Enhancing Cognitive Functioning:
Medium-Term Effects of a Health and Family Planning Program in Matlab

WEB APPENDIX
Results Using Sampling Weights

Tania Bahram
June 3, 2010

This appendix presents tables using weighted regressions to account for the sampling weights. The table names and titles are the same as in the main paper. The weights have been top coded at the 90 percentile, and outliers with sampling weights greater than 42 (90^{th} percentile) are dropped. Outliers were determined based on the main regression model to determine program impacts in equation 1. Regression specific outliers have not been found for each of the robustness checks and models used. It is important to note that the sample size for some of the regressions are small for the 8-14 year old groups so may be more susceptible to bias due to weighting. In particular, regressions run on the random kid sample (e.g. Table 4 column 4), regressions where the 8-14 year old group is broken into two groups (Table 6 panel A), regressions restricted to just the treatment area (Table 6 panel B), or for the spillover effects when we split the control group into villages that are close versus further away from the treatment village. Below I explain why the weighted regressions may be biased and outliers are dropped, as well as changes made to the actual individual sampling weights. The tables of results are presented after these explanations starting on page 3.

There is concern that using weighted regressions for these analyses may lead to biased results. This occurs for two reasons. First, the sample size, particularly for the 8-14 year olds group, is not large and identification is by birth year due to the fixed effects. There are 492 observations for the 8-14 year old, and 63 observations on those born in 1987. The sample size is particularly small for the 8-14 year olds due to the data collection error where the cognitive functioning data was not collected on all children under the age of 15. So, this may well be a problem only for analysis of the MMSE and not for studies using other variables.

Second, the sampling weight has a large variance. The weight that accounts for both the individual and household sampling in my sample ranges from 2.5 to 464.5, the median is 10.3, the standard deviation 26, the 90 percentile is 4 times the median (42) and the 95 percentile 6 times the median (60). It is important to remember because birth year fixed-effects are included in the regression, identification comes from comparing treatment and comparison groups for each age, and thus a few large weights within an age group which is relatively small can easily bias the results. To help address this issue the weights have been top coded at the 90th percentile.

Since the results are more likely to be biased if outliers have large weights I examined the difference in the weighted regressions results with and without the outliers with large weights. To determine which observations are potential outliers, I run the main unweighted regression equation (equation 1 in the paper) without controlling for pre-intervention characteristics. Observations that were outliers according to both the leverage and Cook's D measures were

dropped if they had a sampling weight of 42 or higher. This lead to the exclusion of just 6 observations.

Comparing the different versions of the weighted regression it is clear weighted regressions are sensitive to outliers with high weights. In particular, the point estimate for the 8-14 year old group controlling for pre-intervention characteristics, drops from 0.39 without weighting to 0.22 when weights are included but not adjusted. It increases to 0.29 when outliers with large weights are dropped, 0.26 if the weight is top coded but no outliers dropped, and to 0.31 when high weights are dropped and the weight is top coded.

None of the point estimates for the specific age groups across the various weighted and unweighted regressions are significantly different from each other, however, the point estimates for the weighted regressions without adjustment are much smaller. The variation in the point estimates for the weighted regressions with the inclusion/exclusion of just 6 observations is concerning. It highlights how influential a few outlier observations with large weights can be and raises the question whether the results provided in the weighted regressions are being driven by a few outliers. While I have identified potential outlier/influential observations for the main regression, I am concerned that which observations are outliers/influential may vary for the other analyses as the data is cut up into slightly different groups.

Lastly, the individual-level sampling weights used in the weighted regressions are not those in the public dataset, they were remade by the author. This is because the individual-level sampling weights are corrected for non-response based on book 3 (questions for adults) and book 5 (questions for children). Individual tests, such as the cognitive test or anthropometrics, are in book 6. Book 6 was implemented after Book 3 and 6 by different enumerators. As a result, the non-response is different. In addition, non-response is different for cognitive functioning for the 8-14 year old group due to the data collection error. Using the individual sampling weights as given in the survey would result in a further lose of 60 observations when analyzing cognitive functioning since there was non-response in book 3 or book 5 but not book 6. The individual level weight I made is based on the sampling scheme but not adjusted for non-response.

TABLE 2: 1996 MHSS CHARACTERISTICS

	Treatment Area			Comparison Area			Difference in Means		
	Mean	SD	Obs	Mean	SD	Obs	Mean	T-stat	Mean/SD
MMSE score 25–49	25.8	(0.3)	1404	26.1	(0.2)	1595	-0.31	-0.78	-0.03
MMSE score 20–24	27.3	(0.4)	249	27.7	(0.3)	294	-0.44	-0.86	-0.07
MMSE score 15–19	26.6	(0.5)	281	27.8	(0.3)	355	-1.28	-2.42	-0.17
MMSE score 8–14	23.2	(0.7)	184	22.2	(0.6)	304	0.98	1.07	0.09
Height (cm) 25–49	154.5	(0.4)	1368	155.2	(0.3)	1558	-0.7	-1.6	-0.10
Height (cm) 20–24	155.4	(0.9)	244	155.6	(0.5)	284	-0.2	0.0	0.00
Height (cm) 15–19	150.0	(0.6)	321	150.0	(0.6)	393	0.0	0.0	0.00
Height (cm) 8–14	125.5	(0.6)	632	124.2	(0.5)	873	1.3	1.7	0.10
Years of education 25–49	2.9	(0.2)	1517	2.7	(0.2)	1659	0.1	0.7	0.00
Years of education 20–24	5.8	(0.3)	274	5.3	(0.3)	311	0.5	1.3	0.10
Years of education 15–19	4.5	(0.2)	363	4.9	(0.2)	446	-0.4	-1.5	-0.10
Years of education 8–14	2.1	(0.1)	748	1.8	(0.1)	967	0.3	2.6	0.10
Age	27.7	(0.4)	2118	26.6	(0.3)	2548	1.12	2.1	0.06
Female (=1)	0.55	(0.01)	2118	0.53	(0.01)	2548	0.03	1.45	0.05
Hindu (=1)	0.16	(0.04)	2118	0.06	(0.02)	2548	0.1	2.45	0.07

Notes: Standard deviations (SD) are clustered at the village level. The standard deviation of the whole sample is used to create the ratio of the difference in means to standard deviation. The cognitive function sample is used to calculate age, female, and hindu.

TABLE 3: 1974 BASELINE CHARACTERISTICS

	Treatment Area			Comparison Area			Difference in Means		
	Mean	SD	Obs	Mean	SD	Obs	Mean	T-stat	Mean/SD
Panel A: Full Sample									
Family size	6.97	(0.15)	2118	6.95	(0.10)	2548	0.02	0.11	0.00
Owns a lamp (=1)	0.62	(0.03)	2118	0.62	(0.02)	2548	0.00	-0.04	0.00
Owns a watch (=1)	0.15	(0.02)	2118	0.15	(0.01)	2548	0.00	-0.08	0.00
Owns a radio (=1)	0.08	(0.01)	2118	0.08	(0.01)	2548	0.00	0.24	0.01
Wall tin or tinmix (=1)	0.32	(0.02)	2118	0.31	(0.02)	2548	0.00	0.14	0.00
Tin roof (=1)	0.82	(0.02)	2118	0.84	(0.01)	2548	-0.02	-1.09	-0.03
Latrine (=1)	0.81	(0.02)	2118	0.86	(0.02)	2548	-0.05	-1.58	-0.04
Number of rooms per capita	0.21	(0.00)	2118	0.21	(0.00)	2548	0.00	0.63	0.02
Number of cows	1.49	(0.08)	2118	1.48	(0.08)	2548	0.01	0.11	0.00
Number of boats	0.64	(0.04)	2118	0.7	(0.03)	2548	-0.06	-1.12	-0.03
Drinking water, tubewell (=1)	0.32	(0.03)	2118	0.16	(0.02)	2548	0.16	4.2	0.11
Drinking water, tank (=1)	0.38	(0.04)	2118	0.34	(0.03)	2548	0.04	0.72	0.02
Drinking water, other (=1)	0.3	(0.05)	2118	0.5	(0.04)	2548	-0.20	-3.41	-0.09
HH age	47.2	(0.51)	2118	46.8	(0.61)	2548	0.40	0.5	0.01
HH years of education (edu.)	2.39	(0.12)	2118	2.3	(0.11)	2548	0.10	0.64	0.02
HH works in agriculture (=1)	0.61	(0.02)	2118	0.6	(0.03)	2548	0.01	0.26	0.01
HH works in fishing (=1)	0.07	(0.02)	2118	0.06	(0.01)	2548	0.00	0.17	0.00
HH spouse's age	36.6	(0.40)	2118	36.5	(0.46)	2548	0.10	0.17	0.01
HH spouse's years of edu.	1.07	(0.06)	2118	1.18	0.06	2548	-0.11	-1.32	-0.04
Panel B: Age 8–14									
Family size	6.33	(0.25)	184	7.00	(0.25)	304	-0.67	-1.92	-0.16
Owns a lamp (=1)	0.59	(0.05)	184	0.61	(0.04)	304	-0.03	-0.41	-0.04
Owns a watch (=1)	0.11	(0.04)	184	0.13	(0.03)	304	-0.02	-0.45	-0.04
Owns a radio (=1)	0.07	(0.04)	184	0.07	(0.02)	304	0.00	0.04	0.00
Wall tin or tinmix (=1)	0.28	(0.06)	184	0.30	(0.03)	304	-0.02	-0.34	-0.03
Tin roof (=1)	0.79	(0.04)	184	0.91	(0.03)	304	-0.12	-2.60	-0.23
Latrine (=1)	0.85	(0.04)	184	0.87	(0.03)	304	-0.01	-0.23	-0.02
Number of rooms per capita	0.21	(0.01)	184	0.21	(0.01)	304	0.00	0.49	0.04
Number of cows	1.39	(0.21)	184	1.55	(0.17)	304	-0.16	-0.60	-0.06
Number of boats	0.58	(0.08)	184	0.71	(0.07)	304	-0.12	-1.15	-0.09
Drinking water, tubewell (=1)	0.26	(0.05)	184	0.16	(0.04)	304	0.10	1.59	0.14
Drinking water, tank (=1)	0.38	(0.06)	184	0.30	(0.05)	304	0.08	0.95	0.08
Drinking water, other (=1)	0.36	(0.08)	184	0.54	(0.05)	304	-0.18	-1.88	-0.16
HH age	49.2	(1.22)	184	47.7	(1.43)	304	1.56	0.83	0.07
HH years of education	1.43	(0.27)	184	2.22	(0.22)	304	-0.79	-2.26	-0.19
HH works in agriculture (=1)	0.66	(0.04)	184	0.55	(0.05)	304	0.10	1.52	0.13
HH works in fishing (=1)	0.06	(0.03)	184	0.06	(0.02)	304	0.01	0.28	0.02
HH spouse's age	37.9	(1.10)	184	37.0	(1.18)	304	0.92	0.57	0.05
HH spouse's years of edu.	0.83	(0.08)	184	1.26	(0.13)	304	-0.43	-2.80	-0.21

Notes: Standard deviations (SD) are clustered at the village level.

TABLE 4: INTENT-TO-TREAT PROGRAM EFFECTS FOR THE MMSE Z-SCORE BY AGE GROUP

	Double Difference OLS						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment Area (=1)	-0.02	-0.02		0.00	-0.02	-0.04	-0.04
	(0.06)	(0.06)		(0.12)	(0.06)	(0.06)	(0.05)
Treatment Area*(Age 8–14)	0.27+	0.31*	0.24+	0.18	0.33*	0.32*	0.28+
	(0.15)	(0.14)	(0.14)	(0.24)	(0.14)	(0.14)	(0.15)
Treatment Area*(Age 15–19)	-0.20+	-0.21*	-0.23*	0.21	-0.21*	-0.23*	-0.16
	(0.10)	(0.10)	(0.09)	(0.22)	(0.10)	(0.11)	(0.09)
Treatment Area*(Age 20–24)	-0.04	-0.06	-0.08	-0.19	-0.06	-0.08	-0.08
	(0.08)	(0.08)	(0.07)	(0.24)	(0.08)	(0.09)	(0.07)
Treatment Area*(Age 8–14) * Tubewell drinking water in 1974					-0.13		
					(0.34)		
Mother ever eligible for FP (=1)						0.05	
						(0.05)	
Individual characteristics	Y	Y	Y	Y	Y	Y	Y
Preintervention characteristics	N	Y	Y	Y	Y	Y	Y
Village fixed-effects	N	N	Y	N	N	N	N
Random child sample	N	N	N	Y	N	N	N
Education fixed-effects	N	N	N	N	N	N	Y
Observations	4,666	4,666	4,666	652	4,666	4,089	4,618
Adjusted R-Squared	0.14	0.20	0.26	0.27	0.20	0.21	0.27

Notes: Standard errors are clustered at the village level. "**" ,"*", or "+" indicates that the difference in the coefficient from zero is statistically significant at the 1 percent, 5 percent, or 10 percent significance level respectively. Individual characteristics include year of birth fixed-effects and controls for gender and religion. Preintervention characteristics include all variables in Table 3. OLS=Ordinary least square, FP = family planning.

TABLE 5: SPILLOVER EFFECTS ON MMSE Z-SCORE

	(1)	(2)
Comparison Area	0.02	0.02
	(0.06)	(0.06)
Comparison Area*(Age 8–14)	-0.38*	-0.36**
	(0.15)	(0.15)
Comparison Area*(Age 15–19)	0.21*	0.22*
	(0.10)	(0.10)
Comparison Area*(Age 20–24)	0.05	0.09
	(0.08)	(0.08)
Comparison Area*Border treatment village	0.00	
	(0.09)	
Comparison Area*(Age 8–14)*Border treatment village	0.27	
	(0.20)	
Comparison Area*(Age15–19)*Border treatment village	-0.01	
	(0.14)	
Comparison Area*(Age 20–24)*Border treatment village	0.02	
	(0.13)	
Comparison Area*Border treatment village - closest quartile		0.00
		(0.11)
Comparison Area*(Age8–14)*Border treatment village - closest quartile		0.47
		(0.30)
Comparison Area*(Age15–19)*Border treatment village - closest quartile		-0.11
		(0.20)
Comparison Area*(Age20–24)*Border treatment village - closest quartile		-0.31**
		(0.11)
Observations	4,666	4,666
Adjusted R-Squared	0.20	0.20

Notes: Standard errors are clustered at the village level. "**" ,"*", or "+" indicates that the difference in the coefficient from zero is statistically significant at the 1 percent, 5 percent, or 10 percent significance level respectively. All regressions include year of birth fixed-effects, controls for gender and religion, and preintervention characteristics from Table 3.

TABLE 6: INTENT-TO-TREAT EFFECTS ON MMSE Z-SCORE
DISAGGREGATED IN THE TREATMENT AREA

Panel A: Full Sample		Panel B: Treatment Area Only Sample	
Treatment Area	-0.02	Treatment Area 1	-0.08
	(0.06)		(0.08)
Treatment Area*(Age 8–11)	0.42+	Treatment Area 1*(Age 8–11)	0.16
	(0.22)		(0.24)
Treatment Area 1*(Age 12–14)	0.25	Treatment Area 1*(Age 12–14)	0.15
	(0.19)		(0.20)
Treatment Area 2*(Age 12–14)	0.18	Treatment Area 1*(Age 15–19)	-0.24+
	(0.15)		(0.14)
Treatment Area*(Age 15–19)	-0.21*	Treatment Area 1*(Age 20–24)	-0.14
	(0.10)		(0.12)
Treatment Area*(Age 20–24)	-0.06		
	(0.08)		
Observations	4,666	Observations	2,118
Adjusted R-Squared	0.20	Adjusted R-Squared	0.19

Notes: Standard errors are clustered at the village level. "**" ,"*", and "+" indicate that the difference in the coefficient from zero is statistically significant at the 1 percent, 5 percent, and 10 percent significance level respectively. All regressions include year of birth fixed-effects, controls for gender and religion, and preintervention characteristics from Table 3.

TABLE 7: EFFECTS BY SUBCOMPONENT OF MMSE (Z-SCORES)

	Orientation	Attention-Concentration	Recall	Registration	Language
Treatment Area	-0.03	0.03	-0.05	-0.07	0.09
	(0.05)	(0.04)	(0.07)	(0.07)	(0.07)
Treatment Area*(Age 8–14)	0.22	0.15	0.13	0.35*	0.27+
	(0.14)	(0.14)	(0.15)	(0.17)	(0.16)
Treatment Area*(Age 15–19)	-0.19+	-0.16	-0.16	-0.21+	-0.01
	(0.10)	(0.09)	(0.11)	(0.11)	(0.10)
Treatment Area*(Age 20–24)	0.02	-0.11	-0.02	-0.04	-0.09
	(0.09)	(0.07)	(0.12)	(0.08)	(0.09)
Observations	4,666	4,666	4,666	4,666	4,666
Adjusted R-Squared	0.23	0.20	0.07	0.10	0.05

Notes: Standard errors are clustered at the village level. "**" ,"*", or "+" indicates that the difference in the coefficient from zero is statistically significant at the 1 percent, 5 percent, or 10 percent significance levels respectively. All regressions include year of birth fixed-effects, controls for gender and religion, and preintervention characteristics from Table 3.

TABLE 8: TREATMENT-ON-THE-TREATED EFFECTS

	First Stage Equations			Second Stage Equations	
	Received 2 plus child health intervention	Received 4 child health intervention	Mother ever used modern contraception	MMSE Z-Score	
	(1)	(2)	(3)	(4)	(5)
Endogenous variables					
Received 2 or more preventative child health interventions (=1)				0.31+	
				(0.19)	
Received 4 preventative child health interventions (=1)					0.51+
					(0.31)
Mother ever used modern contraception (=1)				-0.14	-0.13
				(0.19)	(0.19)
Instruments					
Eligible*(Age 8-14) (=1)	0.84**	0.50**	-0.01		
	(0.04)	(0.04)	(0.05)		
Mother Eligible for MCH-FP (=1)	0.02	0.01	0.28**		
	(0.00)	(0.00)	(0.02)		
F-statistic on excluded instruments	255	79	70		
Observations	3752	3752	3752	3752	3752

Notes: Standard errors are clustered at the village level. "**" ,"*", or "+" indicates that the difference in the coefficient from zero is statistically significant at the 1 percent, 5 percent, or 10 percent significance levels respectively. All regressions include year of birth fixed-effects, controls for gender and religion, and preintervention characteristics from Table 3.

TABLE 9: INTENT-TO-TREAT EFFECTS ON HEIGHT AND EDUCATIONAL ATTAINMENT Z-SCORES

	Double-Difference OLS Height Z-score		Single-Difference OLS Education Attainment Z-score	
	(1)	(2)	(3)	(4)
Treatment Area (=1)	-0.11	-0.05		
	(0.09)	(0.06)		
Treatment Area*(Age 8–14)	0.25*	0.18	0.25**	0.40**
	(0.10)	(0.12)	(0.06)	(0.12)
Treatment Area*(Age 15–19)	0.08	-0.03	-0.21*	-0.14
	(0.13)	(0.11)	(0.10)	(0.10)
Treatment Area*(Age 20–24)	-0.01	-0.00	0.10	0.11
	(0.14)	(0.10)	(0.10)	(0.09)
Treatment Area*(Age 25–49)			0.05	0.05
			(0.05)	(0.05)
MMSE Sample	N	Y	N	Y
Observations	5,725	4,462	6,342	4,557
Adjusted R-Squared	0.05	0.02	0.24	0.25

Notes: Standard errors are clustered at the village level. "**" ,"*", or "+" indicates that the difference in the coefficient from zero is statistically significant at the 1 percent, 5 percent, or 10 percent significance levels respectively. All regressions include year of birth fixed-effects, controls for gender and religion, and preintervention characteristics from Table 3.