Effects of a Fertility and Health Intervention on Household Income: A Longitudinal Analysis Using Corrective Weights

Andrew Foster and Sveta Milusheva

October 15, 2014

Abstract

We examine the long term effects of a well-known maternal and child health and family planning program in a rural area of Bangladesh on household income. While the population economics of this program have been examined elsewhere, none of the studies to date has fully recognized the implications for the analysis of program effects of the gap between the introduction of the program in 1978 and the collection of the first round of detailed economic data as part of a representative survey in 1996. We argue in particular that there is a need to account for the process of household formation and recombination. Indeed, this problem arises quite commonly in observational studies of programs because baseline surveys as well as information on subsequent household composition changes of the study population are not typically available. We are able to address these issues in this particular population because we have access to a complete registry of migration and household residence changes carried out in the study area since 1974. Specifically, we develop specialized weights using a novel resampling procedure. Our analysis uses those weights in order to examine the effect of treatment on household income changes between 1974 and 1996. We show a positive impact of treatment on income for the lowest income third of the population in 1974 and no effect on the middle and high income thirds. We also demonstrate the effect of the weighting procedure on results, underscoring the importance for long-term evaluations of development interventions to consider the process of household formation and recombination.

1 Introduction

The increased availability of household surveys with a longitudinal component has had a major impact on the design and implementation of empirical analysis in population research. The ability to examine changes in individual and household behavior over time makes it possible to better control for fixed individual attributes as well as to relate decisions at one point in time to outcomes at another point in time. The fact that many such data sets now span multiple generations makes it possible, at least in principle, to look at sources of long term inequality and mobility. However, in trying to assess these broader macro-processes, issues of assignment of sample weights, which are sometimes ignored in micro-based studies, take on added significance.

The role of weights to adjust for sample attrition has achieved substantial attention in the literature (Fitzgerald, Gottschalk, and Moffitt 1998, Moffit, Fitzgerald, and Gottschalk 1999) . It is understood that for the most part there are not simple solutions, particularly in cases where the attrition is not random with respect to processes of interest in the data. Less attention has been given to the question of weighting in a setting in which the problem is not sample attrition but the process of household formation and recombination. In this case observability may not be as much of an issue as in the case of attrition, but the flow of people across households and the fact that the sampling unit for a survey is generally the household creates a new set of problems.

To understand this point, consider a random sample of households collected at time t and assume that larger (at time t) households are more likely to divide. If all descendant households are followed, then one will correctly measure at time t+1 the distribution of t+1 attributes such as household size. However, if these households are used to retrospectively construct the mean household size at time t, then the estimate will be too large because large t households are overrepresented in the sample of t+1 households. A similar bias would arise if household size at time t was estimated from a random sample of households in period t+1. A simple correction in each case would be to inverse weight each household by the number, if available, of co-descendant households with the same antecedent household.

The situation is complicated in the presence of joining or recombining households. The set of all descendant households in t+1 generated from an initial sample in period t will no longer yield an unbiased estimate of average household size at time t+1 because households composed of members of multiple antecedent households will be more likely to be selected than would be the case if t+1 households were selected randomly. To correct this bias one would need to know about the number of co-descendants of each antecedent household for a particular descendant household, inclusive of those households that were not part of the initial period t sample. Similarly, constructing an estimate of the average size of households at period t from a period t+1 sample is complicated because the set of antecedent households at time t is not necessarily representative of time t households.

This is not just a curiosity. Surveys are at times used to evaluate the consequences of interventions that were introduced at a previous period, and in some cases retrospective or previously collected data are incorporated into the analysis in order to establish differential change over time. In such cases we may ask when, if ever, it is possible in such cases to mimic the results of a randomized trial in which a baseline is collected from a random sample at a particular point in time, a set of treatments is assigned to the participants, and then outcomes are evaluated at some endline. Our answer is a tentative yes, but as our application suggests, the data requirements for doing so are extremely demanding.

In this paper, we tackle these larger questions of sample selection bias and weighting mechanisms in panel datasets by looking at a specific example of panel data where all of these issues arise. We devise a weighting scheme that mitigates these problems and show how the weights help to improve results in certain situations. We focus on the Matlab Health and Socio-economic Survey done in Matlab, Bangladesh in 1996, which can be linked to a Health and Demographic Surveillance System (HDSS) in Matlab that has been ongoing since 1966, allowing for a panel structure to the data. There are dozens of papers and dissertations using this data, many of which use it to examine the effects of a Maternal and Child Health and Family Planning Program begun in 1978. This combination of an HDSS with an intervention or randomized control trial (RCT) has been used in various other regions, including Navrongo (Ghana), Rakai (Uganda) and Filabavi (Vietnam) among others. These provide the perfect set-up to study the long-run effects of the interventions as the populations are traced over time, but it is not always clear how to make the link between the sample and the HDSS. Our paper demonstrates how in the case of an RCT or other intervention, it is possible to make use of the long-run HDSS data by creating weights that help to study the long-term impacts.

There are two specific issues with the dataset we focus on: first,the sample studied in

1996 is not representative of the pre-intervention population, making the study of the 1978 intervention inaccurate, and second, due to selection bias in which descendants are picked to be in the sample, intergenerational analyses in particular are skewed. We attempt to correct for both the representative nature of the 1996 sample and the ability to conduct intergenerational analyses by devising weights that take both of these problems into account. We create weights to make the sample representative of 1974 using a unique resampling procedure that is possible due to the nature of HDSS data, which in the case of Matlab included data on the full 1974 population and all descendants in 1993. This procedure is compared to the standard procedure of using propensity score weights. We then take on the task of creating weights that allow for intergenerational analysis and explore when it is appropriate to use those weights.

We examine how the weights can help with one of the most basic questions for any intervention–what are the effects on income for the participants? Although this question has been examined in Joshi and Schultz 2007, we study the heterogeneous effects of treatment based on income in 1974. Specifically, we are interested in examining the effect of the program on the correlation between households' income in 1974 and their descendants' income in 1996. There are two main problems with studying this question that the longitudinal analysis using our weights helps to resolve. First, given that the 1996 sample is not necessarily representative of the 1978 population, if, for example, richer households tended to split and have more descendants in different baris, then we have an over-representation of wealthy households. The effect of the program measured would only be for the richer subpopulation that might be less likely to see an income effect from the program, leading to an underestimate of the program impact. Second, when doing an intergenerational analysis, we are interested in the effect of the program on descendants. The problem is that the sampling strategy in 1996 based on the way related families choose to group and live together affects the probability of being picked for the sample. If this probability is correlated with certain characteristics of 1996 households, such as income, this would affect how representative the 1996 sample is of the descendants. We elaborate on this in section II.

We use our corrective weights to look at the effect of the program on income. As a measure of income we focus on consumption, which tends to be less influenced by temporary shocks (Hall and Mishkin 1982). Using a crude measure of consumption in 1974, we divide the 1974 population into thirds based on this measure. We find that those households in the bottom third have higher consumption in 1996 as compared to households in the bottom third in the control area. There is no such difference between treatment and control households for the middle income or high income groups. This implies that the bottom of the distribution gained the most from the treatment. In addition, using our weights leads to bigger coefficients as compared to using the original 1996 weights, implying that without the weights the effect of the program would be underestimated. This finding underscores the importance of considering the weights that should be used when conducting longitudinal analyses, and evaluating bias that could arise from a lack of appropriate weights.

The paper proceeds by describing the data in the following section and laying out the problems with how the sampling procedure could bias outcomes if it is not corrected. In section 3, we describe the process we used to create weights for 1974 and compare our weights to propensity score weights. We then go into the creation of the 1996 weights in section 4. In section 5 we present the results of using our weights to analyze the effect of the program on income in the intergenerational context, and we conclude in section 6.

2 Background and Data

The International Centre for Diarrhoeal Disease Research (ICDDR,B) in the Matlab region of Bangladesh began to maintain a Health and Demographic Surveillance System registering all births, deaths and migrations starting in 1966. There is data available for the full period on 149 villages, which include over 200,000 people. In 1974 the ICDDR,B conducted the first comprehensive census of the region. Censuses were again conducted in 1982 and 1993. From this census data we have information on every single household in the region including basic demographic information and some information on assets.

Along with collecting detailed demographic data, the ICDDR,B initiated a Maternal and Child Health and Family Planning Program (MCH/FP) intervention in 1977. This intervention was implemented in 70 of the villages, and as part of it, women of childbearing age received doorstep delivery of contraceptives and antenatal care, children received in-home vaccinations delivery, and there were increased services for the prevention and management of childhood diarrheal and acute respiratory illnesses (Fauveau 1994, Phillips et al. 1982). The family planning and maternal health portion of the intervention was first rolled out in 1977. The services were then expanded in 1982 to include vaccinating children under 5 for measles, DPT, polio and tuberculosis, as well as providing vitamin A supplementation. Such services were not provided by the government to the rest of the region until 1988.

For this intervention, contiguous villages were grouped together into treatment and control areas, partly to control for contamination and spillover effects between treated and not treated villages. Nevertheless, it is a case-control study where treatment and control areas were reasonably matched on observables, though the matching was not perfect. Much of the research on Matlab using ICDDR,B data focuses on the outcomes of this intervention.

Matlab Health and Socio-economic Survey

In addition to the censuses that were collected, in 1996 a Health and Socio-economic Survey (MHSS) was conducted in Matlab. This survey collected detailed economic and social data on a sample of the population, which had not been done before. The goal was to use this data in order to look at the effect of the MCH/FP intervention on a wide range of outcomes and over a long period (Rahman et al. 1999).

The MHSS sample was selected from the population in 1993. In 1993, there were 38,489 households split among 7,440 baris, clusters of households in close physical proximity that are usually linked in a kin network. Of those, 2,883 baris were randomly picked to be part of the sample. Within each bari, one household was randomly selected for the detailed interview. Sampling was done at the bari level because it provides a better representation of family networks as compared to sampling households. There were 102 baris that no longer existed in 1996, and therefore the final number of baris sampled was 2,781. A second household in each bari was also interviewed, but this was not done randomly, so most researchers conduct analyses using only the first household, and we also focus on this primary sample.

The detailed level of the data collected by the MHSS on income, consumption, education, health, and other outcomes has allowed researchers to examine the long run impact of the 1978 MCH/FP program. There are dozens of papers and dissertations that use the data to look at various outcomes (Joshi and Schultz 2013, Field and Ambrus 2008, Maitra 2003, Barham 2012). In order to calculate the effect of the treatment on the whole population, these papers use the sampling weights included in the dataset. As we will describe in the following subsection, certain aspects of the sampling procedure could bias the results depending on the questions of interest. In particular, it is not necessarily appropriate to use the weights given in the dataset when looking at economic mobility and intergenerational questions.

Problems with the Sampling and Sample Weights

For a researcher trying to evaluate program effects, two problems arise with using the data, and specifically using the sampling weights that are included in the MHSS. These relate to how the sampling was conducted, and the fact that households were picked based on the 1993 bari configurations and not based on the pre-treatment 1974 population. Because households were picked based on 1993 configurations, the inverse probability weights available in the dataset make the sample representative of the 1993 population. But in order to be able to evaluate the effectiveness of a program on different groups, or the results from an experiment, one wants the sample picked post intervention to be representative of the population before the intervention. The sampling also does not account for systematic household formation and recombination correlated with household attributes.

The main problem comes from the fact that when sampling descendants, one needs to account for the fact that different original households will have different numbers of descendants. If one household has ten descendants while a second has only two descendants and the sample is chosen by randomly picking several households from the twelve descendants, the first household has a much higher chance of being represented in the sample. If the number of descendants that a household has is random, this would not be a problem. The issue is that certain household characteristics can lead to a household having more descendants, so the number of descendants is not random. Therefore, rather than getting the effect for the population, we get it for a certain subset of the population that is more likely to have more children. In addition to this issue, if the characteristics are correlated with outcomes, then the probability of being in the sample is correlated with outcomes, leading to biased results.

One example of how this could manifest itself is if wealthier households in 1974 were more likely to have more children. In that case, in the 1993 population there would be more descendants from wealthier households. Suppose a surveyor randomly selects households from 1993; there is a higher probability that these households will have come from wealthier antecedents in 1974. If the treatment differentially affected poorer and wealthier households, for example if wealthier households got more out of the treatment and had better outcomes than poorer households, then using the 1996 sample would overstate the effect of the treatment.

In order for the sample to be representative of the population pre-treatment, it would have been necessary to randomly select baris from the 1974 population. Then we could pick one household per bari, trace which households descended from that household in 1993 and randomly pick a descendant per 1974 household. Instead, by randomly selecting descendants from the 1993 population for the 1996 sample without taking into account how many of them came from each 1974 household, the MHSS team exposed the sample to the potential bias that has been outlined. [1] By sampling baris rather than households, the bias could be mitigated because baris tend to be made up of households that are linked by kinship. Yet, women are likely to join the bari of their husband upon marriage, so a 1974 household with several daughters would have descendants in several baris. In addition, the decision of some descendants to split and form their own bari or to join a different bari could also be dependent on the observable and unobservable characteristics of the 1974 household. This dynamic would again affect the probability of a 1974 household being represented in the sample.

The second problem with using the data without corrective weights is related to this last issue of the splitting of households and the bari structure of the sample. Because baris were the unit that was randomly sampled, and only one household was picked from each bari, if two descendants from the same 1974 household were in the same bari, they would never both be picked to be in the sample. If, on the other hand, two descendants from the same 1974 household were in different baris, then it is possible that both could be picked for the sample, and even more so if they are in small or single household baris. We already mentioned how this could affect the representativeness of the 1974 population if the characteristics of 1974 households are correlated with the decision of their descendants to stay in the same bari or split off. In addition, if the decision to stay in the same bari as other descendants, split off into a new bari or join a different bari is correlated with the characteristics of the 1993 households, then it is not possible to accurately estimate average descendant outcomes, distorting the results of intergenerational analyses. This is especially a problem in evaluating the effect of the treatment because the treatment directly affects fertility and the number of descendants, which could lead to different patterns of household formation and recombination that could affect the probability of being selected into the sample. Both Foster 1993 and Foster and Rosenzweig 2002 examine some of the effects of household divisions and the importance of factoring those in when looking at longitudinal data.

[1] One of the coauthors was on the original MHSS team and now recognizes the issues with the way the sampling was done, but at the time, the focus was on getting a representation of kin networks, which were assumed to be manifested in the bari structure, without considering the endogeneity of how kin networks might spread to other baris due to the formation and recombination of households.

To illustrate the problem, suppose a 1974 household has three descendants and we are interested in the effect of the intervention on the income of descendants from this household. We need an accurate estimate of the average income for the descendants of the household, but in most cases all three descendants will not be in the sample. Now suppose that certain attributes determine whether households remain in the same bari or split off. For example, it is possible that the poorest descendant household of a family might choose to split off and look for better opportunities in a different location, while the two richer households remain in the same bari because they are already well off and would not want to leave their land, assets, network, etc. If this behavior were systematic in the population, it would mean that two richer descendants would never both be in the sample, instead there would tend to be a richer and a poorer descendant. Thus, if the average income of descendants for particular households is calculated by taking the arithmetic average of the descendants that show up in the sample, then the sample will consistently underestimate the true average income of descendants. If the treatment led to an increase in the income of descendants, the treatment area would be impacted by this underestimate more than the non treatment area, leading to an underestimate of the program effect on income.

If the mechanism behind household formation and recombination is random, so that the probability of getting any combination of households with certain attributes is equally possible, then there should be no bias, and the arithmetic average will be the average effect on descendants. This seems unlikely though, given that certain attributes such as wealth have been shown to play some role in household formation and recombination. Thus, it is important to attempt to correct for the bias that might arise if simple means are used to calculate average descendant outcomes.

3 Correcting for the Representativeness of the Sample

This section will describe the procedure used to correct for the fact that the 1996 sample is not representative of the 1974 population. We will then compare the results to a different common method used to create weights that corrects for sampling issues, which is propensity score weighting.

Re-weighting Procedure

The basic method used to calculate new weights that are representative of 1974 is to create a representative sample of 1974 households linked to the 1996 sample and assign weights based on the probability that a particular 1974 household is represented in the sample. To understand the procedure, it is important to understand how "antecedent" and a "descendant" are defined. The 1974 households that made up the original population of Matlab that was surveyed are considered the antecedents. A 1993 household is considered a descendant if at least one member in the household satisfies one of four criteria:

1. Any individual that was a member of a 1974 household and is still in the population in 1993

2. The spouse of any individual who was a member of a 1974 household

3. Any child, grandchild, great-grandchild, and so forth of an individual who was a member of a 1974 household

4. The spouse of a child, grandchild, or great-grandchild of a member of a 1974 household

Given that there is data on the full population of 1974 households and the Demographic Surveillance System provides data on all births, deaths, marriages and migrations, including a unique ID number that allows individuals to be traced over time, we can link 1974 households with all of their descendants in 1993. These antecedent-descendant links can be used in order to calculate the probability that a particular 1974 antecedent household has a descendant that appears in the 1996 sample.

Calculating these probabilities theoretically is complicated because it is not possible to sum the probabilities that each of the descendants of a particular antecedent is picked, because if two descendants are in the same bari, both will never be picked in the same sample. Instead, we calculate the probabilities using the procedure implemented in 1996 to pick the sample, and conduct it 1000 times in order to find the probability that out of 1000 samples a particular 1974 household is represented.

To elaborate, we took the 1993 population and randomly picked 2,781 baris from the total 7,440 baris, and then picked one household at random from each bari. This creates a sample of 2,781 households. This process was done 1000 times to create

1000 different samples. The antecedent-descendant links were used to establish which 1974 households were represented by at least one descendant household in each sample. The probability of a 1974 household being represented in the 1996 sample is the number of samples in which the household has at least one descendant out of 1000 possible samples.[2] We created probability weights by taking the inverse of the calculated probability and assigning that as the weight to each 1974 household.[3]

Having assigned each 1974 household a weight, we created a sample of 1974 households that is linked to the actual 1996 sample by taking all of the antecedents of the 1996 sample and grouping them together into what we are calling the "1974 sample." This consists of 6,044 households that all have at least one descendant in the 1996 sample. Using the 1974 probability weights calculated earlier, we then can get a representation of the full 1974 population.

Table 1 shows the mean value for a number of variables in 1974 for the full population as well as for the sample both weighted and unweighted. For all the variables, the weighted sample is representative of the full population. The unweighted sample, on the other hand, has significantly different means for every variable except highest education in the household. This implies that as expected, the 1996 sample is not linked to a representative set of 1974 households, and instead certain types of households were more likely to be represented in the 1996 sample. The unweighted sample has a higher average family size, which seems intuitive because a household with more family members is likely to have more descendants. On average, the households also have more cows and more rooms, both indicative of higher wealth. It seems that the 1996 sample is indeed representative of a distinct set of 1974 households, that among other things are wealthier on average than the population, but using the weights calculated, we are able to make the sample representative of the 1974 population.

[2]Note that we also conducted this analysis by doing 100,000 samples instead of only 1000, and the results were very similar. We are presenting everything here for the set of 1000 samples

[3]There were 3202 households in 1974 that did not have any descendants in 1993. In this current paper we only focus on the 1974 households which have a descendant in 1993 because we cannot follow up those 3202 households, although it is possible to examine and compare their characteristics with those of the households that do have descendants in order to determine whether their omission causes a bias.

	(1)	(2)	(3)	(4)
	All of 1974 Hholds	Unweighted Sample	Weighted Sample	Weight vs No weight (p value)
Highest Edu in the Household	4.14	4.17	4.07	0.223
Number of Cows	1.16	1.38	1.13	0.000
Edu of the Household Head	2.27	2.09	2.25	0.090
Age of Head of Household	45.7	46.3	45.3	0.003
Family Size	6.12	6.83	6.05	0.000
Num of Rooms in the Household	1.22	1.29	1.20	0.000
Number of Observations	24,779	6,044	6,044	
Total Weight	24,779	6,044	25,516	

Table 1: Comparison of Population with Weighted and Unweighted Sample

Now a researcher looking at the effect of the treatment on certain outcomes can use the linked sample of 1974 antecedent households in order to see what happens to the descendants of treated versus not treated households. By applying the weights, it is possible to obtain results that are representative for the full 1974 population. This is especially helpful when interested in heterogeneous effects of the treatment.

Comparison to Propensity Score Weighting

Another common method for calculating weights to correct for nonrandom sampling is to create a propensity score weight. The idea behind this is to use observable characteristics to calculate the probability that someone is picked to be in the sample.

We calculate propensity score weights for the 1974 population and compare them to the weights assigned using the sampling method. In order to do this, we take our sample of 1974 households linked to descendants in the 1996 sample and assign them a value of one for being linked, while all other 1974 households get a value of zero. This variable is used as the dependent variable in a logit regression. The controls used are observable characteristics of the 1974 households including highest education of anyone in the household, number of cows, number of boats, education of the head of household, the age of the head of household, the size of the family and the number of rooms in the household dwelling.[4] The coefficients

[4] We also conducted the propensity score analysis with all of these variables as well as including the number of descendants. In a regular propensity score analysis this variable would not be

from the regression are used to calculate predicted values for each 1974 household, which are equivalent to the probability that a certain 1974 household is likely to be linked to the 1996 sample based on their observable characteristics. The weight is the inverse of this predicted probability.

Table 2 shows the results of the propensity score weighting and how it compares to the sample weights from the previous section, the scenario with no weights and the actual population means. The top panel shows the means for the population and the different weighting strategies. Both our sampling weights and the propensity weights come very close to approximating the true population means. In the bottom panel we present p-values for a comparison between the population means and the differently weighted samples. As in Table 1, not using weights leads to means that are all statistically different from the population means except in the case of highest education in the household. Our weights, in comparison, yield means that are not significantly different from the population means. The propensity weights are also not significantly different from the population means except in the case of family size. Although the propensity weights are fairly representative of the full population, the difference in family size is worrisome because it could mean there are other unobservables that are also significantly different from the population averages. Therefore, our weights yield the most representative sample weighting structure.

---

available, but given that we have a full census in 1993, we have it and tried using it to see if the accuracy of the propensity score weights increased. There is no significant difference between the weights using the number of descendants variable and those not using it, so we only show the results for the propensity score weighting procedure where we do not include number of descendants.

	Mean Values for 1974 Population and Weighted Samples			
	Actual Full Data	No Weight	Our Resampling Weight	Propensity Score Weight
Highest Edu in Household	4.14	4.17	4.07	4.19
Number of Cows	1.16	1.38	1.13	1.18
Edu of Head of Household	2.27	2.09	2.25	2.29
Age of Head of Household	45.7	46.3	45.3	45.7
Family Size	6.12	6.83	6.05	6.21
Num of Rooms in Household	1.22	1.29	1.20	1.22
Number of Observations	24,779	6,044	6,044	6,044
Total Weight	24,779	6,044	25,516	24,577

	P Values for Difference between Actual Data and Weighted Data		
	No Weight	Our Resampling Weight	Propensity Score Weight
Highest Edu in Household	0.481	0.422	0.37
Number of Cows	0.000	0.369	0.419
Edu of Head of Household	0.000	0.708	0.591
Age of Head of Household	0.000	0.272	0.943
Family Size	0.000	0.315	0.010
Num of Rooms in Household	0.000	0.055	0.637

Table 2: Propensity Score Weights vs Sampling Weights

4 Correcting for Bias in Descendant Selection

We now turn to the second problem that is posed by the structure of the sampling framework. If the probability that a certain descendant or combination of descendants connected to a 1974 antecedent gets selected is correlated with attributes of descendant households in 1996, this could lead to a bias in the calculation of average characteristics for households descended from the same antecedent. This was illustrated by our example of the 1974 household with three descendants where only one rich and one poor descendant will ever show up in the sample because the two rich ones are in the same bari and so can never both be in the sample. If this is true of all 1974 households with three descendants, then our results would estimate that descendants have a lower average income than they actually do. This arises because the probability of being picked is correlated with income (lower income households have a higher probability of being picked on average than higher income households).

To understand the problem more clearly, imagine the following scenario. Suppose there are two descendants from a 1974 household with incomes y_1 and y_2. Theoretically we could see just y_1, just y_2 or both in the sample. In this example, household y_1 is never picked alone and there is a .5 chance of picking both y_1 and y_2 and a .5 chance of picking just y_2.[5] If we were to take the average of the households if they show up together and take the value of b when we only have b, then we get the following expected income:

$$\begin{aligned}\mathbb{E}(y) &= 0 * y_1 + \frac{1}{2}\left(\frac{y_1 + y_2}{2}\right) + \frac{1}{2}y_2 \qquad (1)\\ \mathbb{E}(y) &= \frac{1}{4}y_1 + \frac{3}{4}y_2 \neq \frac{1}{2}y_1 + \frac{1}{2}y_2 = \bar{y}\end{aligned}$$

In expectation, we are not getting the average for the descendants. If the probabilities were random, then with a large enough sample, this inconsistency would average out. The problem arises if the probability is correlated with attributes of the 1996 households. For example, if the data consisted of two descendants for every antecedent and the probability of being picked is correlated with income so that y_2 is always the poorer antecedent and y_1 is always the richer one, this would result in a lower estimate of average income for descendants.

There are various ways to tackle this problem. One simple way is that in this case, we could take y_2 if we only have y_2 and only take y_1 if we pick both y_1 and y_2, which would give us an expectation equal to the average. Yet in doing that, information on descendant y_2 would be thrown out if both descendants are in the sample. In addition, this is a solution for this particular set of probabilities. There are also more complicated probabilities in the data where there is a probability of seeing all three combinations, but the expectation still does not equal the average of the descendants.

We attempt to correct for the potential bias by finding weights that can work for any probabilities. There are many ways to come up with these weights, but our goal

[5] This scenario might seem unlikely, but its purpose is to illustrate the more complicated case of several descendants where some live in the same bari and therefore the probability of picking two descendants living in the same bari is 0. Doing the example with more descendants makes the calculations messier and detracts from the point of simply illustrating why it is important to consider how descendants are weighted.

is to come up with simple ones that resolve the potential bias if the probabilities are correlated with the characteristics of the descendants. We do not claim that these are the "best" weights that can be used, but they are simple, straight-forward, and we show that they can be an improvement over not using weights. In this section, we first describe how the weights were created. Then, we show how the weights compare to not using weights and discuss in what situations the weights are proper to use and when they could actually further distort the results.

Creating the Weights

We want to create weights that are based on the probability of a certain combination of descendants appearing in a sample. As discussed, a variety of weights could be created that attempt to do this. Our weights are based on two particular criteria. Considering the paper's focus on changes in income over generations, the weighting procedure and discussion are centered on income, though they are just as applicable to other intergenerational characteristics of interest. In addition, most of the examples and illustrations will involve only two descendant households for the sake of simplicity, but the application to a higher number of descendants is analogous.

The first criteria we want to meet is that any weights should lead to an outcome where the expected value for descendants' income is equal to the actual mean of the incomes. The way this would look in the case of two descendants with incomes y_1 and y_2 and probability p_a of picking just household 1 in the sample, probability p_b of picking both household 1 and household 2, and probability p_c of picking only household 2 in the sample is as follows:

$$\mathbb{E}(y|p_a, p_b, p_c) = p_a w_a y_1 + p_b(w_{b1} y_1 + w_{b2} y_2) + p_c w_c y_2 = \frac{1}{2} y_1 + \frac{1}{2} y_2 = \bar{y} \qquad (2)$$

where w_a, w_{b1}, w_{b2}, and w_c are the weights.

There are many different possible ways of weighting the observations in order to get an expected value for descendants equal to the actual average. Some of these weights might lead to a large variance in the estimates, depending on the probabilities. Therefore, in addition to getting the correct mean income in expectation using our weights, we also want to minimize the variance from different probabilities of descendant combinations being picked for different antecedents. We want to minimize the following:

$$
\begin{aligned}
Z = [&var(p_a)(w_a^2 y_1^2) + var(p_b)(w_{b1}y_1 + w_{b2}y_2)^2 \\
&+var(p_c)(w_c^2 y_2^2) - 2cov(p_a, p_b)(w_a y_1)(w_{b1}y_1 + w_{b2}y_2) \\
&-2cov(p_a, p_c)(w_a y_1)(w_c y_2) - 2cov(p_b, p_c)(w_{b1}y_1 + w_{b2}y_2)(w_c y_2)]
\end{aligned} \tag{3}
$$

Both equation 2 and equation 3 depend on the actual mean and variance of the income values. We do not have all of the income values, which is why we are devising these weights in the first place. Therefore, the weights must work more generally and not be sensitive to the mean and variance of income. Again, this could be achieved in different ways. A sufficient condition for equation 2 to hold is that it holds for all incomes, which can be ensured by taking derivatives with respect to the income values y_1 and y_2. In our simple two descendant example, this yields the following two equations, both of which need to hold in order for our first condition to be met:

$$
\begin{aligned}
p_a w_a + p_b w_{b1} &= \frac{1}{2} \\
p_b w_{b2} + p_c w_c &= \frac{1}{2}
\end{aligned} \tag{4}
$$

We apply a similar logic to our second condition and take second derivatives with respect to y_1 and y_2 in order to come up with the following objective function:

$$
\underset{w_a, w_{b1}, w_{b2}, w_c}{min} \frac{d^2 Z}{dy_1^2} + \frac{d^2 Z}{dy_2^2} \tag{5}
$$

In this way we are trying to ensure that the variation in the fraction of households in each sample has a small impact on the computed average income for descendants because the weights are applicable and minimize variance no matter what the actual incomes are. We calculate the weights by minimizing equation 5 conditional on equations 4. As mentioned, there are other, more complicated, criterion functions that we could have used, but we believe that this simple one still allows us to find weights that help to mitigate the potential bias arising from the bari structure. We will show how our weights using this procedure compare to not using weights in the following section and that indeed they can help to address the potential bias.

Solving the minimization problem, we find that the weights which minimize the variance of the estimates and in expectation yield the true income are based on the probability of sampling a 1996 household. The combination in which a household appears (whether a descendant appears alone in the sample or if there are several other descendants in the sample from the same antecedent) does not affect the weight. This is surprising because as we saw in our example in equation 1, the combination of descendants in a bari affects the expected value we get. Yet in trying to minimize the variance in a manner general enough to apply to all income values, the combination in which the descendants appear is no longer important. Nevertheless, the weight not only depends on the probability of being picked in 1996, but also on the total number of descendants. With the number of descendants in the denominator of the weight, those households who come from an antecedent with many descendants receive a smaller weight. Finally, the probability of the 1974 household being represented in 1996 also factors in because there is a chance that there is no descendant in the sample at all. Therefore, for our two descendant example we get:

$$w_a = w_{b1} = \frac{\Pr(Y)}{2 * \Pr(1)} = w_1$$
$$w_c = w_{b2} = \frac{\Pr(Y)}{2 * \Pr(2)} = w_2$$

Where Pr(Y) is the probability of the antecedent Y of the household being represented by a descendant in the 1996 sample. We can generalize this result to assign a weight to every descendant i of a 1974 household Y with X descendants:

$$w_i = \frac{\Pr(Y)}{X * \Pr(i)}$$

Robustness of the Weights

Without the incomes of descendants to help assign the weights directly based on our original two specifications, it was necessary for us to come up with weights that are generalizable no matter what the incomes might be. Given that there are a number of ways we could have devised the weights, it is important to show that using the weights improves estimates. To do this, we have conducted a simulation to demonstrate how the weights compare to not using weights.

The simulation is a simplified case of our data in order to focus on the effect of the weights when the probability of being picked is correlated with income, and how the performance of the weights depends on the extent of the correlation. We do not incorporate the sampling structure but instead look at how our weights perform in the case where we have 1000 antecedents and each has at least one descendant chosen for the sample. Therefore, here our Pr(Y) is equal to 1 because each antecedent has probability 1 of having a descendant in the sample. [6]

In the simulated data, each antecedent has exactly two descendants and each of their descendants has randomly been assigned a log income from a normal distribution with mean 8.58 and variance 1.15.[7] This mean and variance were chosen as they were the mean and variance of the actual log consumption variable for the 1996 sample.

We assigned probabilities for the following three events: household 1 is selected, household 2 is selected, both household 1 and household 2 are selected. The probabilities of being chosen are based on the random incomes using a logistic function in order to ensure a correlation between incomes and probabilities. We varied the size of that correlation by multiplying incomes in the logistic expression times a coefficient δ, which is manipulated. The same δ is used for all three

[6]Expanding this simulation to include the sampling structure that determines the probability that an antecedent household is selected does not change the results of how our weighting scheme compares to not using weights. This is because if we used the original sampling structure, we would then need to multiply times the 1974 weight of each antecedent in order to get the results for the population. But due to the fact that we had to create the 1974 weight based on the probability that an antecedent's descendant is in the 1996 sample, this is the same as Pr(Y), and so multiplying times the inverse of this, which is our 1974 weight, we have:

$$
\begin{aligned}
W_{final}(i) &= w_{1974} * \frac{\Pr(Y)}{X * \Pr(i)} \qquad (6) \\
w_{1974} &= \frac{1}{\Pr(Y)} \Longrightarrow \\
W_{final}(i) &= \frac{1}{\Pr(Y)} * \frac{\Pr(Y)}{X * \Pr(i)} = \frac{1}{X * \Pr(i)}
\end{aligned}
$$

Therefore, the final weight that is used is the same as our simplified weight in the simulation, so it is not necessary to complicate things by including the sampling structure.

[7]Although we only do the simulation with two descendants per antecedent household, the results are generalizable to more descendants and we have done some simulations including more than two descendants, but do not include the results here. We have also done simulations where we change the variance of the income variable, and this has also not affected the general result, so we omit those results.

probabilities. The three probabilities are:

$$
\begin{aligned}
Pr(1) &= \frac{e^{\delta * y_1}}{e^{\delta * y_1} + e^{\delta * y_2} + e^{\delta * \bar{y}}} \\
Pr(2) &= \frac{e^{\delta * y_2}}{e^{\delta * y_1} + e^{\delta * y_2} + e^{\delta * \bar{y}}} \\
Pr(1\&2) &= \frac{e^{\delta * \bar{y}}}{e^{\delta * y_1} + e^{\delta * y_2} + e^{\delta * \bar{y}}}
\end{aligned}
\tag{7}
$$

where $\bar{y}$ is the arithmetic mean of the two incomes. The coefficient delta varies from 0 to 1 in .01 intervals. A coefficient of 0 implies that each event has a one third probability of occurring irrespective of income, so there is zero correlation between incomes and probability. As the coefficient grows, the dependence between income and probability increases up to when the coefficient becomes 1, which gives the highest dependence between income and probability.

When the correlation is positive, if one household has a higher income than another, it will always have a higher probability of being selected alone, the probability of selecting both households will be next highest, and the probability of selecting the poorest household alone will be smallest. As the coefficient grows, this ordering of probabilities does not change, but the differences in probabilities become starker.

A sample of descendants is chosen based on the probabilities. One of the events is randomly chosen based on the probability of each event occurring.[8]Depending on the combination of descendants chosen for each antecedent, the weighted mean income is calculated based on the weights ($w_i = \frac{1}{2 * \Pr(i)}$). The mean income is also calculated with no weights, which entails taking the arithmetic mean if both descendants are in the sample, and taking the plain value of the descendant chosen if only one is in the sample. These two means are compared to the actual mean income for each descendant. Actual mean income is subtracted from the simulated mean income with and without weights and averaged to find the mean difference between actual and sample descendant income for the 1000 antecedents.

In order to make sure the simulation is robust to outliers in the events picked, a set of events was chosen 500 times. The average of the absolute mean difference and

[8]This is done by assigning each event a piece of the unit interval equal to its probability, and then randomly choosing a number on the unit interval which determines the event based on which piece the number falls into.

squared error was calculated for each sample. This analysis was done with 1000 samples having different income values (and thus probabilities). This procedure was done for each δ from 0 to 1 in .01 intervals.

Figure 1 shows the sample average difference between the income calculated with our weights and the actual income, as well as the sample average difference between the income calculated without weights and the actual income.This is graphed for various deltas which represent how dependent the probabilities are on income. A delta of 0 signifies that the probability is not dependent on income at all, and a delta of 1 signifies a high degree of dependence between incomes and the probabilities. The figure demonstrates higher variability in the average error when no weights are used versus when weights are used. Although the weights do not always perform better than not using weights, on average they are consistent in the level of error no matter what the correlation between income and probability, and this level is relatively low. The level of error ranges from almost none to almost .35 when no weights are used, while the error with the weights remains consistently under or close to .1.

If there is no correlation between income and the probability of certain combinations of households being picked, it would be better to use no weights. In that scenario, the mean descendant incomes calculated using no weights are very close to the actual mean incomes for the sample. When delta goes above .17, then the weights become better to use. This switch occurs for a relatively small delta, so if there is reason to believe a link exists between income and the probability of being chosen in the sample, then it is better to use the weights as compared to no weights.

The extent of the correlation can be tested using a Socioeconomic Status (SES) survey conducted in Matlab in 1996 that collected data from most of the households that were present in 1993. This survey has very limited data, but it does include several variables on household assets and household infrastructure which can be used to create an index as a proxy for household income. [9] The index is based on the importance of these assets and infrastructure for predicting income in the MHSS. Although the MHSS does not have an income variable, it provides data on consumption and the value of assets, which are regressed on the assets

[9] The variables used for the index are whether a household has a cow, boat, clock, or radio; whether it gets its drinking water from a tubewell; and whether the roof is made of tin.

and infrastructure variables in 1996 that are analogous to the 1996 SES ones.[10] The coefficients from these regressions give two sets of weights for how important different assets are in predicting the income of a household. These weights are applied to the asset and infrastructure variables in the 1996 SES survey and create an index based on consumption and an index based on value of assets. These measures of consumption and wealth, used as proxies for income, can be used to calculate the correlation between income and the probability of being chosen for the sample. The probability values come from the simulations done for the first set of weights. The log probability of being picked to be in the MHSS is regressed on the log income index to get an elasticity. Using the consumption based index gives an elasticity of 0.339 significant at the .01 level, and using the assets value based index gives an elasticity of 0.402 significant at the .01 level as well.[11]

To put this in the context of where these coefficients fall in terms of δ, similar regressions were run using the simulated data. Regressions of the probability of being picked on the simulated income were run for each δ between 0 and 1 at .01 intervals. Figure 2 plots how the coefficient grows linearly as δ increases. The horizontal dashed and dotted lines plot the coefficient values based on the regressions using the consumption index and the asset index respectively. The vertical dash and dot line marks the δ at which the no weight and the weight lines crossed in Figure 1. The coefficient from the consumption regression crosses a little below where it becomes better to use the weights, while the coefficient from the asset regression crosses above where it becomes better to use the weights. This makes it difficult to make a definitive statement on whether it is better to use the weights. Nevertheless, due to the use of an index to proxy for consumption and an index to proxy for the value of assets, there is a possibility of measurement error. Testing this using the actual consumption and value of assets in the 1996 MHSS and the indexes created, there is evidence of classical measurement error. Assuming this is true of the full population data, it would imply our coefficients are biased downward. [12] The

[10]The assets used to compute the value of assets are homestead land, ornaments (gold, silver) savings in bank, television, radio, clock, electric fan, cycle, and furniture.

[11]These are the values from running a regression using all of the descendant households of the 1974 households that are present in 1993 and that have data in the SES in 1996. Given that there are 2,599 households in 1993 that are not in 1996 (6 percent of households present in 1993), it means that there is not data on the full set of descendants for a number of 1974 households. The regressions were restricted to 1974 households for whom there was data on every one of their descendant households in 1993, and the results were very similar (.339 for consumption and .401 for assets value, both significant at the .01 level).

[12]Evidence of classical measurement error can be supplied by the authors upon request.

assets coefficient is already at a point that would imply that using the weights is better, and if the coefficients are biased downward, then it is likely the consumption coefficient would also be above the point marking indifference between using and not using weights. Therefore, in the case of the Matlab data, it makes sense to use the weights devised when conducting aggregate analyses.

Up to now, the discussion has focused on the performance of the weights on average for the sample. Except in the cases where there is very little correlation between income and the probability of being selected, the weights give a good approximation of the mean income of descendants on average for the sample. If we are interested in how they perform for individual households though, the average squared error is much bigger when using the weights as compared to not using weights. This is because as the sample gets bigger, the average income with the weights will be close to the expectation (in accordance with the law of large numbers), which due to the construction of the weights should be equal to the actual mean income. So with a sample of 1000, this holds true. But when looking at each individual antecedent and how the income calculated with the weights compares to the actual income, the squared difference is much bigger because the weights can cause some outliers. This is because chance means that sometimes even an event with a small probability will be picked, but that means it will have an extremely large and distorting weight.

Figure 3 shows how the weights compare to using no weights for individual antecedent households. The average squared error for an antecedent when not using weights is around 0.4, while the average squared error per antecedent starts out a little less than 9 and grows to almost 20 when using the weights. This is because, especially as the probabilities become more dependent on the income draws, there are more likely to be outliers with a very small or very large probability, which means a very large or small weight. Using such a large or small weight will give a more skewed average income than if no weight is applied. Nevertheless, even though for individual antecedents there is higher variability in the mean income calculated, looking at the whole sample, the very small incomes (due to very small weights) will be balanced by the antecedents that get very large incomes due to large weights, and in this way the average income for the sample will approach the actual average income.

What this means is that it is important for a researcher to think about the type of analysis he or she is running in order to determine whether using the weights

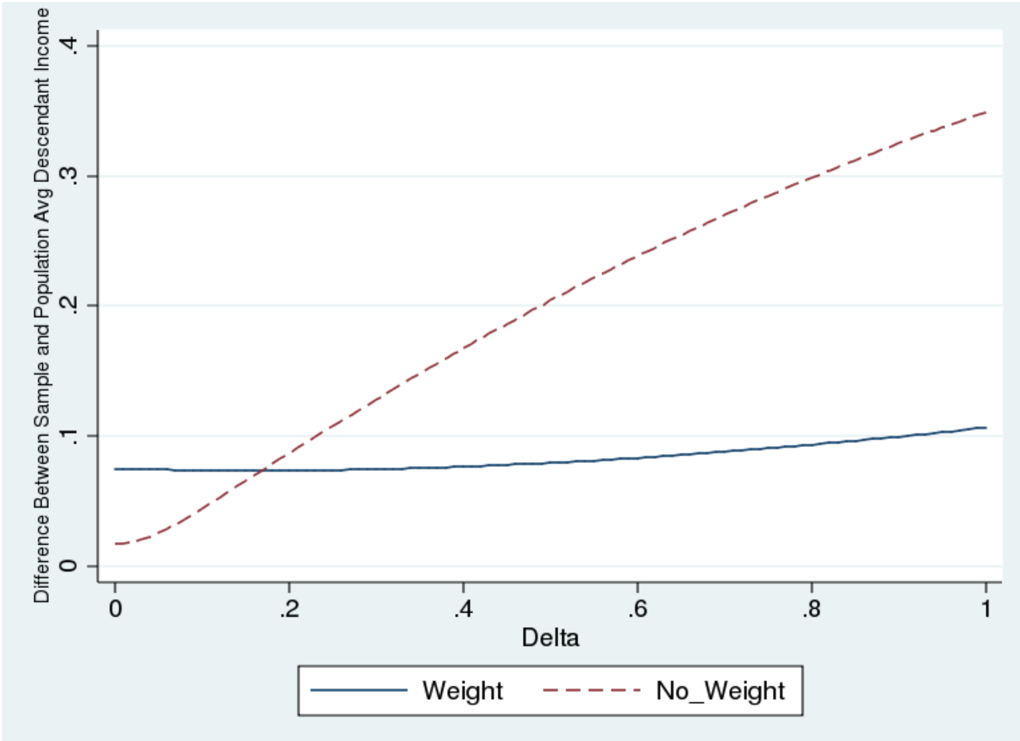

Figure 1: Average absolute difference between sample and actual descendant income means for different levels of correlation

is appropriate. In the case where one is interested in average effects, such as running regressions, using the weights would lead to more accurate results (if there is a link between the probabilities of being picked and the variable of interest). If, instead, one is interested in the effects on say certain quintiles of the population, which would involve using the average income to break people up into those quintiles, then the weights would distort the data. Therefore, it is important to be aware of the goal of any analysis and how using these weights might affect it in order to make sure that the weights are used correctly and are helping to improve the accuracy of the results rather than leading to additional bias or distortions[13].

[13] Although in the empirical section we break up the 1974 data into thirds based on income, we use the full population and not the weights to do this. We only use the weights in running the regressions

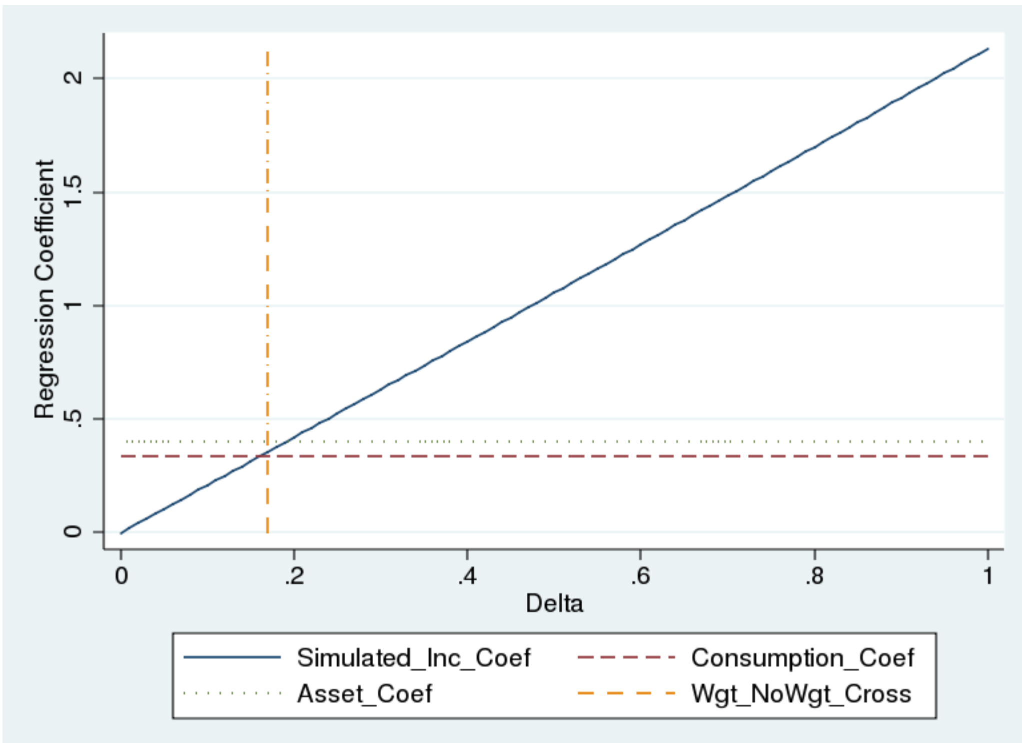

Figure 2: Coefficients of Probability Regressed on Income for the Simulated Data and the 1996 SES Survey Data

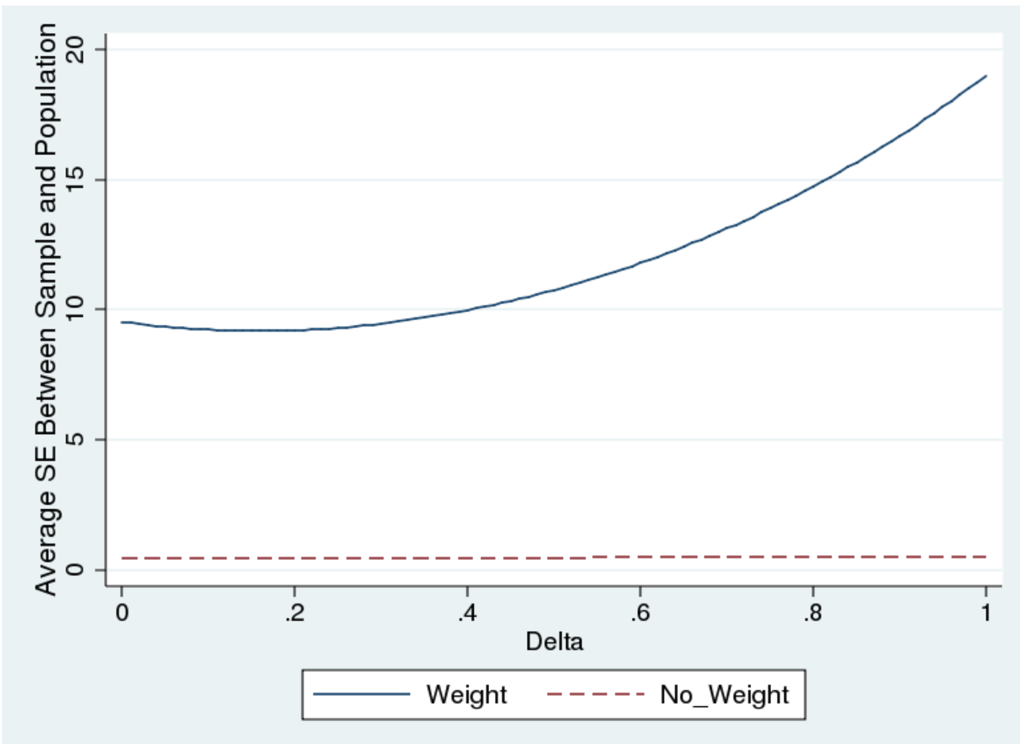

Figure 3:
Average squared difference between sample and actual descendant incomes

5 Household Income Analysis Using the Sample Weights

Having created these weights, we can use them to determine the long term effect of the Maternal and Child Health and Family Planning program on income, specifically looking at changes over time as we compare households in 1974 with their descendants in 1996. There are various channels through which the program could impact household income. The reduction in fertility and the change in household composition could lead to increased investment in human capital, which could increase income and wealth (Rosenzweig and Wolpin 1980, Becker and Lewis 1974). Improvements in child health could also increase cognitive ability and educational attainment, eventually leading to higher income(Case, Lubotsky, and Paxson 2001, Weil 2005, Smith 1999). Both of these channels would take longer to show results in measures of household income because it would be necessary to wait until the children affected by the health and fertility measures in the 1980s were at a wage earning age. Therefore, we might not see strong results from these channels looking at the 1996 data.

There could be more immediate effects on household income due to changes in marriage prospects. This would be the case if women in the treatment area are able to find higher income husbands due to their participation in the program or if poorer men in the treatment area receive a higher dowry because families want their daughters to move into the treatment area. For example, Arunachalam and Naidu 2008 show that the program caused an increase in bride-to-groom dowry transfers. We believe this increase in dowry affected women moving into the treatment area because families valued the MCH/FP program and were willing to pay more in order to have their daughters migrate to that area through marriage. This preference could positively impact the income of men in the treatment area. If there is similarly a value placed on marrying women who grew up in the MCH/FP area and were exposed to the health and fertility services, then they might have the ability to marry into better households with higher income. Both of these effects would manifest themselves into higher average income for the treatment area.

In addition, there could be heterogeneous effects of the treatment if certain income groups in 1974 experience a bigger effect from the treatment. For example, high income households might already have better marriage prospects, and the treatment might not affect them. Low income households in the treatment area, on the other hand, would have low marriage prospects, but the treatment could improve those substantially. In this way, the lower income households would experience an increase in income from the treatment and higher income households would not. Similarly, the intervention could have differential impacts on fertility based on 1974 income, which would translate into differential income effects.

The effect of the program on assets and income have been looked at by Joshi and Schultz in several of their papers, which analyze a number of long-term effects of the MCH/FP intervention (Joshi and Schultz 2007, Schultz 2009). Joshi and Schultz 2007 analyze women's income, but also look at household assets. They find that all but non agricultural assets are significantly larger in treatment areas than comparison areas, but that gains are generally concentrated among better educated older women. They describe this as being consistent with physical assets being a substitute for a decrease in children for better educated women. Asadullah 2012 looks specifically at household wealth mobility using the Matlab data and focusing on the relationship between fathers' and sons' wealth as well as changes in ones own wealth between 1974 and 1996 for men who were a head of household at both times. Asadullah finds limited intergenerational wealth mobility and suggests that this might be due to limited schooling mobility.

When analyzing income using just the 1996 data and weights, it can give an answer to the question of what is the effect of the treatment on income in 1993 because the sample is representative of the 1993 population. What we are unable to answer is how the treatment affects households that were in different parts of the income distribution prior to the intervention. This question is important when considering intergenerational income mobility and changes in income over time. Asadullah 2012 tackles this question, but does not consider the sampling bias and household formation and recombination, which we have corrected using our weights. If there is bias in the selection of the sample so that those selected are not representative of the population, then the treatment effect calculated is only for that selected group, which might experience a differential impact from the program. In addition, household formation and recombination leads to a sample of descendants which is not representative because the changes in household structure could be correlated with characteristics of interest such as income in 1996, leading to a correlation with the probability of being in the sample. We conduct an analysis of the effect of treatment on low, middle and high income households in 1974 using the corrective weights. The results are compared with results derived using three other sets of weights: the current 1996 weights, no weights, and weights that only correct for the sampling bias.

Empirical Specification

For the longitudinal analysis, we used the dataset of 1974 households linked to 1996 households based on descendant relationships. This means that some 1996 households appear multiple times if they are linked to more than one 1974 household. In addition, some 1974 households also appear multiple times if they are linked to more than one household in the MHSS sample. All of the weights used take this into account. There are four sets of weights used to run the analyses: the main weights of interest that correct for the 1974 sampling bias and 1996 bari selection issues, along with weights that only correct the 1974 sampling bias, no weights, and the original 1996 weights. These weights look as follows:

$$
\begin{aligned}
W_{74\_96} &= w_{1974} * \frac{\Pr(Y)}{D_{tot} * \Pr(i)} && (8) \\
W_{74} &= w_{1974} * \frac{1}{D_{MHSS}} && (9) \\
W_{none} &= \frac{1}{D_{MHSS}} && (10) \\
W_{96} &= \frac{1}{\Pr(i)} * \frac{1}{A} && (11)
\end{aligned}
$$

where w_{1974} is the weight for the 1974 household Y, Pr(Y) is the probability that the 1996 household i's antecedent Y has at least one descendant in the sample [14], D_{MHSS} is the number of descendants in the MHSS sample that are linked to the 1974 household Y, D_{tot} is the total number of descendants that the 1974 household Y has in 1993, A is the number of antecedents in 1974 that household i in the 1996 sample has, and $\Pr(i)$ is the probability that household i in 1996 is selected to be in the MHSS sample.

Devising a single measure of income from the data would be difficult and often income can be more sensitive to temporary shocks, so we instead focus on household consumption. In 1996 there is data on the value of total household food consumption, value of food given to others, and total spent by all household members on non-food items. These three variables are combined and divided by household size to come up with an aggregate household per capita consumption

[14] Note that in our weights, Pr(Y)=w_{1974} due to the way we created the 1974 weights, and therefore the two terms cancel out.

measure. To compensate for the lack of consumption data in 1974, we create an index to proxy for consumption in 1974. There is data on several assets in 1974 that are also in the 1996 survey. The consumption measure is regressed on the 1996 assets and the coefficients are used as weights to create a consumption index for 1974 (this is the same procedure used when checking the robustness of the weights using the 1996 SES).[15] The 1974 sample is divided into thirds based on this consumption index. The cutoffs are based on the full 1974 population, and households are labeled as low, middle or high income according to these cutoffs.

We estimate the following equation using generalized least squares in order to determine the effect of the treatment and income status in 1974 on income in 1996:

$$c_{i,t} = \beta_0 + \beta_1 c2_{ij,t-1} + \beta_2 c3_{ij,t-1} + \beta_4 (Tc1)_{ij,t-1} + \beta_5 (Tc2)_{ij,t-1} + \beta_6 (Tc3)_{ij,t-1} + \epsilon_{ij,t} \quad (12)$$

where $c_{i,t}$ is the consumption per capita for a household in 1996 descended from household i in 1974, $c2_{ij,t-1}$ is a dummy for household i in village j in 1974 belonging to the middle consumption group, $c3_{ij,t-1}$ is a dummy for household i in village j in 1974 belonging to the highest consumption group, $(Tc1)_{j,t-1}$ is an interaction between a treatment dummy for whether village j received the program in 1978 and a dummy for the low consumption group, $(Tc2)_{ij,t-1}$ is an interaction between the treatment dummy and the dummy for the middle consumption group, $(Tc3)_{ij,t-1}$ is an interaction between the treatment dummy and the dummy for the high consumption group, and $\epsilon_{ij,t}$ is an error term.

Results

The results of the reduced form model are shown in Table 3. Column 1 shows the results using our weights, which correct for both the issue with sampling from 1993 rather than 1974 and the bari nature of the sampling. Looking at the coefficient on the interaction term of treatment and low income in 1974, there is a positive and significant effect, implying that the households in 1974 in the lowest group in the treatment area have higher consumption than the equivalent group in the control area. In contrast, there is no significant effect of the treatment on the middle and high income groups as compared to how they do in the control area. This is in line with the potential channel of improved marriage prospects for low income households in the treatment area and no such changes for middle and high income households.

[15] Again, the variables used for the index are whether a household has a cow, boat, clock, or radio; whether it gets its drinking water from a tubewell; and whether the roof is made of tin.

Looking at the control group, the descendants of households with middle income in 1974 have higher consumption than households descending from the lowest consumption group in 1974, and those descending from the highest consumption group in 1974, have even higher consumption compared to the lowest group. This is not necessarily surprising, because it implies that if one is well off in 1974, his or her descendants continue to be well off, so there is persistence in the income distribution. Nevertheless, the fact that those in the lowest group in the treatment area experience higher income could mean that in the treatment area, the gap between the groups is lower. It would be possible to examine whether income mobility is increasing or decreasing in the treatment and control areas by comparing the gap between the different groups in 1974 to the gap in 1996, but considering the crude measure of income in 1974, this result would by a very rough estimate.

	(1)	(2)	(3)	(4)
VARIABLES	74 and 96 Weight	1974 Weight	No Weight	1996 Weight (original)
Treat x Lowest Consump	0.238***	0.0496	0.0462	0.151***
	(0.0446)	(0.0438)	(0.0469)	(0.0469)
Treat x Middle Consump	-0.0136	-0.0280	0.0577	-0.0227
	(0.0424)	(0.0410)	(0.0400)	(0.0400)
Treat x Highest Consump	0.0776	0.102**	0.0709*	0.0609
	(0.0500)	(0.0478)	(0.0428)	(0.0423)
Middle Consump	0.422***	0.334***	0.161***	0.283***
	(0.0427)	(0.0420)	(0.0402)	(0.0431)
Highest Consump	0.652***	0.541***	0.353***	0.462***
	(0.0490)	(0.0471)	(0.0429)	(0.0456)
Constant	6.450***	6.581***	6.721***	6.592***
	(0.0300)	(0.0299)	(0.0299)	(0.0320)
Observations	7,105	7,110	7,110	7,110

Standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 3: Log Per Capita Consumption in 1996 Regressed on the Consumption Third in 1974

Now, correcting only for the sampling problem and using a representative population of 1974 without taking into account household formation and recombination and their correlation with income in 1996, we find very different results in column 2. There is no longer a positive effect on low income households in the treatment area, and instead there is only a positive impact on the highest income households, suggesting increased income inequality in the treatment area. This could be explained if richer households tend to remain in the same bari, while poorer households split off. Foster 1993 showed that wealth is negatively correlated with household partition using Matlab data from the 1974 census and the 1982 census. It would therefore be reasonable that there could be a negative correlation between wealth and splitting off from a bari. In this case, if the treatment led to an increase in the income of the descendants in the lowest income third, those descendants that are better off would tend to live together in the same bari. So if a poor 1974 household in the treatment area had three descendants and two of them became well off due to the treatment and one remained poor, the bari sampling which does not allow for the two better off households to both be picked would lead to the estimation of a lower average consumption. Using our weights, we correct for this and weigh the households in a way that tries to incorporate the number of descendants and the probability of different combinations of descendants being selected, and therefore it would give a higher weight to the richer household and allow us to be closer to the actual average consumption of the descendants.

The results in column 3 show the effect of the treatment without using any weights. The results are very similar to those in column 2, but lower in magnitude. Table 1 showed how based on assets in 1974, the 1974 sample linked to the MHSS is not representative of the 1974 population and instead overrepresents wealthier households in 1974. Therefore, the lowest income group in the sample would be higher income than the actual lowest income group. This explains why the difference between the lowest and middle group and the lowest and highest group are now smaller in magnitude because the comparison lowest group is not representative of the lowest income group in 1974.

Finally, turning to the results using the original 1996 weights to look at changes in income due to the treatment, there are similar results to the findings using our corrective weights, although smaller in magnitude. Having seen the effects of not using our 1996 corrective weights in column 2 and then not using the 1974 corrective weights in column 3, this outcome makes sense. Although the current 1996 weights might not per se correct for household recombination and formation being correlated with outcomes, they do take into account the probability of being chosen based on bari size. Therefore, we can go back to our example of having a low income household in the treatment area that has two high income descendants living together and a low income descendant living separately, so that the two high income descendants would never both be chosen. The original 1996 weights would place a higher weight on the rich household because it has a smaller probability of being chosen. In this way, only using the current 1996 weights still can yield the effect on the lowest income households in 1974.

At the same time, the magnitude of all the coefficients is lower (just as is the case when comparing columns 2 and 3) because using the original 1996 weights does not account for the bias in the sample. Therefore, the households descend from higher income households and so the effects are smaller because the biggest effect is on the lowest income households in 1974, who are underrepresented.

According to these estimates, the treatment led to a 24 percent increase in consumption for the descendants of low income households and it did not affect the consumption of descendants from middle and high income households. This would imply that the treatment leads to a decrease in income inequality as the lowest income are the ones who benefit most. If corrective weights were not used, and instead only the original 1996 weights were applied, we would reach a similar conclusion, although the effect would be only a 15 percent increase in consumption. In conducting this type of intergenerational analysis, it therefore would be prudent to use the combination of 1974 and 1996 corrective weights in order to correct for both types of bias.

6 Conclusion

We started out wanting to answer the question of what are the long term effects of a maternal and child health and family planning program on household income. In order to properly conduct this analysis, though, we had to think through the broad question of how to devise the appropriate weights for panel data when there might be bias in the selection of the sample, and we created those weights based on the MHSS and HDSS data from Matlab, Bangladesh. We first laid out the various issues that arise with the Matlab data due to the post-1978 selection of the MHSS sample. Although this is a problem specific to the Matlab data, it is one that might apply in any of the other HDSS sites where an intervention was conducted on a sample of the population, or only a sample of the HDSS population was later tracked after an intervention in the region. It could arise even in the case of regular panel data if the formation and recombination of households combined with the choice of descendants picked to be surveyed leads to selection bias in the sample that is followed up. Therefore, in the case of any development intervention where there are such data limitations, we have created a possible framework for weights that can help to mitigate the bias.

We devised two sets of weights to help solve the two main problems with the MHSS/HDSS data. For the first, to make the 1996 sample representative of the 1974 population, we used the nature of the HDSS data which allowed us to mimic the process that had been used to create the actual sample in order to come up with probability weights. Even in the case of panel data where the full population is not available to conduct this sort of resampling procedure, propensity score weights also give extremely good results in helping to correct the sampling bias. For the second problem, we have found a formula for weights that can be universally applied in the case of multiple descendants where not all descendants have the same probability of being picked. Nevertheless, the application of these weights is not advisable if there is no correlation between the probability of being selected and the characteristics of interest, or if the analysis is not focused on aggregate data.

Using the sample weights to look at the main question of interest, we found that the maternal and child health and family planning program led to an increase in household consumption for the lowest income third of the treatment population. Among other things, this could be due to better marriage prospects for descendants from poor households in the treatment area, or an increase in consumption due to a decrease in the number of children. It would be necessary to further explore changes in marriage partners along with other variables in order to determine the mechanism leading to this finding.

In addition, there was an improvement in estimates by using our weights as compared to using the current 1996 weights. We further break down the effect of using the weights by looking at the impact of using just the 1974 weights or no weights, which gives us more insight into the type of bias that the sampling issues could be leading to. The sample of 1974 households linked to the 1996 sample is wealthier, and the data can be used to further analyze whether richer households tend to live in the same bari while poorer ones split off.

As discussed, there are many similar interventions where baseline surveys and subsequent household composition changes of the study population are not available. Employing similar weights to the ones we have created here, which take into account the various issues created by the process of household formation and recombination, could lead to significant improvements in the evaluation of these development interventions.

References

Arunachalam, Raj and Suresh Naidu (2008). "The Price of Fertility: Marriage Markets and Family Planning in Bangladesh". In: *University of California, Berkeley.*

Asadullah, M Niaz (2012). "Intergenerational wealth mobility in rural Bangladesh". In: *Journal of Development Studies* 48.9, pp. 1193–1208.

Barham, Tania (2012). "Enhancing cognitive functioning: Medium-term effects of a health and family planning program in Matlab". In: *American Economic Journal: Applied Economics* 4.1, pp. 245–273.

Becker, Gary S and H Gregg Lewis (1974). "Interaction between quantity and quality of children". In: *Economics of the family: Marriage, children, and human capital.* UMI, pp. 81–90.

Case, Anne, Darren Lubotsky, and Christina Paxson (2001). *Economic status and health in childhood: The origins of the gradient.* Tech. rep. National Bureau of Economic Research.

Fauveau, Vincent (1994). "Women, Children and Health". In: *ICDDR, B special publication* 35.

Field, Erica and Attila Ambrus (2008). "Early marriage, age of menarche, and female schooling attainment in Bangladesh". In: *Journal of Political Economy* 116.5, pp. 881–930.

Fitzgerald, John, Peter Gottschalk, and Robert A Moffitt (1998). *An analysis of sample attrition in panel data: The Michigan Panel Study of Income Dynamics.*

Foster, Andrew D (1993). "Household partition in rural Bangladesh". In: *Population Studies* 47.1, pp. 97–114.

Foster, Andrew D and Mark R Rosenzweig (2002). "Household division and rural economic growth". In: *Review of Economic Studies*, pp. 839–869.

Hall, Robert E and Frederic S Mishkin (1982). *The sensitivity of consumption to transitory income: estimates from panel data on households.*

Joshi, Shareen and T Paul Schultz (2007). *Family planning as an investment in development: evaluation of a program's consequences in Matlab, Bangladesh.* Tech. rep. IZA Discussion Papers.

Joshi, Shareen and T Paul Schultz (2013). “Family Planning and Women’s and Children’s Health: Long-Term Consequences of an Outreach Program in Matlab, Bangladesh”. In: *Demography* 50.1, pp. 149–180.

Maitra, Pushkar (2003). “Schooling and educational attainment: evidence from Bangladesh”. In: *Education Economics* 11.2, pp. 129–153.

Moffit, Robert, John Fitzgerald, and Peter Gottschalk (1999). “Sample attrition in panel data: The role of selection on observables”. In: *Annales d’Économie et de Statistique*, pp. 129–152.

Phillips, James F et al. (1982). “The Demographic Impact of the Family Planning–Health Services Project in Matlab, Bangladesh”. In: *Studies in Family Planning*, pp. 131–140.

Rahman, Omar et al. (1999). “The 1996 Matlab Health and Socioeconomic Survey: Overview and User’s Guide”. In: *RAND Corporation Draft Series DRU-2018/1-NIA*.

Rosenzweig, Mark R and Kenneth I Wolpin (1980). “Testing the quantity-quality fertility model: The use of twins as a natural experiment”. In: *Econometrica: journal of the Econometric Society*, pp. 227–240.

Schultz, T Paul (2009). “How Does Family Planning Promote Development? Evidence from a Social Experiment in Matlab, Bangladesh 1977-1996”. In: *Yale University, Economic Growth Center, New Haven, Conn.*

Smith, James P (1999). “Healthy bodies and thick wallets: the dual relation between health and economic status”. In: *The journal of economic perspectives: a journal of the American Economic Association* 13.2, p. 144.

Weil, David N (2005). *Accounting for the effect of health on economic growth*. Tech. rep. National Bureau of Economic Research.