas), the Indonesian Demographic and Health Surveys (DHS), and many other rounds of the Susenas data, making them unavailable to use to analyze the impacts of the school construction program. However, the Susenas 2016 is one exception to this, as there is information on every individual's district of birth.

Second, it is important that the data include a sufficiently large sample of individuals from these specific birth cohorts (1957-1962 and 1968-1972). The Indonesia Family Life Survey (IFLS) does contain information on each individual's region of residence at birth, thus satisfying the first criteria we outline above. We use the IFLS data to estimate our main difference-in-difference specification exploiting variation across birth cohorts and regions in the number of schools built. The IFLS is a longitudinal survey, and the first round was collected in 1993/1994. Subsequent rounds were collected in 1997, 2000, 2007/2008, and most recently in 2014/2015. Tracking across rounds has been extremely successful, with rates between 92 to 95 percent for each IFLS round (Thomas et. al., 2012). Almost 88 percent of households in survey round one were subsequently interviewed in all of the five survey rounds. In columns 1-3 of Appendix Table B.1, we use the most recent survey round collected in 2014/2015 (IFLS 5) and include all individuals interviewed in that round in the regressions. In columns 4-6, we begin with the IFLS 5 and then add in any other individuals from the other four rounds who might no longer be present in the final round of the panel survey. We estimate regressions with different control variables to see if that has any influence on the results. Column 3 (IFLS 5 only) and column 6 (IFLS 5 plus last observed round) correspond with our main results for men and women in Table 1 row 1. We do not observe any statistically significant relationship between exposure to the school construction and increased years of schooling.

The IFLS and Susenas data have two key differences that might be relevant to explain this situation. First, the Susenas data is nationally representative covering all 34 provinces and all 511 districts in the country. IFLS is representative of only 83% of the Indonesian population and covers individuals living in 13 out of 27 provinces in the country. Appendix Figure B.1 shows a map of Indonesian districts with the districts shaded in gray indicating which ones the IFLS survey covers. Comparing Figure 2 (map of Indonesia indicating the spatial distribution of school constructed per 1,000 children) and Appendix Figure B.1 highlights that many of the

[1] Based on the IFLS data, almost 92 percent of children at age 12 still live in the same district where they were born (Duflo, 2001). Likewise, in the Susenas 2016, 93.2 percent of children at age 12 live in the same district where they were born.

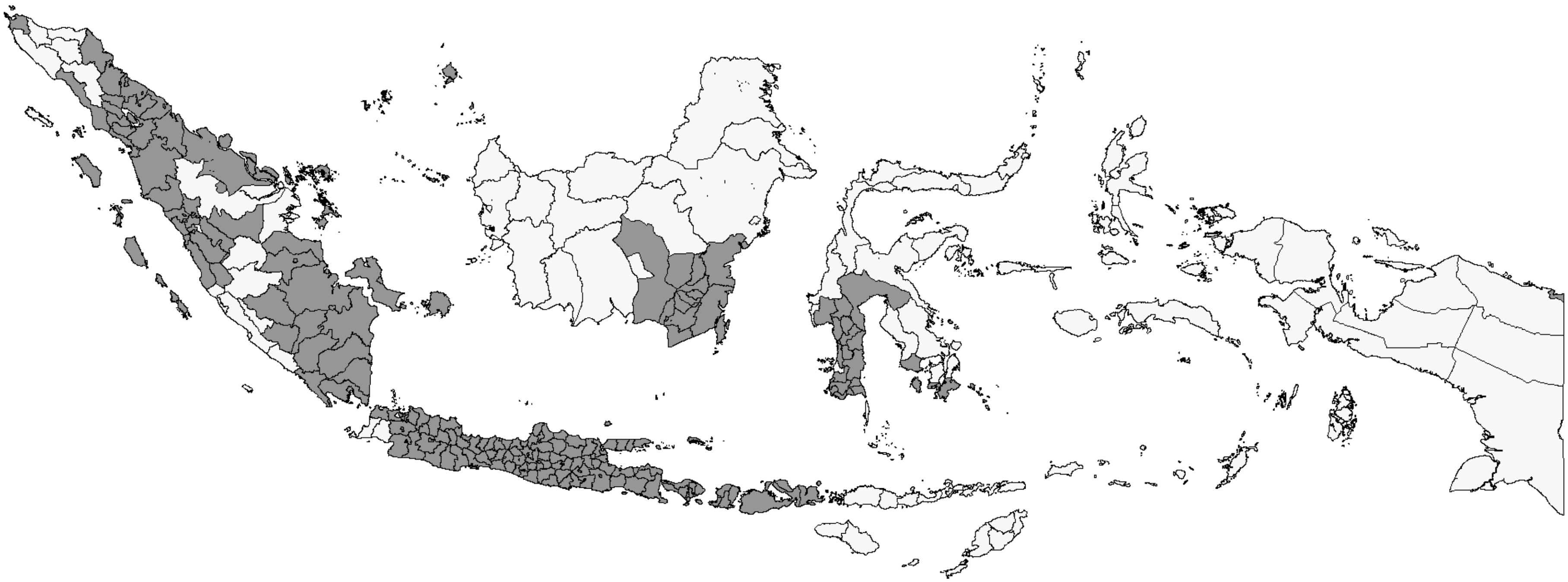
districts that had many schools constructed are not included in the IFLS survey. In column 7 of Appendix Table B.1, we present results using the Susenas 2016 data but restricting the analysis to only those districts covered in the IFLS survey. The coefficients from the regression with this restricted sample are somewhat smaller (0.211 for men and 0.166 for women) compared to the full sample from Table 1 (0.268 for men and 0.234 for women), but the results are still statistically significant and economically meaningful. This is evidence that the different geographic coverage of the IFLS and the Susenas is unlikely to explain the lack of relationship between school construction and years of schooling in the IFLS data (columns 1-6). Second, note that the number of observations in the IFLS regressions for women is only 2,546 if using only IFLS 5 or 2,783 if using IFLS 5 plus the last observed round for any individual.[2] This compares with 71,423 observations for women in the regression using the Susenas data. While the point estimates for women are similar across the two datasets, this difference in sample size could explain the much larger standard errors in the regressions using IFLS data.

[2] Using the extended cohort of individuals born between 1950 and 1980 roughly triples the sample size (for men to 7,093 and 7,666 and for women to 7,382 and 8,018 in the IFLS 5 and IFLS 5 plus last observed round, respectively), but the results are still not statistically significant.

Appendix Table B.1. Effect of school construction on first generation's education using IFLS data

Data source:	IFLS 5 (2014/2015)			IFLS 5 + last observed round			Susenas 2016 restricted to IFLS districts
Dependent variable: Years of schooling	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Male							
Schools constructed * Young cohort	-0.144	-0.122	-0.032	-0.224	-0.186	-0.220	0.211***
	(0.141)	(0.148)	(0.187)	(0.160)	(0.173)	(0.204)	(0.063)
Observations	2,389	2,389	2,389	2,609	2,609	2,609	54,646
Children population in 1971	X	X	X	X	X	X	X
Enrollment in 1971		X	X		X	X	X
Water and sanitation program			X			X	X
Panel B: Female							
Schools constructed * Young cohort	0.037	0.053	0.207	0.215	0.229	0.318	0.166***
	(0.174)	(0.176)	(0.214)	(0.151)	(0.148)	(0.198)	(0.056)
Observations	2,546	2,546	2,546	2,783	2,783	2,783	54,508
Children population in 1971	X	X	X	X	X	X	X
Enrollment in 1971		X	X		X	X	X
Water and sanitation program			X			X	X

Notes: Effects of program exposure are the regression coefficients of young cohort dummy interacted with the number of schools constructed in region of birth. Standard errors clustered at region of birth are shown in parentheses. Stars denote statistical significance at 1, 5, and 10% levels based on regular p-values. Columns 1-3 uses Indonesia Family Life Survey data, Round 5 (2014/2015). Columns 4-6 uses IFLS round 5 data plus the observation from the last observed round for any individual not in round 5. Column 7 uses Susenas 2016 data that is restricted to the IFLS districts, which cover 83% of the Indonesian population.



Appendix Figure B.1. Map of Indonesia with districts shaded in gray indicating coverage in Indonesia Family Life Survey (IFLS)

Notes: IFLS survey is representative of 83% of the Indonesian population and covers individuals living in 13 of the 27 provinces in the country. The districts shaded in gray are included in the IFLS household survey, while the Susenas 2016 used in the main analysis in the paper is nationally representative and includes all districts in the country.

C. Cost-benefit Calculations Appendix

Discount rate

World Development Indicators collects real interest rates in Indonesia between 1987 and 2017. It averages 5.77 percent per year. Since it does not extend as far as our sample period in 1973, we assume a constant annual discount rate of 5 percent.

Teachers' salary growth

We first assume there is no real salary growth over the years and use Duflo (2001)'s reported teacher's salary in 1973. Subsequently, we allow for linear growth using teacher's salary observations in 1970 by Daroesman (1972), 1973 by Duflo (2001), Intercensal Surveys 1976 and 1995, and Labor Force Surveys 2000, 2005 and 2010.[3] Teachers are paid for the lifetime of the schools

Lifetime curvature

Individuals' tax payments and living standards generally follow an inverted-U shape, where they peak at around age 40-50. In our Susenas data, we observe individuals at their peak. To model the lifetime curvature of tax payments and living standards, we assume the same average effect on taxes and living standards across ages but different means at different ages. A 20-year old male, for instance, only spends $2,373 annually, compared to the mean in of our observed sample, $3,531, as implied in Table 3.

GDP/capita growth

GDP per-capita growth is obtained from World Development Indicators. We took the average between 1961 and 2017: 3.25 percent per year.

Number of students and teachers per school and recurrent costs/salaries multiplier

We follow Duflo (2001) in assuming 120 students/school, 3 teachers/school, and 25% recurrent administrative costs in addition to teachers' salaries. These imply a class size of 20 students across 6 grades of primary education and 1 teacher per grade. The latter is reasonable given that schools often run two sessions per day: morning and afternoon classes.

Individuals start paying taxes after age 18

We first assume that individuals start paying taxes after finishing Upper Secondary education at age 18. We subsequently relax this assumption to age 22, after individuals finish Tertiary education.

School lifetime

Daroesman (1971) and Duflo (2001) report that schools were expected to last for 20 years. We first use this assumption. We subsequently relax this assumption to 40 years because many INPRES schools are still in-use as of 2016.

Life expectancy

[3] We drop Duflo (2001)'s reported salary in 1995 because it implies a 9 percent real growth per year and it is much higher than the linear fit would have predicted. It is also higher than observations in 2000, 2005, and 2010.

World Development Indicators suggest an average of 56.6 years of life expectancy at birth for individuals born between 1968 and 1980. Conditional on making it to primary school age, the life expectancy is likely higher. We assume a life expectancy of 60 years throughout and then relax this assumption in the final column.

Share of men and women in affected cohorts

We construct a weighted average of the treatment effects on men and women. The share of women in the affected cohort is 0.498. For simplicity, we assume an equal share of men and women.