Assessing the Frontiers of Ultra-Poverty Reduction: Evidence from CFPR/TUP, an Innovative program in Bangladesh

M. Shahe Emran [1]
Virginia Robano
Stephen C. Smith

Department of Economics,
George Washington University

This version: November, 2008

Abstract

This paper uses household panel data to provide evidence on the effects of a pioneering anti-poverty program of BRAC in Bangladesh (called Challenging the Frontiers of Poverty Reduction/Targeting the Ultrapoor, or CFPR/TUP) that attempts to target the poorest of the poor. We focus on the effects of program participation on a set of household outcomes including food security, income, and asset accumulation. To construct appropriate treatment-control groups, we partition the sample using type 1 and type 2 assignment errors according to BRAC's criteria for inclusion into and exclusion from the TUP program. We use a wide set of econometric approaches including the difference-in-difference matching estimator to identify and estimate the average treatment effect. To capture the potential heterogeneity of the treatment effect, we use a quantile difference in difference approach. The evidence shows that program participation had a significant positive effect on income, and food security of the ultra poor, but there is weak or no evidence in favor of a program impact on health related outcomes and ownership of homestead land. The quantile difference-in-difference results show that the lowest two deciles get much less benefit from program participation compared with the two highest deciles among the ultrapoor.

[1]We would like to thank the Research and Evaluation Division (RED) of BRAC for providing us with the panel data used in this paper. Special thanks to Munshi Sulaiman for help at different stages of the project. We thank participants at the Northeast Universities Development Conference at Boston University (Nov. 2008) for their comments. The standard disclaimers apply. Emails for correspondence: shahe.emran@gmail.com (M. Shahe Emran), vrobano@googlemail.com (Virginia Robano), and ssmith@gwu.edu (Stephen C. Smith).

Introduction

It is increasingly appreciated, by both practitioners and academics alike, that extreme poverty (or ultra-poverty) is qualitatively different from other forms of poverty and deprivation (see, for example, IFPRI (2007), Matin et. al. (2008), WDR (2006), Lipton (1983)).[2] Ultra-poverty differs from conventional poverty in terms of depth (degree of deprivation), length (duration of time) and breadth (the number of dimensions such as illiteracy, malnutrition etc.).[3] The possible complementarity among the different dimensions is argued to potentially result in multiple mutually reinforcing poverty traps. This makes ultra-poverty a qualitatively different problem to address than conventional poverty.

The experience of last few decades suggests that while the poverty programs of NGOs including microcredit programs have, in general, been successful in reaching the moderate poor (i.e., households below poverty line, but relatively close to it), the poorest of the poor are more often inadequately served or completely bypassed by such programs. This appreciation led to the development and implementation of innovative anti-poverty programs that are designed especially for the ultra poor. These programs address the multitude of interrelated factors that create the conditions of extreme poverty and make it a trap difficult to escape from.

BRAC in Bangladesh, one of the largest NGOs in the world, had been at the forefront of such innovative programs for addressing extreme poverty. In 2002, BRAC developed and implemented a unique anti-poverty program called "Challenging the Frontiers of Poverty Reduction: Targeting Ultra Poor, Targetting Social Constraints" (henceforth TUP). TUP is a multidimensional program that incorporates both livelihood protection and promotion components. The program combines consumption support with human capital development in an initial phase with the ultimate goal of helping the ultra-poor graduate to the standard micro-credit program of BRAC.[4] TUP as a strategy to tackle ultra-poverty has attracted increasing attention in the literature.

BRAC is the world's largest NGO by some measures (membership, scope, and budget). Founded 1972, it started microfinance in 1974, which now includes approximately seven million women members. The BRAC Education Program (BEP) serves over 1 million (10%)

[2] Although there is a growing consensus that exterme or ultra-poverty is an important and difficult problem requiring novel intervention strategies, the concept of "ultra-poor" remains unsettled. There are different definitions in the literature: Lipton (1983) defines ultra poor in terms of a calorie intake threshold (a person is ultra poor if he/she gets 80 percent or less calorie of an appropriate poverty line calorie benchmark); a recent IFPRI report (2007) identifies an individual as ultra poor if he/she lives on less than 54 cents a day. Emran, Shilpi, and Stiglitz (2008) define ultra poor in terms of lack of physical and human capital endowments that results in exclusion from both labor and formal credit markets. In this paper, we do not focus on how to define or identify ultra poor, taking the BRAC identification scheme as given for the empirical analysis. BRAC definition refers to "not being able to meet even the barest of the basic needs". For recent analysis of issues related to identification and proper targetting of ultra poor, see Banrejee et. al. (2008) and Sulaiman and Matin (2006).

[3] For discussions, see World Bank (2000), Smith (2005), and Chronic Poverty Research center (2008).

[4] Other examples of programs for ultra poverty include the Grameen beggars program and the Bandhan "Chartering into Unventured Frontiers- Tragetting the Hardcore Poor (CUF-THP) program.

Banglaeshi primary students in some 35,000 informal schools. Over 110 million receive BRAC health and other services in Bangladesh. BRAC features such diverse activities as development-oriented enterprises, legal education for the poor, a bank, a university, and an internet service provider, among others. BRAC is now expanding abroad including activities in Afghanistan, Sri Lanka, Uganda, Tanzania, and South Sudan. The NGO has won numerous prizes and awards, and its innovations have been widely replicated. Thus, studying BRAC programs is unusually important.

Through the TUP program, BRAC is seeking to reach and serve the most marginalized and disadvantaged women (in this case in Bangladesh). Despite great progress, some 20% of Bangladeshis are "ultrapoor," spending 80% of income on food yet consuming less than 80% of minimum caloric intake. BRAC claims that the ultrapoor are generally left out of NGO and other mainstream poverty and development programs: microfinance institutions (MFIs) view the ultrapoor as too risky; the ultraoor lack the political connections needed to participate in government programs; and NGOs find it hard to identify the ultrpoor for NGO relief programs. TUP is the latest major BRAC initiative; so, based on past patterns, we can anticipate widespread replication and indeed similar programs are already being replicated in several other countries. The program features significant innovations for example in targeting, selected asset transfer with human and social capital support, graduated transition into microfinance, and an active "elite" community role. A careful analysis of the BRAC TUP program is particularly important because ultra-poverty manifests the worst form of human deprivation, so a better understanding of what kind of intervention is effective has very high social returns. Also, such interventions are very expensive and thus robust evidence on its effects would be important both for improving the program design and stimulating new approaches. In sum, the program is innovative, promising, well-funded, likely replicated, and not inexpensive; thus, the program and its underlying ideas deserve good and careful evaluations and analysis.

This paper uses a two period panel data set (2002, 2005) to analyze the effects of the TUP program participation on a set of household outcomes including income, food security, health and asset accumulation. To provide robust evidence on the treatment effect of participation in the TUP program, we use a rich set of econometric techniques. Starting from a simple difference-in-difference approach (DID), we allow for different time trends in different districts and control for selection on observables. In particular, we use the difference-in-difference matching estimator (Heckman et. al., 1998, Todd, 2007) that combines a difference-in-difference approach with the matching technique to eliminate the time invariant unobserved heterogeneity (henceforth called DIDM estimator).[5] The assignment errors in the selection of participants in the TUP program (according to the

[5]BRAC's in-house research and evaluation division (RED) used the same panel data set to do an impact evaluation of the TUP program using the standard difference-in-difference approach (see Rabbani et. al. 2006). In this paper, we use a rich set of econometric approaches to provide robust evidence on the treatment effects of the TUP program participation with careful considerations of selection issues (both BRAC's selection and the household's participation decision).

inclusion and exclusion criteria set out by BRAC) are used to partition the sample to generate alternative control and treatment groups, based on type 1 and type 2 errors.[6] The evidence from the DIDM approach as applied to the alternative treatment-control schemes show that there is significant positive effect of participation in the TUP program on net income and food security of the ultra poor.[7] There is also strong evidence of improvements in housing quality (presence of a tin roof) for BRAC's treatment group, the "selected ultra poor" (called the SUP group) and for the households incorrectly included into the program according to the stated selection criteria (called SNB_1 group), but the evidence for the treatment group most representative of the ultra poor (called SB_1) is not equally strong. Although the estimates from the DID and DIDM approaches show a statistically significant and numerically important effect for the ultra poor group SB_1, it is not always robust to allowing for even a small amount of selection on unobservables. There is very weak or no evidence of any significant effect of the TUP program on subjective health outcomes, women's empowerment (as measured by the ratio of number of sari (women's clothing) to lungi (men's clothing)), and on ownership of homestead land.

Although the average treatment effect[8] estimates from the DIDM approach are useful as summary measures of the effects of the TUP program participation, they are unable to shed much light on the possible heterogeneity in the treatment effects. To analyze the issue of heterogeneous treatment effect in depth, we implement the Quantile Difference-in-Difference estimator (henceforth QDID) that allows for different treatment effects across different parts of the distribution (only for continuous outcome variables like income). The results show that although, in general, strict monotonicity in the treatment effect does not hold, the lowest two deciles of of these ultra poor households always benefit much less (in absolute terms) from the program participation compared to the top two deciles of ultra poor households.

The rest of the paper is structured as follows. The first section provides a brief discussion of the BRAC TUP program. The second section discusses the empirical strategy for identification and estimation of the treatment effects in greater detail. The next section reports the results of empirical analysis on the treatment effect of program participation in a sequential manner starting from a simple difference-in-difference approach. The paper concludes with a summary of the findings.

2. The BRAC Ultra-poverty Program

[6]BRAC's in-house evaluation mentioned in footnote 5 above uses the selected ultra poor (SUP) as the treatment group and the not selected ultra-poor (NSUP) as the control group.

[7]The DIDM approach takes care of time invariant unobserved heterogeneity, but still relies on selection on observables to defend the common time trend assumption. However, there can be time variant unobserved heterogeneity that gives rise to differential time trends across the treatment and control groups. We thus use the recently proposed sensitivity analysis of matching estimators to see if the estimated treatment effects can be swamped by low to moderate selection on unobservables.

[8]In this paper, our focus is on the average treatment effect on treated (ATT). So the term treatment effect refers to ATT throughout the paper.

One of the most comprehensive approaches to redressing ultra-poverty has been developed and implemented by BRAC, the largest NGO in Bangladesh (and the world). BRAC had concluded that its previous efforts had helped the poor, but not the poorest of the poor, because of their malnutrition, poor health, illiteracy, and generally vulnerable status. TUP is a multidimensional attack on poverty, introduced in Bangladesh by BRAC in 2002, focusing on developing human capital (health, education, and training) and social capital (village support networks and sponsorship of community leaders) for poor women. It pairs these efforts with provision of specific physical assets such as livestock with a goal of later mainstreaming these clients into microfinance. The latter "graduation" to BRAC's mainstream village organizations is a goal of the TUP program, and a potential key to its sustainability. Three relatively poor districts in Northwest Bangladesh (Rangpur, Kurigram, and Nilphamari) were initially selected on the basis of poverty mapping.

TUP was launched in these three districts, with 5000 women selected from a larger group of potential participants, who together form the basis for our panel data set. All members of participant and control groups were selected by villagers as among the poorest local families. A subset was selected by BRAC according to exclusion and inclusion criteria. The exclusion criteria required that participating women must be capable of doing work outside the home, must not belong to another NGO program and must not receive a food benefits card. In the inclusion criteria, participating women have to meet three of the following: child labor is present; ownership of less than 10 decimals of land (a tenth of an acre), lack of a male earner at home, adult women selling labor outside of the household, and lack of any productive assets.

Understanding that the poorest were often not ready for microcredit, BRAC, in collaboration with WFP, ran the highly regarded program called Income Generation for Vulnerable Group Development (IGVGD; see Matin and Hulme 2003). The idea was to feed and provide health care for people until they have the strength for education and microenterprise development. Thus, IGVGD serves as a remedial program for the extremely poor, preparing them to graduate into mainstream programs for the moderately poor.

Nevertheless, BRAC founder Fazle Hasan Abed lamented that, "despite our efforts, we have not succeeded in reaching the ultra poor" (Smith 2005 p. 90). Thus, despite its reputation as an effective, poverty-focused NGO, BRAC was not quite reaching its goal. In 2002, BRAC decided to design an entirely new program to reach the ultra poor. Rather than simply modify existing programs, BRAC developed a new approach, "Challenging the Frontiers of Poverty Reduction: Targeting the Ultra Poor, Targetting Social Constraints". The TUP program theme is "pushing down and pushing out." It pushes down, to those with lower incomes, and it pushes out, by moving beyond the bounds of conventional programs to focus on the constraints that keep the ultra poor trapped in poverty. Behind this theme lies an important assumption for which BRAC has accumulated experiential evidence. The ultra poor face qualitatively different deprivations of opportunities and basic security, as compared with those who are poor but above this desperate state. Differences include the lack of able-bodied men in the household, social exclusion (that may among other things make it very difficult for them to successfully join group based microfinance village organizations), and other vulnerabilities to exploitation.

To find the poorest of such women, several strategies were used. One is "Participatory Wealth Ranking" that utilizes local information available to the villagers. A meeting is held in which a village map is drawn on the ground with each household labeled. The villagers agree on a wealth ranking among the households, to identify those who are the poorest of the poor. Those who can afford tin plate walls or roofs are less poor than those with straw walls or thatched roofs. Those who are known to have a steady, formal job are categorized as among the well off. To keep the process manageable, only about 150 households were included in each wealth ranking exercise.

There is an incentive for people to rank themselves as poor enough to receive assistance; but the multiple checks done on family status means their ability to get away with this is sharply limited. The mechanism is not perfect; better off people may find ways to convince BRAC people that they should be counted among the poor, and conversely, those in the most extreme poverty may not come forward at all; and people may forget their small huts when drawing village maps. Or, the poorest may not be identified because they are viewed as a part of the household of distant relatives who function in a clientalistic relationship with the poor (Matin et al 2008). Indeed, the more socially excluded among the poor may be less likely to be picked–yet their social exclusion is a fundamental cause of their poverty. To supplement community meetings, BRAC staff members walk through the village, looking for any hut that gives the appearance of extreme poverty. They then try to bring potentially overlooked ultra poor people to the attention of the community meetings.

Village leaders are actively involved in all stages of the process, generally people who are relatively well educated such as the schoolteacher. Some of these leaders are also clearly ineligible for TUP assistance, and their impartiality and knowledge tends to be respected, so they can help to mediate disagreements about who should be included as among the poorest. Some of these village leaders then help form a TUP support committee that monitors progress and intervenes to prevent personal crises that can overwhelm participants and otherwise lead them to drop out of the program. This feature is based on BRAC's observation that without achieving reliable social support in village organization and governance–"socio-political assets"–the ultra poor often cannot break the cycle of poverty. However, the functioning of this village elites program is outside the scope of this paper.

3. The Data and Variables Description

For the empirical analysis, we use the BRAC TUP panel data set. This is a two-year panel of about 5000 households. The baseline survey of 5626 households was done in 2002. In 2005, 5288 households were resurveyed, along with 278 newly formed households that had split from the initial set of households. Attrition was moderate and was due to migration, death, and marriage. The final matched panel contained 5067 households.

The BRAC TUP panel data set provides information on a wide range of household characteristics and outcomes. The survey contains a rich body of information regarding the asset base of the household that includes natural (land), physical, human, financial and social assets (capital). We estimate the causal effects of program participation on income and physical assets, food security, health, and women's empowerment. Although our analysis covers both the flow and stock variables, one might argue that three years may not be enough to capture long term effects of the program, and thus the evidence on the stock variables should be interpreted with appropriate caveats.

The variables that we have used for matching are (at 2002 levels): gender, body mass index, age of the household head, each of the inclusion criteria, a dummy for whether the main source of income was from day-labor activities, a dummy for whether the person owned homestead land, and a variable that measures the amount of land that the individual owned. Appendix 2 provides a more complete description of the variables used in the empirical analysis.

4. Empirical Strategy

For a proper analysis of the treatment effect of the TUP program, we need to construct the treatment and control groups carefully so that any potential selection bias can be minimized. BRAC's own treatment and control groups are called "selected ultra poor" (SUP) and "nonselected ultra poor" (NSUP). Although both the treatment group (SUP) and the control group (NSUP) in the BRAC panel data set are drawn from among extremely poor households identified by villagers (thus reflecting local knowledge), they are differentiated by BRAC's systematic inclusion and exclusion criteria, and may suffer from other selection biases for a variety of reasons (see below). So the SUP-NSUP subsets may not be the best possible treatment and control groups for estimating the treatment effects, especially when the interest lies in understanding the effects of the program on the ultra poor. We utilize errors in assignment in BRAC's selection to construct alternative treatment and control groups based on type 1 and type 2 errors. Based on the formal selection criteria of BRAC, we partition the sample of households in the panel data set into four sub-sets. They are: (*i*) households that are eligible according to the stated criteria and are included in the program (subset called the "should be, one" (SB_1) group henceforth), (*ii*) the eligible households not selected (called the "should be, zero" group (SB_0)), (*iii*) households ineligible according to formal criteria but selected in the program (called the "should not be, one" group (SNB_1)), and (*iv*) households ineligible and not selected (called the "should not be, zero" group (SNB_0)). For details on the construction of these four subsets, please see Appendix 1.

There are two levels of selection problems that we have to consider: (i) BRAC's selection process, and (ii) the participation decision by households. As discussed earlier, BRAC's selection process was based on a set of explicit inclusion and exclusion criteria. To understand the nature of potential selection bias arising from BRAC's selection process we need to have an implicit model of the actual decision making by BRAC employees. The simplest model is to assume that the BRAC employees were following the set of inclusion and exclusion criteria strictly, and thus the *assignment errors* discovered in the data are either completely random or due to the fact that some eligible households declined to participate in the program. If systematic self-selection out of the program by eligible households is important then households in group SB_0 are likely to differ systematically both in observed and unobserved dimensions from eligible households that participate in the program, i.e., SB_1. The alternative model is to assume that BRAC employees were using both the formal criteria and private signals available to them. In this case, the objective function of the BRAC employees become critical. If the objective was to identify the true ultra-poor, then the group of households who should have been in the program according to the set of formal criteria but were not selected (i.e., SB_0) must be relatively well off (more advantaged) in terms of initial economic conditions and characteristics in 2002. Under the alternative assumption that the objective was to identify and exclude potentially high risk households so as to help ensure the success of the program, then the SB_0 group is likely to be systematically more disadvantaged in 2002. Of course, it is possible that both of these effects are present and largely wash each other out.

Table 1a reports the difference in means and the associated standard errors for a set of observable characteristics in 2002 across different pairs of treatment-control groups. The first column gives the initial difference in means for the $SUP - NSUP$, the second for $SB_1 - SB_0$, the third for $SNB_1 - SNB_0$, and the last for $SB_1 - SNB_0$. The evidence in Table 1a clearly shows that the initial difference in the means is, in general, much lower for the treatment-control pair $SB_1 - SB_0$. In contrast, there are some significant and relatively large differences in the initial conditions in 2002 between the treatment and control groups as defined by BRAC (i.e., the subsets SUP and $NSUP$) and used by BRAC's Research and Evaluation Division (RED) in its "descriptive analysis" of the TUP program (Rabbani, Prakash, and Sulaiman, 2006). This confirms the possibility that the NSUP may not be an appropriate control group for the treatment group SUP. Consider for example, the variable "change in net income over the last year" in the last row. The differences in means are: $Tk.162(SB_1 - SB_0), Tk.1924(SUP - NSUP)$, $Tk.1362(SNB_1 - SNB_0)$, and $Tk.5289(SB_1 - SNB_0)$. It is interesting that the subsamples $SB_1$and SB_0 look much more similar according to the observable characteristics reported in Table 1. Since selection on observables and selection on unobservables are likely to be related (a point emphasized recently by Altonji et. al. (2005)), it is reasonable to assume that SB_0 constitutes an appropriate control group to estimate the treatment effect when the treatment group is SB_1. The fact that the groups SB_1 and SB_0 look similar to each other is, however, not consistent with the hypothesis that BRAC employees were systematically excluding some of the eligible households.[9] This evidence also does not lend support to the hypothesis that the SB_0 households self-selected out of the program because they constitute very

[9]Note that although BRAC employees may have more information, some of the most important individual characteristics like ability are unobservable to both the BRAC employees and the econometrician.

different types of households compared with eligible participants (i.e., SB_1 households).[10] If BRAC's inclusion and exclusion criteria were implemented perfectly we would have the treatment group SB_1 and the control group SNB_0. The difference in means in 2002 between these two groups is much more pronounced than the differences across SUP and NSUP (see Table 1a).

The evidence also indicates that there are important differences in the initial conditions across the three different treatment groups. Table 1b reports the group averages of a set of variables in 2002 across the groups. Although the groups are similarly situated according to some observables like food availability, and quality of houses as indicated by the roof made of tin, the SB_1 is clearly the poorest among them. While the percentage of households who own their homestead land is 39 percent for the SB_1 group, the corresponding numbers for SUP and SNB_1 are 47 percent and 53 percent respectively. The increase in net income from 2001-2002 was $Tk.5860$ for an average SB_1 household, $Tk.8150$ for SUP, and $Tk.9787$ for SNB_1. This implies that if one is interested in understanding the treatment effect of the TUP program on the poorest of the poor, SB_1 is the most appropriate treatment group to focus on with the appropriate control group SB_0. Given the above analysis, we put more weight on the estimates of treatment effects from the combination of SB_1 (treatment) and SB_0 (control), although we also report the estimates from other possible treatment-control combinations including BRAC's own classification (i.e., SUP (treatment) and $NSUP$ (control)).

To estimate the treatment effect using the alternative treatment-control groups as discussed above, we use difference in difference (DID) with and without differential time trends in different districts (i.e., Rangpur, Kurigram, and Nilphamari). Moreover, additional controls are included in the DID regressions which might affect both the treatment decision and the outcome variables to account for selection on observables. We also combine the difference-in-difference approach with propensity score matching (the DIDM estimator). As mentioned earlier, the DIDM approach purges any time invariant heterogeneity at the individual level by time differencing; and then matching takes care of selection on observables in a flexible way without imposing any particular functional form. This, however, does not address the possibility that the estimated treatment effect may be contaminated by selection on unobservables that vary over time and thus may result in differential time trends across treatment and control groups. We implement sensitivity analysis for the DIDM results to see if the estimated treatment effects can be solely due to reasonable magnitudes of selection on time varying unobservable factors.

There has been a growing appreciation in the recent literature that treatment effects are, in general, heterogeneous in a non-trivial way (Ravallion, 2007, Heckman, et. al. 1998). We implement a quantile difference in difference approach (QDID) to provide evidence on

[10]It is possible, at least in theory, that the selection is dominated by unobservable characteristics that are not in the information set of the BRAC employee. We implement some sensitivity tests for the DIDM results ro see if the estimated treatment effects can be solely due to selection on unobservables (see, Table 5 below).

heterogeneity in the treatment effects in different parts of the distribution. We report results from the alternative specifications of the QDID: common time trends, differential time trends in different districts, and also with and without additional controls to take into account selection on observables.

5.Treatment Effects of the TUP Program

(5.1) Results from the Difference-in-Difference Approach

In this section, we report the estimated treatment effects on a set of household outcomes including income, assets, and health related indicators using alternative specifications of the difference-in-difference approach. The standard difference-in-difference specification is based on the following model of the treatment effect:

$$Y_{it} = \alpha_0 + \alpha_1 d_{05} + \alpha_2 d_T + \beta \ (d_T * d_{05}) + \epsilon_{it} \tag{1}$$

where Y_{it} is the outcome variable of interest for household i in year t, d_{05} is a dummy that equals 1 for the year 2005, and d_T is a dummy that equals 1 when household i belongs to an appropriately defined treatment group (i.e., SB_1, SUP, SNB_1) and equals zero when a household belongs to the corresponding control group (i.e., $SB_0, NSUP, SNB_0$). The parameter of interest is β that isolates the treatment effect on outcome Y under certain assumptions. The crucial difference-in-difference estimation assumption is that the treatment and control groups would follow the same trend in the absence of the program. If this assumption is not satisfied, the estimate of the treatment effect $\hat{\beta}$ will be biased when we use OLS to estimate equation (1). We augment the basic DID specification in two ways to make it more plausible that the counterfactual trend for the treatment group is well represented by the actual trend in the control group. First, we allow for differential time trends in the different districts in our data set. This leads to the following specification:

$$Y_{it} = \alpha_0 + \alpha_1 d_{05} + \alpha_{1R} (d_{05} * d_R) + \alpha_{1K} (d_{05} * d_K) + \alpha_2 d_T + \beta \ (d_T * d_{05}) + \epsilon_{it} \tag{2}$$

where d_R and d_K are dummies for Rangpur and Kurigram districts respectively.[11] This is relevant because evidence reported in Sen and Hulme (2005) indicates that a measure of human poverty fell in the 1995-2000 period by 3.57% in Nilphamari, but only 1.73% in Kurigram and 1.65% in Rangpur. In addition, we allow for the possibility that the trends might differ across households based on selection on observables. Thus, we also control for a set of observables that are likely to be important for selection into the treatment (either because of BRAC's criteria, or the household's own outside option). Controlling for selection on observables results in the following specification of the DID regression:

$$Y_{it} = \alpha_0 + \alpha_1 d_{05} + \alpha_{1R} (d_{05} * d_R) + \alpha_{1K} (d_{05} * d_K) + \alpha_2 d_T + X_{02}' \Pi + \beta \ (d_T * d_{05}) + \epsilon_{it} \tag{3}$$

[11] The omitted district is thus Nilphamari.

where X_{02} is the set of controls in 2002 added to equation (2) above.[12]

Table 2a presents the estimated treatment effect from specifications (1)-(3) across three different treatment-control pairs discussed before. For binary outcome variables such as food deficit or homestead ownership, we report the estimates from probit regressions, although the estimates from linear probability models are, in general, very similar.[13] A few general patterns emerge from the estimates reported in table 2a. Although the magnitudes of the estimated treatment effects vary across different specifications of the difference-in-difference regression, they, in general, fall within tight bounds (an important exception is the 'change in net income'). The estimates of the treatment effect vary a little bit more across the different treatment-control pairs, although for some outcome variables like 'two meals a day', the difference in the magnitude is small, especially between $SB_1 - SB_0$ and $SUP - NSUP$. Although the estimates from the BRAC classification $(SUP - NSUP)$ and our preferred classification $(SB_1 - SB_0)$ differ significantly for some of the outcomes (for example, change in net income), the treatment effect estimates are broadly similar. It is, however, important to appreciate that this similarity can be misleading when the households in the SB_1 group start from much lower initial conditions in 2002 in terms of a given indicator like 'change in net income'. The treatment effect is substantially higher for the SB_1 group relative to SUP when we normalize by the mean in 2002 for different groups for such outcomes (reported in Table 2.b). The results in Table 2a-2b provide evidence that TUP program participation has significant positive effects on a number of important household outcomes including income and food security, especially for the target group, i.e., the participating households that satisfy the BRAC selection criteria (SB_1). In contrast, there is no significant effect on any of the health related indicators nor on ownership of homestead land.

We now turn to the details of the results. Considering first the impact on 'change in reported income over the last year', income gains are consistently higher for the participants in the TUP program across all three different specifications of DID for SB_1 and SUP. But for the group of participants who fail to satisfy the BRAC eligibility criteria (SNB_1), the estimate of the treatment effect becomes small and statistically insignificant when we include differential time trends and control for selection on observables in the DID regression. The estimates for our preferred treatment group SB_1 show that controlling for selection on observables substantially reduces the estimate of the treatment effect while allowing for a differential time trend does not have any appreciable effect. The estimate from the most general specification of DID shows a slightly higher impact of the program on the SB_1 group $(Tk.3131)$ compared to the BRAC treatment group SUP $(Tk.2941)$. When expressed as percentage of the group mean in 2002, the treatment effect for SB_1 (53

[12]The set of variables used for selection on observables is discussed in detail in the following section where the results from the DIDM approach are presented.

[13]The estimates for "food deficit" are somewhat different in terms of their numerical magnitude when we use a linear probability model. However, the main conclusions of the paper remain intact irrespective of the estimation method.

percent) is much higher than that for SUP (36 percent) (see Table 2b).

Perhaps the most important impacts in terms of human welfare were found in alleviating the problem of food insecurity. We use two indicators of food security: 'food availability' and 'ability to obtain two meals a day', as reported in Tables 2a-2b. The fraction reporting better food availability was significantly higher, especially among the SB_1 and the SUP households. It is interesting that the estimates for SUP vary significantly across different specifications of DID regressions, but are identical for SB_1. This can be interpreted as evidence that selection on observables is important for the SUP group with regards to this particular outcome. Also, the treatment effects are similar for SUP and SB_1 for the food security outcomes, especially if we focus on the estimates from the most general specification of the DID regressions (see column 3 for each treatment group in table 2a).

Failing to own the land on which ones house is located is a basic determinant (and indicator) of lacking even the most minimal wealth and security. The houses of participants are generally little more than one room shacks, so lack of ownership of these tiny plots is a signal of extreme poverty, insecurity, and general vulnerability. The estimates show a weak program impact on this outcome variable. There is a small but statistically significant effect according to the simple DID and DID with differential time trend estimates. However, once we control for selection on observables, the impact is reduced more and becomes statistically insignificant at the 10 percent level, for the SB and SNB groups. A related outcome variable is tin material for roofs, a positive indicator of the overall housing quality in rural Bangladesh. Interestingly, there is a significant program impact irrespective of DID specifications for both SB_1 and SUP groups, but the program impact becomes statistically insignificant at the 5 percent level for the SNB_1 group once we control for selection on observables. Again, the magnitude of the program effect does not vary significantly between the SUP and SB_1 groups (see Tables 2a and 2b).

Next, consider the survey questions on subjective self-reported health status and health improvement over last year (rows 1 and 2 respectively in Tables 2a and 2b). The estimated treatment effects are very small across the board and not significant at 5 percent level. Moreover, the estimates are small in terms of numerical magnitude. We thus do not find any evidence of any significant effect of TUP program participation on the subjective health outcomes. Although this might reflect the fact that health improvements take time and may be subject to threshold effects, the evidence should be interpreted with additional caveats. The reported health indicators may have significant measurement error or reporting bias, in part due to its subjective nature, and in part because better health training, as provided in the program, can lead to increased awareness of participants' conditions as health problems. This effect would bias downward responses from actual improvements.

Possessing sandals/shoes is important not only to protect feet from cuts but also to prevent other infections including parasites (see e.g. Smith 2005), and to improve speed and flexibility of movement in an environment in which the poor largely travel on foot. A significant and substantial positive effect is found whether looking at the SUP or SB_1

households, but the effect becomes statistically weak for the SNB_1 when selection on observables is controlled for. The numerical magnitude of the treatment effect is smaller for SB_1 compared with SUP, both in terms of absolute and normalized treatment effects (see Tables 2a and 2b).

Another indicator of basic wellbeing for women in Bangladesh is the number of saris (dresses) a woman owns. Following the literature, this can also be viewed as an indicator of a woman's bargaining power in the household.[14] It is interesting that the treatment effect is significant both numerically and statistically across treatment groups and DID specifications. The normalized treatment effects are also similar (see Table 2b). The results thus provide evidence that TUP program participation significantly improves the clothing of women in the household across different treatment groups.

The evidence on number of sari discussed above is, however, a noisy indicator of a woman's bargaining power at best. Even if there is no change in the bargaining power of women due to participation in the TUP program, the number of shari a woman owns may be higher because of income effect reflecting higher income gains discussed earlier. A better indicator of women's household bargaining power is the ratio of saris to lungis (male clothing). The estimates in Table 2a show that there is a significant effect of TUP program participation on the ratio of saris to lungis for the BRAC treatment group SUP. The evidence of a positive effect for the SB_1 group is, however, weak; the estimated treatment effect is consistently smaller and becomes statistically insignificant when the observable characteristics in 2002 are controlled for.

(5.2) Difference-in-Difference Matching Approach (DIDM)

The estimates reported in Tables 2a and 2b and discussed in the preceding section provide us with robust evidence on the treatment effect of TUP program participation. The results, however, rely on two restrictive assumptions: (i) the selection on observables is adequately controlled for by the postulated linear effects of the varibles included in the DID regressions (i.e., the vector X_{02}), and (ii) selection on unobservables is not strong enough to dominate the estimated treatment effects. However, as widely discussed in the literature, both of these assumptions may not be tenable in many applications. The DIDM approach gets around the first problem by using matching techniques to control for selection on observables. As mentioned earlier, the DIDM approach still relies on selection on observables for identification and thus assumes implicitly that the degree of selection on unobservables is not significant. In this section, we first report the estimated treatment effects from the DIDM approach and then provide evidence on the importance of selection on unobservables using sensitivity analysis.

A critical step in implementing the DIDM approach is to choose an appropriate set of observable characteristics that are important in determining the selection into treatment

[14] There is now a large literature that interprets expenditure on feminine goods as an indicator of women's intra-household bargaining power. See, for example, Deaton (??), other references.

and may also affect the outcome variables. As discussed before, we need to consider two levels of selection: BRAC's selection process and also the participation decisions of the households. We thus use observables that reflect these two levels of selection problems for matching. To account for the BRAC selection process we use the set of inclusion criteria. We also include indicators of a household's physical and human capital (for example, land owned, household size, BMI, age, and the indicator of women working as day laborers). As emphasized recently by Emran, Morshed and Stiglitz (2007), and Emran, Shilpi, and Stiglitz (2008), the outside option of a household and thus the net return they get from participation in the TUP or other NGO programs depends on the nature of labor market interactions and the shadow value of labor, especially of women's labor. We thus include household size as an indicator of labor endowment of the household, and the variable "day labor" as a measure of labor market participation by women. We also include "land owned," as it is a crucial variable for the determination of the shadow price of labor and also whether a woman is excluded from critical markets such as the formal credit market and the labor market.

Table 3a reports the estimated treatment effects from the DIDM estimator; and the corresponding normalized treatment effects are reported in Table 3b. The results are, in general, consistent with the conclusions reached above on the basis of the DID approach. The estimated treatment effects vary depending on the matching algorithm used, but they are, in general, confined within reasonable bounds. For example, consider the estimated treatment effect on food security as measured by "food availability" and "two meals a day". The intervals of the estimated treatment effects in the case of the SB_1 treatment group are [0.26, -0.31] and [0.39, 0.42] for food availability and two meals a day respectively. The estimate from the general DID regression is -0.24 (food availability) and 0.42 (two meals a day) . The corresponding intervals in case of the treatment group SUP are [0.25, -0.27] and [0.38, -0.40] respectively, while the estimates from the general DID are -0.25 and 0.42. The DIDM results, however, contradict the earlier conclusion that for SUP treatment group, there is significant positive effect of program participation on the ratio of saris to lungis which can be viewed as an indicator of women's relative bargaining power within the household. The evidence from DIDM shows that there is no significant effect on this outcome variable across different treatment groups including SUP.

How Does Selection on Unobservables Affect the Results? Evidence from Sensitivity Analysis

The evidence presented above in Tables 2a-2b, and 3a-3b does not take into account the implications of potential selection on unobservables for the estimated treatment effects. In this section, we present Rosenbaum bounds on the estimated treatment effects to provide evidence on the importance of selection on unobservables using the methodology developed by Aakvik (2001, DiPrete and Gangl (2004)) and Becker and Caliendo (2007). Table 4a presents the results from the sensitivity analysis for the binary outcome variables and Table 4b for the continuous outcome variable (net income). We concentrate on the outcome variables that indicate significant program effects according to the DIDM estimates in Table 3a. For the binary outcome variables, we present Mantel-Haenszel statistics (see Becker and Caliendo (2007) for details). There are two test statistics: $Q_m h^+$ is the Mantel-Haenszel statistic under the assumption that the estimated treatment effect is overestimated, and $Q_m h^-$ under the assumption that the estimated treatment effect is underestimated. The corresponding P-values are reported as $P_m h^+$ and $P_m h^-$ respectively. For continuous outcome variables, we present Wilcoxon signrank tests that give the upper and lower bounds on significance of the treatment effects for a given level of selection on unobservables (i.e., hidden bias). The results show that the estimated program impact on net income and food security (food availability and "two meals a day") are robust to allowing for a significant level of selection on unobservables irrespective of the treatment group. This can be seen from the P-values of the Mantel-Haenszel statistics under the assumption that the estimates are biased upward (i.e., the $P_m h^+$) for different levels of selection on unobservables represented by different values of Gamma. For example, when the odds of participation is 25 percent higher (i.e., Gamma=1.25) for the treatment group, the $P_m h^+$ is zero for food availability and two meals a day for all three treatment groups providing strong evidence that the estimated treatment effects cannot be driven by selection on unobservables. For other outcome variables such as number of saris and having a roof made of tin, the estimated program effects survive for the SUP and SNB_1 when we allow for selection on unobservables; but the program impact is insignificant for the SB_1 group in the presence of even a small amount of selection on unobservables.[15]

(5.3) Heterogeneity in Treatment Effects: Quantile Difference-in-Difference Approach (QDID)

Tables 5 report the results from estimating specifications (1)-(3) in the text using quantile regressions. This gives us a way to assess the extent of variations in the treatment effects across the distribution.[16] There are plausible theoretical reasons to expect that households who start at lower initial conditions may benefit less from the TUP program participation, at least in absolute terms. In principle this can be due to threshold effects and the myriad of interlocking constraints that create and sustain poverty traps for the poorest of the poor. If, on the other hand, one entertains a standard neoclassical view with concave production functions satisfying the Inada conditions, we would expect that the poorest of the poor would benefit the most.

The estimated treatment effects on the outcome variable 'change in net income over the last year' by QDID are reported in Table 5 for three different specifications of the DID model used in Table 2a. The most striking conclusion that holds across the board is that the households in the bottom two deciles of the ultra poor get substantially less

[15] We note the caveat that this cannot be taken as definitive evidence against a program impact, as we have no way to determine if there is actually significant selection on unobservables. One can interpret this evidence saying that even if there is a low degree of selection on unobservables, then the evidence in favor of a program impact is not strong.

[16] QDID has been used recently by Song and Manchester (2007), among others.

benefit from the TUP program participation compared to the households in the top two deciles of the ultra poord. Also, according to the estimates, the treatment effect is not statistically significant for the lowest decile households for the SB_1 and SUP groups, which reinforces the conclusion that the poorest of the ultra poor seems to be facing additional constraints. A second interesting pattern is that the magnitude of the program effect goes down substantially across the distribution when we control for selection on observables.

Conclusions:

Using a two-period household level panel data set, this paper provides robust evidence on the effects of TUP program participation on a set of important household outcomes. We use the errors in assignment in BRAC's selection to create alternative treatment and control groups. A rich set of econometric approaches are used to estimate the average treatment effect on treated (ATT) that takes into account both 'selection on observables' and 'selection on unobservables'. The results show that there is significant impact of program participation on net income and food security of the ultra poor households. The evidence also indicate that the TUP program does not have any significant effects on health related outcomes, women's empowerment as measured by the ratio of sari (women's dress) to lungi (male clothing), and on the ownership of homestead land. The estimates from quantile difference in difference approach show that the lowest two deciles of the ultra poor households get much lower benefit in terms of income compared to the households in top two deciles.

Appendices.

Appendix 1: Creating Variables for the Errors in Assignment Analysis

Initial eligibility for people living in poverty to join the program is based upon selection at a meeting of the village, which designates individuals in the lower two socioeconomic strata; but among those selected as potentially eligible ultra poor by the village, the NGO then selects participants according to three exclusion criteria and the presence of at least 3 out of 5 inclusion criteria (Noor et al 2004). The exclusion criteria are ECI (the individual is not a member of another NGO), EC2 (the individual is not a recipient of a government welfare food distribution program, and EC3 (there is no female able to work in the household.)

We created our own designation of those eligible using the survey data. To do so for the case of NGO membership we used the responses to: i) whether the person had NGO savings (variable "ngos" - selected 340 observations); ii) whether the person had a loan from NGO (variable "ngoln" - selected 64 observations); iii) whether the materials for the house wall and roof were provided by an NGO – tins1=3, (selected 32 observations); iv) whether the source of a loan was from an NGO (variable srln – selected 1 observation), and v) whether the individual was recorded as either a member of the BRAC Development Program or indicated as a member of more than one NGO (selecting 100 and 23 observations respectively). This classification selected 447 observations for the year 2002, of which 57 had been selected as SUP members for the program despite apparent ineligibility.

Exclusion criterion 2 was composed of the following variables: i) whether the person had government benefits (gprben1=2), which selected 30 observations; ii) whether main source of income was government benefits, in main source of income, for three primary sources (variables msoi1, msoi2, msoi3), which selected 3, 11 and 7 observations respectively. This classification selected 127 observations, of which 38 had been selected as SUP members for the program.

To create EC3 we used the variable disabl, those women who presented a disability. This selected 48 observations, of which 24 previously had been selected as SUP members. Overall, according to the exclusion criteria, we identified 116 participants who were selected despite being ineligible.

With respect to the inclusion criteria, the household had to meet at least three out of five conditions in order to be considered for the TUP program. They were: IC1: owning less than 10 decimals of land (a tenth of an acre), including homestead; IC2: no male income earner at home; IC3 children of school-age working; IC4: adult women of household selling labor outside homestead; and IC5: household having no productive assets.

With respect to the first inclusion criterion (ownership of less than 10 decimals of land, including for their homestead), we created a dummy variable for whether the household owns self cultivated land, own lands that others cultivate, own homestead land, or owns land that is uncultivated. This criterion selects (as eligible) 4624 out of the 5067 for the year 2002.

For the second inclusion criterion, no male income earner present at home, we first created a dummy variable for the presence of no male income earner at home, as the intersection of males of working age (more than 14 years old) that are not working. There are 66 observations that fulfill this criterion, of which 27 already had been selected as SUP. The second auxiliary variable constructed was a dummy for the presence of no male at home (additional to the previous one, no male earner). This variable selects 1893 observations, of which 1147 had been selected for SUP participation.

For the third inclusion variable, that school-age children are working, we used questionnaire data to that effect, which selected 167 observations, of which 81 had been selected as an SUP.

To parallel the fourth inclusion criterion, that there are adult women selling labor outside the homestead, we selected those observations for which the main source of income (for the first three primary occupations) were: 5 =daylabour (agriculture), 6=daylabour (non-agriculture), 7=small business/trading, 9=begging, 10=servant, 11=professional. This selected 1627 observations, of which 1047 had been already selected as SUP.

For the fifth inclusion criterion that the household had no productive assets, we used the dummy variable "prodasst", which selected 2791 observations, of which 1614 were already SUP members.

Finally, to construct the inclusion criteria, we consider those observations that fulfill at least three out of the five conditions. According to these data, there were 1760 observations that should have been classified as SUP, of which 647 were not.

Accordingly to the exclusion and inclusion criteria, we have created the following groups: SB_1 (selected as SUP, and fulfilling both inclusion and exclusion criteria), SNB_1 (selected as SUP, not fulfilling the criteria), SNB_0 (correctly not selected as SUP, criteria not met), and SB_0 (not selected as SUP, but fulfilling criteria).

References

(1)Aakvik, Arild, "Bounding a Matching Estimator: The Case of a Norwegian Training Program," Oxford Bulletin of Economics and Statistics, February 2001, v. 63, iss. 1, pp. 115-43

(2)Altonji, Joseph G., Elder, Todd E., and Taber, Christopher R., "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools," Journal of Political Economy, February 2005, v. 113, iss. 1, pp. 151-84

(3)Banerjee, A, E. Duflo, R. Shattopadhya, J. Mukhopadhya, J. Shapiro, 2008: Targeting Efficiency: How Well Can we Identify the Poor?, presented at CMFR, India.

(4)Becker, Sascha O. and Marco Caliendo, "Sensitivity analysis for average treatment effects," Stata Journal, Stata Corp LP, vol. 7(1), pages 71-83, 2007.

(5)Chronic Poverty Research Center, Chronic Poverty Report 2008-09, Manchester, 2008

(6)Deaton, Angus (1989), Looking for Boy-Girl Discrimination in Household Expenditure Data, World bank Economic Review, Januray, 1989.

(7)Emran, M. Shahe, A. K. M. Mahbub Morshed, and Joseph E. Stiglitz, "Microfinance and Missing Markets," working paper, George Washington University, 2007

(8)Emran, M. Shahe, Forhad Shilpi, and Joseph E. Stiglitz, "Microfinance, Property Rights and Ultra Poor," working paper, George Washington University, 2008

(9)DiPrete, T and Gangl, Markus, 2004, Assessing Bias in the Estimation of Causal Effects: Rosenbaum Bounds on Matching Estimators and Instrumental variables Estimation with Imperfect Instruments, Working Paper.

(10)Heckman, James, H. Ichimura, P. Todd, 1998, "Matching as an Econometric Evaluation Estimator", Review of Economic Studies, vol. 65, No. 2.

(11)IFPRI, "The World's Most Deprived. Characteristics and Causes of Extreme Poverty and Hunger," IFPRI Oct 2007

(12)Lipton, Michael (1986), 'Seasonality and Ultra poverty', Sussex, IDS Bulletin 17.3

(13)Matin, Imran, et al 2008, Crafting a Graduation Pathway for the Ultra poor: Lessons and Evidence from a BRAC Programme, RED Working Paper, BRAC.

(14)Noor, Marufia, et al, "Stories of Targeting: Process Documentation of Selecting the Ultra Poor for CFPR/TUP Programme", BRAC TUP Working Paper No. 1, 2004

(15)Rabbani, Mehnaz, Vivek A. Prakash, and Munshi Sulaiman, Impact Assessment of CFPR/TUP: A Descriptive Analysis Based on 2002-2005 Panel Data, BRAC TUP Working Paper No. 12, July 2006

(16)Ravallion, Martin, 2007, Evaluating Anti-Poverty Programs, World Bank.

(17)Sen, Binayak ,and David Hulme, eds., Chronic Poverty in Bangladesh: Tales of Ascent, Descent, Marginality, and Persistence, Dhaka: Bangladesh Institute of Development Studies, 2006

(18)Smith, Stephen C., Ending Global Poverty, Palgrave Macmillan 2005

(19)Song, Jae G. and Manchester, Joyce; , "New Evidence on Earnings and Benefit Claims Following Changes in the Retirement Earnings Test in 2000," Journal of Public

Economics, April 2007, v. 91, iss. 3-4, pp. 669-700

(20)Sulaiman, Munshi and Imran Matin, 2006: Targeting Effectiveness of CFPR/TUP in Scaled Up Environment, CFPR Working Paper Series 8, BRAC-RED.

(21)Todd, Petra, 2007, Matching Estimators, Entry for New Palgrave Dictionary of Economics, Second Edition.

(22)World Bank, World Development Report, Oxford: World Bank, 2006

(23)Yunus, Muhammad, Creating a World Without Poverty: Social Business and the Future of Capitalism, NY: Public Affairs, 2007