Institute for Fiscal Studies

Household Responses to Information on Child Nutrition: Experimental Evidence from Malawi

Emla Fitzsimons, Bansi Malde, Alice Mesnard, Marcos Vera-Hernandez[1]

PEPA Inference Workshop, October 9 2013

[1] We gratefully acknowledge funding from ESRC-Hewlett 183-25-0008 and the ESRC Centre for the Microeconomic Analysis of Public Policy

Motivation

- Households make health and non-health related choices based on their knowledge of how health is produced (Gronau 1972, Rosenzweig and Schultz 1983)
- Incorrect knowledge of the health production function leads to inefficient choices and consequently sub-optimal levels of health
- Establishing empirically the consequences of deficiencies of knowledge is difficult, because of the endogeneity of knowledge:
  - (Unobserved) parental preferences or health endowments may drive both accumulation of knowledge and health input choices

This Paper

- Overcome this challenge by exploiting a cluster randomised control trial in rural Malawi which provided mothers with information on child nutrition
  - Trained local women visited mothers living in randomly chosen clusters and provided information and advice on infant feeding
- Uses a simple theoretical model to show how households' choices on both health and non-health dimensions should adjust to improved knowledge
- Special attention is paid to inference since the trial includes 24 clusters only
  - Use 2 leading methods for inference in such designs:
    - Wild cluster bootstrap-t (Cameron et al. 2008)
    - Randomisation inference (Fisher 1935, Rosenbaum 2002)
  - Investigate how both of these methods perform in our data

Related Literature

- Effects of information on health
  - Improved health: Madajewicz et al (2007), Jalan and Somanathan (2008), Dupas (2011)
  - No Effect: Kremer and Miguel (2007), Luo et al (2012)
- Effects of nutrition information on child health practices and child health
  - RCTs: Morrow et al (1999) for Mexico, Haider et al (2000) for Bangladesh
  - Non-experimental studies: Alderman (2007), Galasso and Umapathi (2009), Linnemayr and Alderman (2011)
- Consider both health and non-health margins of household behaviours

Setting: Mchinji (Malawi)

- Child health is very poor in Malawi
  - Infant mortality rate of 133 per 1000 births
  - 48% of kids aged < 5 years are stunted
- Misperceptions on child nutrition are widespread
  - Common to give porridge with unsterilized water to infants as young as 1 week
  - Widespread belief that eggs are harmful to 9-month old infants
  - Common belief that children should be given the broth in which vegetables/meat are cooked, instead of the vegetables/meat themselves

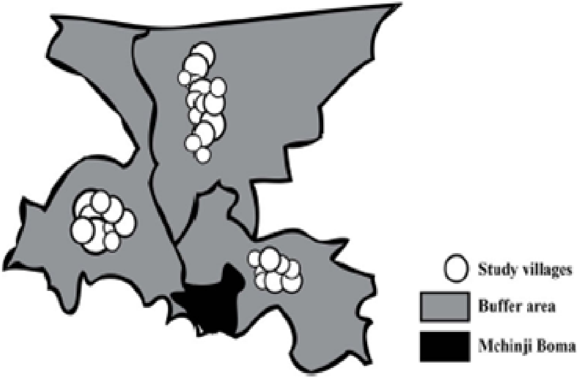
The Intervention

- Set up in 2005 by Mai Mwana, a research and development project that aims to improve maternal and child health
- Trained local women ("peer counsellors") provide information and advice on infant feeding to mothers of babies aged < 6 months
    - 5 visits: once before birth, 4 times after birth
    - Peer counsellors cover a population of around 1000 individuals
    - Visit content focused on exclusive breastfeeding and post-breastfeeding nutrition
- All pregnant women in a cluster are eligible for the intervention, but in practice around 60% are visited by the peer counsellors
- Intervention began in July 2005 and is still on-going

Experimental Design

- Mchinji District divided into 48 clusters, each with a population of around 8000 individuals
- Within each cluster, the villages closest to the geographical centre of the cluster were chosen to be part of the study area (approx. 3000 individuals)
- Creates a natural buffer area, limiting contamination between neighbouring clusters
- 12 clusters randomly chosen to receive the intervention, 12 clusters serve as controls
- Remaining 24 clusters received another intervention - women's groups

Experimental Design



Data

- Baseline census of all women aged 10-49 years in the study areas conducted by Mai Mwana, pre-intervention
  - Limited number of socio-economic variables
  - Sampling frame for follow-up data
- Follow-up data collected in 2008-09 and 2009-10
  - Random sample of 104 women drawn from each cluster from the baseline census, regardless of their fertility
  - Target sample of 2496 women

Data

- Succeeded in interviewing 2/3 of the 2496 women
    - Final sample of 1660 women and their households
    - Robustness checks show that all results are robust to the encountered attrition
- Attrition rates of around 9% between the two follow-ups

Sample Balance

	Full Sample			Interviewed Sample		
	Control Group	Diff: Treatment - Control	p-value	Control Group	Diff: Treatment - Control	p-value
Woman's Characteristics						
Married (dv = 1)	0.615	-0.021	0.386	0.661	-0.034	0.184
Some Primary Schooling or Higher	0.707	0.033	0.402	0.682	0.040	0.340
Some Secondary Schooling or Higher	0.066	0.010	0.535	0.060	-0.007	0.545
Age (years)	24.571	-0.180	0.637	25.492	-0.429	0.376
Chewa	0.948	-0.044	0.330	0.957	-0.050	0.246
Christian	0.977	0.006	0.476	0.979	0.008	0.336
Farmer	0.661	-0.075	0.108	0.688	-0.060	0.128
Student	0.236	0.015	0.438	0.204	0.022	0.274
Small Business/Rural Artisan	0.036	0.030	0.129	0.037	0.024	0.220
Household Characteristics						
Agricultural household	0.995	-0.005	0.471	0.995	0.002	0.591
Main Flooring Material: Dirt, sand or du	0.913	-0.041	0.232	0.916	-0.027	0.474
Main roofing Material: Natural Material	0.853	-0.018	0.697	0.857	-0.004	0.891
HH Members Work on Own Agricultura	0.942	-0.057	0.124	0.950	-0.056	0.120
Piped water	0.011	0.040	0.314	0.009	0.032	0.340
Traditional pit toilet (dv = 1)	0.772	0.054	0.218	0.791	0.054	0.182
# of hh members	5.771	0.066	0.817	5.848	0.132	0.863
# of sleeping rooms	2.116	0.199	0.038*	2.152	0.166	0.128

Model

Simple model where 1-parent 1-child households choose child consumption, adult consumption and adult leisure to maximise household utility

Providing the parent with information on child nutrition will:

1. Increase child consumption
2. Increase adult labour supply (assuming that leisure and adult consumption are complements, or have a limited degree of substitutability)
3. Reduce adult consumption
4. Increase total household consumption

Empirical Model

$$Y_{ict} = \alpha + \beta_1 T_c + X_{ict}\beta_2 + Z_{c0}\beta_3 + \mu_t + u_{ict}$$

- $T_c = 1$ if main respondent in the follow-up survey resided in a treated cluster in 2004
    - Identify an Intention-to-Treat (ITT) parameter
- X_{ict} are individual level covariates; Z_{c0} cluster-level baseline covariates
- Pool data from both follow-up surveys in our estimation

Inference

- Must consider that observations are not independent within clusters
- Huber-White clustered standard error estimates downward-biased in samples with small numbers of clusters
- Use two leading inference methods for such designs
  - Wild cluster bootstrap-t procedure recommended by Cameron, Gelbach and Miller (2008)
  - Randomisation Inference (Fisher 1935; Rosenbaum 2002)

Inference

- Wild cluster bootstrap-t procedure covered in previous presentations
- Randomisation inference is non-parametric and exploits the randomisation to conduct inference
  - Tests a sharp null hypothesis of no effect for any unit in the data, rather than a zero average intention-to-treat effect
  - Permutes the randomisation allocation Method
- Report p-values computed using both methods
- Conduct a Monte Carlo experiment to compare how methods perform in our data

Multiple Outcomes

- Interested in testing the effects of the intervention on 6 domains: health knowledge, child consumption, household consumption, labor supply, child growth and child morbidity
- For each of these domains, we have multiple measures → almost 30 outcomes in total
- Concerns about multiple inference
  - The probability of rejecting a test is increasing in the number of tests carried out
- Deal with this problem by reducing the number of tests we carry out
- Aggregate multiple outcome measures in a domain into a summary index following Anderson (2008)
- Testing a summary index provides a test for whether the intervention has a "general effect" on each domain
  - But, magnitude of effect cannot be assessed
  - So, show outcomes for domains with statistically significant effects

Index

Results

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]
	Increases in Main Respondent Knowledge	Improvements in Child Food Consumption		Increases in Household Food Consumption	Increases in labor Supply		Improvements in Child Physical Growth		Reductions in Child Morbidity	
		< 6 months	> 6 months		Adult Males	Adult Females	< 6 months	> 6 months	< 6 months	> 6 months
T_c	0.169+	0.250*	0.143+	0.218*	0.262+	0.018	0.066	0.102*	0.058	-0.013
Standard Error	[0.086]	[0.098]	[0.074]	[0.082]	[0.131]	[0.165]	[0.056]	[0.036]	[0.070]	[0.102]
Wild Cluster Bootstrap p-value	{0.058}	{0.016}	{0.076}	{0.018}	{0.086}	{0.955}	{0.293}	{0.022}	{0.438}	{0.861}
Randomization Inference p-value	{0.065}	{0.028}	{0.099}	{0.037}	{0.062}	{0.903}	{0.366}	{0.035}	{0.509}	{0.920}
Observations	1512	151	1280	3200	3642	4138	312	2175	376	2356
R-squared	0.107	0.214	0.099	0.063	0.183	0.136	0.062	0.026	0.059	0.053
IntraCluster Correlation	0.169	0.041	0.085	0.087	0.146	0.140	0.019	0.021	0.021	0.150
Mean Control Areas	-0.040	-0.109	-0.054	-0.099	-0.135	-0.050	0.245	0.266	-0.034	0.022

Knowledge

Child Consumption

	[1]	[2]	[3]
	Summary Index	Water	Milk other than maternal
		< 6 months	
T_z	0.250*	-0.144+	-0.082*
Standard Error	[0.098]	[0.081]	[0.034]
Wild Cluster Bootstrap p-value	{0.016}	{0.106}	{0.020}
Randomization Inference p-value	{0.028}	{0.077}	{0.112}
Observations	151	359	151
R-squared	0.214	0.249	0.087
IntraCluster Correlation	0.0405	0.024	0.060
Mean, Control	-0.109	0.488	0.101

Child Consumption II

	Summary Index	Any beans	Any meat	Any fish	Any eggs	Any vegetables	Any fruit	Any nsima	Any porridge
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
T_z	0.143+	0.225**	0.091	0.007	0.026	-0.010	-0.011	0.025	0.094
Standard Error	[0.074]	[0.056]	[0.096]	[0.098]	[0.052]	[0.020]	[0.057]	[0.015]	[0.064]
Wild Cluster Bootstrap p-value	{0.076}	{0.006}	{0.563}	{0.927}	{0.637}	{0.643}	{0.825}	{0.134}	{0.246}
Randomization Inference p-value	{0.099}	{0.007}	{0.279}	{0.947}	{0.624}	{0.627}	{0.869}	{0.142}	{0.261}
Observations	1280	1288	1287	1289	1288	1,291	1,288	1,290	1,294
R-squared	0.10	0.07	0.02	0.01	0.011	0.141	0.153	0.143	0.035
IntraCluster Correlation	0.085	0.113	0.085	0.111	0.0502	0.0181	0.0923	0	0.136
Mean, Control	-0.054	0.258	0.291	0.463	0.164	0.959	0.700	0.930	0.800

Household Consumption

	[1]	[2]	[3]	[4]	[5]	[6]
		Per Capita Monthly Food Consumption for:				
	Food Index	Cereals	Proteins	Fruit and Vegetables	Other Foods	Total Non-durable
T_z	0.218*	-9.878	128.359*	269.819+	60.453	526.601*
Standard Error	[0.082]	[52.450]	[54.798]	[108.600]	[33.561]	[193.981]
Wild Cluster Bootstrap p-value	{0.018}	{0.931}	{0.022}	{0.060}	{0.150}	{0.038}
Randomization Inference p-value	{0.037}	{0.952}	{0.016}	{0.042}	{0.020}	{0.006}
Observations	3200	3205	3202	3204	3204	3190
R-squared	0.063	0.118	0.02	0.195	0.024	0.06
IntraCluster Correlation	0.087	0.074	0.042	0.172	0.053	0.095
Mean Control Areas	-0.10	606.00	349.80	679.70	149.70	2146.00

Male Labor Supply

	Male Adults			
	[1]	[2]	[3]	[4]
	Summary Index	Works	Has at least 2 jobs	Weekly Hours Worked
T_z	0.262+	0.096	0.072*	4.31
Standard Error	[0.131]	[0.078]	[0.028]	[2.918]
Wild Cluster Bootstrap p-value	{0.074}	{0.303}	{0.020}	{0.230}
Randomization Inference p-value	{0.062}	{0.251}	{0.057}	{0.202}
Observations	3642	3961	3958	3642
R-squared	0.183	0.17	0.05	0.16
IntraCluster Correlation	0.146	0.208	0.036	0.100
Mean, Control	-0.135	0.836	0.122	25.740

Child Physical Growth

	Age > 6 months			
	Summary Index	Height for Age	Healthy weight for age	Healthy weight for height
T_z	0.102*	0.271*	0.030	0.048
Standard Error	[0.036]	[0.102]	[0.019]	[0.027]
Wild Cluster Bootstrap p-value	{0.022}	{0.022}	{0.150}	{0.132}
Randomization Inference p-value	{0.035}	{0.055}	{0.312}	{0.147}
Observations	2175	2192	2265	2217
R-squared	0.026	0.046	0.024	0.029
IntraCluster Correlation	0.021	0.022	0.018	0.017
Average, Control	0.266	-2.338	0.817	0.845

Comparing Inference Methods

- Compare the performance of the various inference methods in our data using Monte Carlo simulations
- Do so for 9 different data generating processes (DGPs), corresponding to our outcomes of interest
  - Each DGP uses the same sample and covariates as our main results
- Each DGP includes a cluster level random effect and an i.i.d. individual error (i.e. homoskedastic std errors)
- The cluster level random effect is constructed such that the intra-cluster correlation in the simulated data matches that in the actual data

Monte Carlo Evidence

Test size ↓	[1] Increases in Main Respondent Knowledge	[2] Improvements in Child Food Intake	[3]	[4] Increases in Household Food Consumption	[5] Increases in Adult Labour Supply	[6]	[7] Improvements in Child Physical Growth	[8]	[9] Reductions in Child Morbidity	[10]
		< 6 months	> 6 months		Males	Females	< 6 months	> 6 months	< 6 months	> 6 months
Method										
Huber-White Clustered Std Errors	0.093	0.088	0.072	0.078	0.081	0.085	n/a	0.086	0.108	0.084
Wild Cluster Bootstrap-t	0.048	0.061	0.061	0.065	0.055	0.047	n/a	0.07	0.073	0.051
Randomization Inference	0.039	0.046	0.052	0.034	0.05	0.041	n/a	0.047	0.037	0.047
IntraCluster Correlation in Data	0.169	0.041	0.033	0.087	0.146	0.140	0.019[a]	0.020	0.021	0.150

Robustness

- Rule out a number of alternative explanations
  - Adult health improvements
  - Fertility
  - Other aspects of the intervention
  - Attrition

Conclusion

- Use variation induced by a randomised control trial to show that improved knowledge of the child health production function influences a broad range of household behaviours
- Health and Non-health behaviours including male labour supply
- Pay careful attention to the important issue of inference
  - Use two leading methods for inference with small number of clusters
  - Monte Carlo experiments to assess their performance in our data
  - Both methods provide similar inference, though randomisation inference has a slight tendency to over-reject the null hypothesis.

Randomisation Inference

- Step 1: Take covariates into account (Small et al (2008)):

$$Y_{ict} = \alpha + X_{ict}\beta_2 + Z_{c0}\beta_3 + \mu_t + \epsilon_{ict}$$

Predict the residuals, $\hat{\epsilon}_{ict}$

- Step 2: Specify the test statistic

$$\sum_{c:T=1} \frac{\hat{\epsilon}_{ict}}{N_1} - \sum_{c:T=0} \frac{\hat{\epsilon}_{ict}}{N_0}$$

- Step 3: Calculate the distribution for the test statistic by permuting the randomisation across clusters
- Step 4: Calculate exact p-value by computing the proportion of test statistic values in the distribution that are greater than the test statistic calculated under the true randomisation
  - In practice, given large set of possible permutations (>2.7m in our case), use 100,000 randomly selected permutations to generate the test statistic distribution.

Aggregating across outcomes

- First, standardise outcomes to have a 0 mean and std deviation of 1
- Re-define outcomes so that a higher value implies a better outcome
- Summary index is calculated as a weighted mean of the standardised outcome values within each domain
  - Weights are calculated to maximise the amount of information captured by the index
  - Less weight is given to highly correlated outcomes
  - Boosts efficiency

Back

Nutritional Knowledge

	Summary Index	Breastfeeding when infant has diarrhoea	Are biscuits or groundnuts/soya more nutritious for kids aged 6 months-3 yrs?	From what age should solid foods be given infants?	How should an HIV positive woman feed her baby?	Is nsima or porridge more nutritious for an infant aged > 6 months?	What is the best way of cooking fish with porridge for an infant aged > 6 months?	Should eggs be given to an infant aged > 9 months?
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
T_z	0.169+	0.253+	-0.052	0.037	0.138	-0.101	0.067**	0.104
Standard Error	[0.086]	[0.115]	[0.041]	[0.026]	[0.150]	[0.078]	[0.019]	[0.069]
Wild Cluster Bootstrap p-value	{0.058}	{0.084}	{0.290}	{0.166}	{0.444}	{0.210}	{0.002}	{0.186}
Randomization Inference p-value	{0.065}	{0.028}	{0.222}	{0.292}	{0.399}	{0.179}	{0.008}	{0.192}
Observations	1512	1512	1512	1512	1512	1512	1512	1512
R-squared	0.11	0.10	0.05	0.04	0.04	0.07	0.04	0.02
IntraCluster Correlation	0.169	0.277	0.082	0.049	0.408	0.183	0.057	0.107
Mean, Control	-0.04	0.217	0.938	0.88	0.393	0.857	0.026	0.719

Back