Almost random: Evaluating a large-scale randomized nutrition program in the presence of crossover

Sebastian Linnemayr[♦]
Harvard University

Harold Alderman[‡]
The World Bank

JOBMARKET PAPER

Abstract
Large-scale randomized interventions have the potential to uncover the causal effect of programs applying to a large population, thereby improving on the insights gained from currently dominant smaller randomized studies. However, the external validity gained through larger interventions typically comes at the cost of more pronounced deviation from the randomization plan that needs to be addressed in order to unleash the full potential of such studies. This article investigates the impact of the Nutrition Enhancement Program that aims to improve child nutrition in Senegal based on a large-scale randomized community intervention. We explicitly deal with deviation from the planned treatment and suggest different approaches for combining ex-post adjustments such as propensity score matching with the randomized treatment plan. We do not detect a strong overall program impact on the outcome measure of weight-for-age based on planned treatment status, but find an effect when using the alternative estimators of two-stage least-squares and propensity score matching. Children who benefitted from the program during their whole lives show a significant nutritional improvement, irrespective of the empirical method used. The findings of this article underscore the importance of addressing the shortcomings of large-scale randomization interventions in a systematic manner in order to understand the selection process that can guide further implementation of such projects, as well as to expose the true, causal effect of such programs.

Key words: Nutrition; Impact evaluation

JEL Codes: I12; O12

[♦] Corresponding author: Sebastian Linnemayr, Harvard School of Public Health, Harvard University, Cambridge, USA; email: slinnema@hsph.harvard.edu.
[‡] Harold Alderman, World Bank, Washington DC, USA; email: halderman@worldbank.org.

The authors would like to thank Stefan Dercon, David Evans, Erica Field, and seminar participants at the CSAE 2008, the PAA 2008, the ASHE 2008, and the World Bank Impact Evaluation Group Lunch Seminar for valuable feedback.

Acknowledgement of Data Organization and Collection; General Acknowledgement

The data for this nutrition intervention program have been organized and collected by the Research Centre for Human Development under the direction of its director Salif Ndiaye, whom we owe thanks. We also would like to thank Claudia Rokx as well as Biram Ndiaye, Coordinator of Cellule de Lutte contre la Malnutrition – Primature (PRN), Dakar during the period of this study, for collaboration on the design and implementation of the research. We would also like to thank the regional coordinators of the PRN and the regional organizers of the Forecasting and Statistics Directorate for their help. We gladly acknowledge the competence, enthusiasm, and availability of the members of the technical teams. The personnel on the ground, in particular the chauffeurs, interviewers, and team leaders have contributed significantly to the success of the survey. Last but not least we would like to thank the people who were interviewed for generously offering their time, energy, and commitment.

1. Introduction

The current study makes two main contributions to the literature evaluating randomized trials: first, we report the results of a large-scale nutrition intervention reaching 200,000 households, one of the first programs of such large scale. Second, the study illustrates the value of a randomized intervention even when there is departure from the initial design (Angrist et al., 2002). This is important since adherence to treatment status becomes increasingly difficult with the size of the project.[2] Nevertheless, large-scale randomized interventions are urgently needed, as the currently predominant relatively small-scale randomized interventions are frequently criticized for perceived lack of external validity. We illustrate that while the main policy result is robust to alternative approaches, the empirical consequences of assumptions common in the literature are not trivial. Given the need for accurate evaluation of large-scale interventions for scaling up, it is imperative to deal with deviations from initial treatment status and find ways to extract information from this situation. We hope that the current study contributes to this much-needed knowledge.

Malnutrition is a persistent problem in developing countries; WHO estimates the fraction of malnourished children in developing countries at 33% as measured by the percentage of children stunted, i.e. children that fall below -2 standard deviations of the United States National Center for Health Statistics / WHO international reference median value (de Onis, Frongillo, and Blössner 2000). This has far-reaching consequences: malnutrition *in utero* or in infancy can have a long-lasting negative impact on cognitive development and on their subsequent capacity to achieve sufficient income to provide for

[2] While close cooperation between the researcher and the implementation team on the ground reduces the deviation from the treatment assignment, it potentially also reduces the external validity of the study; such close supervision potentially also reduces the internal validity if the study objects feel closely observed.

their own children when adults (Alderman, Hoddinott, and Kinsey 2006; Victora et al., 2008).

There is substantial agreement on the efficacy of a number of nutrition interventions (for a review see Allen and Gillespie 2001). There is little doubt, for example, that breastfeeding promotion saves infant lives, or that vitamin A prophylaxis reduces child mortality. However, there is little consensus that community based programs can reduce stunting or being underweight (Bhutta et al., 2008, Alderman, 2007).

This study assesses the impact of a package of health inputs on the anthropometric status of children in three regions in rural Senegal as evidenced by their weight-for-age. The relatively short time between the baseline and follow-up survey motivates the focus on two further measures of program success: health inputs and nutritional knowledge of the mothers. The program's ability to influence these intermediary inputs not only helps explain the final outcomes but also indicates whether other health measures can be improved in the course of a program with a primary objective framed in terms of nutritional status. Furthermore, we stratify the sample by child age, i.e. whether the child was exposed to the program in utero as well as during her lifetime. The short intervention duration makes it likely that the youngest children show the largest gain from the program that for longer durations could be expected to also raise the nutritional status of older children.

The outline of the paper is as follows: in section 2, we describe the program. We then briefly summarize the identification strategies for the estimators in section 3 that we will use in the subsequent analysis. In section 4, we give an overview of the main variables used. Section 5 presents the results for nutritional status expressed as weight for age as well as looking at measures of behavioral change and the availability of health inputs based on the planned treatment status. Alternative estimators based on the actual receipt of the intervention are also presented. Section 6 concludes and discusses the differences in outcome between the approaches used to address imperfect randomization.

2. Description of the intervention and its implementation

2.1 Data setting

The government of Senegal designed a strategy in 2002 to fight malnutrition that is scheduled to reach 50% of children under five years of age by the year 2011, the *Programme de Renforcement de la Nutrition* – Nutrition Enhancement Program (PRN). The first phase of the Nutrition Enhancement Program targeted 20% of children under the age of five with growth promotion and the integrated service of child diseases at the community level. In the three regions under consideration (Fatick, Kaolack, and Kolda), child nutritional status is low and knowledge of best practices very limited: the percentage of children who are exclusively breastfed in the first four months of their lives varies between 1.1% in Kaolack and 2.6% in Kolda, even though the WHO recommends exclusive breastfeeding up to at least six months of age. Children in Senegal are also suffering from the lack of micronutrients that remains widespread despite interventions that have taken place in the past. 84% of children less than 5 years of age suffer from anemia, as do 61% of women. Vitamin A deficiency is a public health problem, with about 61% of children under the age of six years suffering from this deficiency. Summary statistics for the sample population are presented in Table 1. The z-scores of weight-for-age is more than one standard deviation lower than the mean for the US reference group and indicates the poor nutritional standard of the children in the sample:

for example, a two year old boy would on average weigh about 12.2 kg in the US; in Senegal, his average weight could be expected to be about 1.5 kg lower at about 10.8 kg (WHO Child Growth Standards).[3]

Over the course of the first wave of the PRN, monthly growth promotion was provided to 200,000 mothers and their children with NGO agencies contracted to supply these services. One of the main pillars of the approach was monthly discussion with mothers related to nutrition organized at the community (i.e. village or village-neighborhood) level (Alderman et al., 2008). Great care was put into involving communities and key figures within these communities such as village elders, the marabou (a religious leader), or grandmothers, who traditionally play a big role in influencing feeding and child care practices. The goal of targeting these people in the implementation strategy is on the one hand to involve the agents actually influencing behavior of mothers, and on the other hand to educate these people who will then pass on the knowledge, thereby supporting the sustainability of the project. Meetings of pregnant women were organized in order to generate a forum where these women can exchange ideas and experiences concerning pregnancy and child-rearing. Another strategy encouraged is the principle of 'positive deviation': individuals who show behavior different form that of the other villagers and who avoid certain health problems are invited to share their experience and to teach the other women these novel strategies.

The following program components were carried out in all three regions:

- Behavioral change towards exclusive breastfeeding without supplementation for at least the first six months of the child's life was promoted in discussion groups
- Growth promotion: during these sessions, the health worker weighs the child and discusses its progress with the mother by comparing it to a growth chart distributed

[3] These standards can be accessed at http://www.who.int/childgrowth/en/.

- Vitamin A supplementation: in the course of the weighing sessions, vitamin A is distributed to children 6 – 59 months and mothers in the 42 days after having given birth
- Iron supplementation: in discussion rounds, pregnant women are encouraged to take iron supplements that are distributed by health centers
- Bednets distribution for a fee (including a subsidy) and demonstration on their use
- Deworming was offered to all children aged 6 – 59 months.
- Cooking workshops were organized to demonstrate the preparation of nutritious foods for the mothers as well as supplements to breastfeeding after six months of age

2.2 Implementation of the randomized treatment status assignment

As part of the first phase of the PRN, a randomized treatment assignment was implemented in three poor rural regions in order to allow for subsequent evaluation before scaling up to the rest of the country's rural regions. The implementing NGOs were asked to provide a list of villages in which they had the means and intention to intervene. From the total list of about 1000 villages, 212 villages were randomly chosen in the three regions. Based on these villages, 220 clusters were built (as some villages are large enough to have 2 clusters, and one village had three clusters), and in each cluster up to 20 households were randomly drawn based on the list of households in the village. The nutrition intervention was randomly assigned to half of these villages; the NGOs were asked to schedule services to the other half in a later wave of implementation. They were free to include other villages not among the 212 in the intervention group at any time.

A baseline survey was conducted in April 2004 in all 212 villages, collecting data about the health status of the children, socioeconomic variables of the households these children are residing in, and extensive information about the nutrition and child care practices of the mother. The survey teams administered three questionnaires: a village questionnaire, a household questionnaire, and an individual questionnaire for the mother of the child. If there was no child under three years of age in the household, the household was dropped in the first round without replacement. In the second round, these households were replaced with other randomly drawn households from the same village list until the fixed number of households was interviewed. This change in the survey design explains the significantly larger number of children measured in the second round. In June 2006 the same information as at baseline was collected in control and treatment villages.

3. Strategy for empirical analysis

When evaluating the effectiveness of a treatment *T*, we would like to compare the difference *D* in the outcome variable of interest *Y* for the same individual *i* once he receives the treatment and once when he does not[4]:

(1) $D = Y_i^T - Y_i^C$,

where the superscript *T* denotes an individual receiving the treatment, and *C* stands for the outcome without the treatment. As we cannot observe the same individual or unit in two states of the world at the same time we face the so-called problem of the missing counterfactual. However, it may be possible to discern the average effect of a certain intervention on a group of individuals:

(2) $E\left[Y_i^T - Y_i^C\right]$.

[4] Key references on this topic include Duflo, Glennerster, and Kremer (2006) or Angrist and Krueger (1999). For the original reference for the Rubin causal model see Rubin (1974).

When subtracting and adding the unobserved but typically well-defined term $E[Y_i^C \mid T]$, i.e. the outcome of the treatment group in the absence of treatment, we can state the evaluation problem as the situation in which the total change in the outcome consists of the treatment effect and the selection bias that confounds causal identification:

(3) $D = E[Y_i^T | T] - E[Y_i^C | T] + E[Y_i^C | T] - E[Y_i^C | C]$.

Much of empirical work is concerned with finding ways to control for selection bias, the difference in the non-treatment outcome between treatment and control individuals. The challenge is to establish a close estimate of the missing observation of the non-treatment outcome of the treatment group.

Randomization guarantees on average for a large sample that in the absence of the intervention the control and treatment groups have the same outcome:

(4) $E[Y_i^C | T] - E[Y_i^C | C] = 0$,

In such a situation, a simple comparison of the sample post-intervention means suffices to measure the average treatment effect of the intervention. In terms of regression analysis one can regress the outcome on covariates and a dummy variable for inclusion in the treatment group:

(5) $Y_{it} = X_{it}\beta + T_t\delta + e_{it}$,

where e is an error term composed of individual, family and community unobserved fixed characteristics as well as a stochastic disturbance term, μ_{it}:

(6) $e_{it} = \nu_i + \eta_i + \varepsilon_i + \mu_{it}$.

In small samples, there is the possibility that villages differ in their characteristics influencing outcomes. Therefore, it is common to include socioeconomic variables *X* at the individual and household level that in previous studies have been shown to influence the outcome of interest (see for example Behrman and Skoufias 2004).

Random assignment is, however, not without its pitfalls. For example, individuals selected for the treatment may not take it up, so that the intention to treat does not provide an accurate assessment of the impact of the treatment on the treated. Or individuals assigned to the control group obtain the service from an alternative source, say a private provider. This is often called a crossover effect. Angrist et al. (2002) provide an illustration in which some individuals who received a randomly assigned school voucher did not utilize it and others in the control group received a functionally similar scholarship from private groups.

A variation of the cross over problem occurs when the implementation does not follow the program assignment strictly, as occurred in the current study. In such a situation, planned and actual treatment status differ, with the intent-to-treat estimator (as defined by the researcher) representing a lower bound on the evaluation estimate. Another problem that can lead to a mitigated program impact is spillovers from treatment to control villages, for example, when information being disseminated during the intervention is shared. In such a situation, we would expect the outcome to improve for both groups, but more so for the treatment units given their longer and more intense exposure to the program.

One approach to the problem of comparability of control and treatment group that is often used is to employ a difference in difference method in the context of panel data. By construction, the fixed effects remove the corresponding fixed component of the error term and thus any correlation between it and the treatment variable T. This simple problem can be implemented in a regression set-up with two data-waves such as:

(7) $Y_{it} = \alpha + \beta \cdot 1(i \in T) + \gamma \cdot 1(t = 2) + \delta \cdot 1(t = 2) * 1(i \in T) + \varepsilon_{it}$,

where we control with the second term on the right hand side for initial differences between the control and the treatment group, the third term controls for a time-trend common to both groups, and the fourth term indicates the treatment effect of actually receiving the intervention. The estimate of interest is *(δ-β)* which measures the treatment effect purged of initial differences under the assumption that in the absence of treatment both groups would experience a similar trend in the outcome variable. When there is deviation from planned treatment it is necessary to assume that these are not correlated with responsiveness to the program that is not removed by differencing.

Moreover, in cases in which the assignment is based on observed values of the outcome desired – say, where the treatment is prioritized to groups with low test scores or nutritional status, for example –it is likely that $[E(\mu_{it} \mid T) \neq 0]$ since measurement error partially determines the assignment. This might be the case in our sample; as discussed further below, villages with lower initial nutritional status had a higher probability to be included in the treatment group. If this reflects fixed effects, then difference in difference will address the bias, but if it reflects time varying factors, including measurement errors, then difference in difference will not solve the bias of reversion to the mean. Chay et al. (2005) present an example of an assignment to treatment based on baseline performance where difference-in-differences results are biased.

For the analysis at hand, the treatment group includes about 30% of households that did not receive the intervention and the intent-to-treat estimator can be viewed as a lower bound for the magnitude of the nutrition intervention. However, the potential bias can be addressed by using the planned treatment status as an instrument for actual receipt of the intervention. Planned treatment status is exogenous by construction yet being a strong predictor of actual treatment fulfills the requirements for instrumental variables. In the first stage, planned treatment status and other village-level variables Z are used as instruments for actual receipt of the intervention:

(8) $T_{it} = X_{it}\alpha + Z_{it}\gamma + u_{it}$.

In the second stage, the fitted value of T is used in the regression:

(9) $Y_{it} = X_{it}\beta + \hat{T}_{it}\,\delta + e_{it}$.

Another way of addressing imperfect execution of random assignment is to use propensity score matching (Rosenbaum and Rubin,1983). If the potential treatment and control outcomes are independent when conditioning on some observable characteristics, then the outcomes are also independent conditional on the propensity score. Smith and Todd (2005) review the performance of matching estimators and come to the conclusion that for their sample a difference-in-difference matching estimator performs best, motivating the use of this estimator in the current study. Hirano et al. (2003) suggest weighting the observations by a function of the estimated propensity score to arrive at a more efficient estimator. Their approach can be implemented by using the weight of $1/\hat{P}$ for villages having actually received the treatment, and $1/(1-\hat{P})$ for both control villages and treatment villages not having received the intervention.

The planned treatment can be used in estimating the weights. This combination of approaches also addresses a problem that is discussed in Heckman et al. (1998), that of common support in matching approaches. When the major explanation for the selection into treatment remains the planned treatment status which was random, there is ample overlap between the treatment and control villages. Additionally, the combination of propensity score estimates with difference in difference helps address the possibility that the assumption of time independent selection does not hold (Ravallion, 2008).

Several studies have used both prospective and retrospective approaches in order to evaluate the degree of selection bias to which observational studies may be subjected. For example, Lalonde (1986) finds that there are often significant differences between prospective and retrospective empirical approaches for the evaluation of programs as does Glewwe et al. (2004).[5] The current paper aims to contribute to this discussion by employing alternative approaches using the same dataset.

4. Empirical Analysis

4.1. Measures of program success used

The main focus of the evaluation of the current nutrition intervention is to assess the impact of the set of services offered on the nutritional status of children. When preparing the intervention, the outcome measure of weight-for-age z-score was determined as the original indicator to determine the sample size of the program as well as the indicator for tracking success of the program on the part of the organizations financing the PRN[6].

The outcome of interest reflects a package of services which are valued not only for their impact on weight but also as indicators of the functioning of a community health program in general. For this reason, we also examine the availability of health care measures before and after birth such as micronutrient supplementation or malaria bednets. The analysis of these measures can provide supportive evidence of program success (in the sense that the health inputs reached the villages/households) that are less

[5] For overviews see Duflo, Glennerster, and Kremer (2006) and Ravallion (2008).
[6] Apart from weight-for-age, we also investigated the impact of the program on height-for-age, another frequently used anthropometric measure that captures more long-term impacts. The results obtained for this measure are virtually identical to the ones for weight-for-age and are therefore not presented to reduce the number of output tables. These omitted tables are available from the authors.

prone to measurement error, rely on faster-moving measures, and indicate potential future success of the program if it takes time to transform these input measures into measurable change in the outcome z-scores. This aspect adds an important dimension to the analysis as measurement error has been cited to be a common problem for the analysis of household surveys involving health measures (Deaton 1997).

The unit of assignment in the current study is the village and not the individual; all households in a village belong either to the control or the treatment group. As a result, actual take-up of the program by individuals is not observed, although the village health workers tried to encourage all mothers in the village to participate in the program. Therefore, it is the impact of the availability of the program in a village rather than its actual take-up that is evaluated. As actual take-up is a choice variable, investigating the effects of mere availability of the program may be less prone to selection bias (Strauss 1990).

In addition to investigating health inputs as well as health outcomes, we also try to take potential heterogeneity of the program impact into account. In particular, we test whether younger children have a different response than the older cohort. Given that the program was available for less than two years in the villages, we test the possibility that the children that were exposed to the intervention already *in utero* have greater benefits. This would be in keeping with evidence that malnutrition occurs at very early ages and is fairly unresponsive by 18 months (Shrimpton et al. 2001). The effectiveness of the program may also depend on village initial infrastructure, a possibility we take into account in the analysis below by stratifying by whether the village has a road to the next village that can be used all year round, and also along the average village wealth at baseline.

4.2. Control of the success of the randomization of planned treatment status

In the next step we investigate whether the assignment of treatment status to the villages happened in a random fashion, in which case the outcome variables as well as the conditioning variables should not differ between the treatment and the control group in the baseline period except for random deviations (Behrman and Todd 1999). The comparison of villages along their planned treatment status in Table 2 shows that the two types of villages show largely similar outcome and control variables.

As discussed above, while the initial assignment of treatment was random, the NGOs implementing the program may not have fully understood the importance of the randomization procedure prescribed, or deviated from the planned treatment status of villages for other reasons. In Table 3, we see that there was considerable deviation of actual treatment status from the status initially assigned. Of the 111 villages that were initially assigned treatment status, 80 ended up receiving treatment, while 31 villages (28%) did not receive the intervention. Of the 100 initial control villages, for one village no second round data were available, and of the 99 remaining villages, 8% received the intervention despite their control status. In the analysis to follow, we therefore have to address the problem of partial compliance.

When we compare the most important group of villages that deviated from the original design, the 31 villages that initially were assigned treatment status but that did subsequently not receive intervention along some key dimensions with the remaining 80 villages from the group initially assigned treatment status, there is some evidence that the deviation did not happen randomly (Table 4). The villages from the 2004 planned treatment list that subsequently did not receive the nutrition intervention were initially somewhat better off in terms of their nutritional status, and had less children that were mildly underweight (below -2 standard deviations from the mean of the US reference population). They more often had a market, and showed a lower presence of NGOs and health posts than the villages that retained their treatment status. These statistics indicate the possibility that the NGOs purposely selected villages by focusing on the worst-off villages in the treatment sample and on those villages in which there was already a NGO present (potentially the intervening NGO itself).

Basing the analysis on actual rather than planned treatment-status could lead to a systematic underestimation of the treatment status if, by targeting the worst-off villages, the service providers gave priority to villages that were less able to profit to the same degree from the nutrition intervention program compared to their better-off peers. On the other hand, if the selection was based on short term fluctuations in nutritional status, we might see an improvement in these villages that was partly due to a reversion to the mean as discussed in section 3. This would tend to overestimate the positive effect of the intervention. While selection on fixed community characteristics can be addressed using difference-in-differences analysis that is employed, as mentioned, this step would not necessarily be the case if the selection criteria included time varying factors. A similar caveat applies for the propensity score estimator that makes the assumption that the selection bias is time-invariant (and can hence be removed in a panel context) conditional on observables. Below, we will contrast the outcomes of the analyses based on planned and actual treatment, using both the intent-to-treat and a 2SLS estimator using planned treatment status as an instrument; for the analysis based on actual treatment status we focus on treatment-on-the-treated as well as a matching estimator combined with difference-in-differences that compares villages with similar probability of having been selected into treatment.

5. Results

Table 5 in columns 1 and 2 presents the results for weight-for-age using the original randomization classification in 2004, irrespective of whether the village actually received the program. The approach uses both the baseline and subsequent data and, thus, combines difference in difference with random assignment. The coefficient for the variable *planned intervention* indicates the difference of the mean value for the villages that were initially assigned treatment status from those expected to be in the control group. We include age dummies of the children in six months age groups in columns 1 and 3, with the children between 30 months and 3 years of age representing the omitted group. The original treatment status assigned to the village irrespective of actual receipt of the program is an imperfect indicator of services actually delivered. As such, using the planned assignment status avoids a correlation with unobserved factors at the possible expense of errors in variable from mismeasurement from which an attenuation bias is expected.

Based on the two assumptions that in the absence of treatment, the villages in both groups would experience a similar trend in malnutrition rates and also that selection was not based on the time varying factors such as the level of the outcome at time the baseline was implemented a difference in difference framework using the actual receipt of treatment would give an accurate assessment of the impact of the services. This approach, however, could introduce a bias and, unlike the bias due to attenuation with the planned treatment, the direction of such a possible bias is unknown. We report the outcome based on the assumption of a time-invariant selection bias in column 3. In all specifications, we allow for the clustering of standard errors at the village level.

For the specification based on planned treatment status in column (1), we do not find a statistically significant impact in the program villages in comparison to the control villages.[7] In contrast, when basing the analysis on actual treatment status, the program is found to increase weight-for-age of children 0-36 months by on tenth of a standard deviation on average. The age dummies (omitted for space reasons) reflect the common finding of a deterioration of the nutritional situation for the children with increasing age when compared to the children in the US reference group. The other control variables mirror the findings in previous studies on the determinants of child nutrition: parent's education and sanitary facilities in the household improve the nutritional status, while the status of being a twin reduces it significantly. The gender dummy is insignificant, as discrimination by gender is typically not observed in Africa (Svedberg, 1990). As indicated by the indicator variable '*second round',* the villages in the sample experience an overall increase in the weight-for-age indicator of about one tenth of a standard deviation for both types of villages that is statistically significant at the 1% level. This finding confirms the general trend observed in the summary statistics in Table 1.

There are several reasons why the program impact may differ by the age at which the children were exposed to treatment (Alderman, 2007). Children who at the time of the baseline survey in April 2004 were six months old were included in the second wave in 2006 although these children likely were weaned by the time the intervention began. In contrast, a child born after April 2004 would have had the additional benefit of their mothers participating in the discussion groups and micronutrient provision for pregnant women, two important program components. There is increasing evidence that the experiences *in utero* can have long-lasting effects (Behrman and Rosenzweig 2004;

[7] As the results for mild and severe malnutrition in almost all cases go in the same direction as the average weight-for-age score, we omit in the subsequent result tables the regressions for mild and severe malnutrition and focus on mild malnutrition only. The omitted output tables are available from the authors.

Strauss 2000). Additionally, these children would benefit from the advice to the mothers on the use of colostrum as well as on exclusive breastfeeding.

We therefore create a dummy variable termed 'full exposure' for children up to 6 months of age as these children have experienced the program impact during their entire life as well as *in utero*. The results are presented in columns 2 (planned treatment status) and 4 (actual receipt of the intervention) of Table 5. It appears that the youngest children, i.e. those whose mothers benefited from the program when they were pregnant, benefitted from the intervention even though the average impact on children less than three years of age was virtually zero.

However, as discussed, the results in Table 5 are likely an underestimate of the true impact if based on planned treatment yet also possibly subject to reversion to the mean if based on actual treatment status. Thus, we instrument actual receipt of the treatment with planned treatment status that fulfills both criteria for a good instrument: it is both of exogenous character (by construction), and it is highly correlated with the actual receipt of the intervention. In addition, to the planned treatment, we add initial village-level characteristics such as distance to the next village, prevalence of female education, or the presence of a market that may have influenced the NGOs' placement decision as well as five interaction of village characteristics and the planned treatment. Each of these interactions as well as the treatment status itself are individually statistically significant at $p<.01$. The r-square of the instrumenting equation is 0.72.

The instrumented treatment status in Table 6 does not have a statistically significant impact on child nutritional status using the full sample, although the sign of the coefficient has changed from negative as in Table 5 column 1 to positive. In addition, in column 2 we stratify by the age of the child, and find again that the youngest children do benefit substantially from the program. As the variables used for instrumentation of actual receipt of the treatment do not show variation at the individual level, for this specification we performed the 2SLS 'by hand', interacted the instrumented treatment status with the age variable of the child, and then derived the standard errors using the method of bootstrap using 100 repetitions. Another indicator of program success comes about when we stratify the sample according to initial conditions: we observe a positive and statistically significant in villages that are deprived of a road connecting them to the next village that is useable all year round. In villages connected with such a road, there is no discernible effect of the program when using planned treatment status as an instrument in the first stage regression. A similar, albeit statistically only borderline significant coefficient (p-value: .11) is observed for villages below the mean of the wealth index created from the possession of physical assets and livestock. For villages over that cut-off, instrumented planned treatment status is not statistically different from zero.

The outcome variable weight-for-age is a function of behavioral inputs as well as health and time inputs. These latter indicators may change in response to the intervention relatively quickly and potentially translate into nutritional improvements with a delay. Health inputs such as drugs or bednets distributed also indicate whether the NGOs actually delivered the services required to the villages. The results for these inputs into the health production function are presented in Table 7, which is, again, an intention-to-treat estimator based on the planned treatment status.[8]

The results in rows 1 and 2 indicate that the wide-spread practices of giving liquids other than breastmilk in the child's first six months as well as the practice of not giving the colostrum after birth are less prevalent in planned intervention villages than in initial control villages in 2006, despite there not having been a significant difference in

[8] The variable definitions for Table 7 are given in the appendix table A.1.

the prevalence between the groups in 2004. When expressing the coefficient in terms of marginal probabilities, we find that there was a reduction in the probability of giving liquids other than breast milk in the first three days following birth of 11% in the treatment group as compared to the control group.

For health inputs, two out of three measures we observe a statistically significant impact of planned treatment status. For the provision of bednets, the coefficient is not significant at the 10% level. Similar results are found for the provision of micronutrients: for vitamin A for infants and iron supplements for pregnant mothers, there is a statistically significant impact of being in a planned treatment village for vitamin A, and a borderline significant effect for iron supplementation. The last two rows in Table 7 show that disease prevalence is not affected by planned treatment status. This finding is not surprising as disease prevalence is likely to be correlated with the outcome measure of weight-for-age, for which we also found no significant impact between the planned treatment and control villages for the pooled sample.

The results for behavioral indicators and health inputs based on actual instead of planned treatment status confirm the findings in Table 7 and are omitted for space reasons. The above results indicate that the NGOs delivered health inputs and improved the knowledge of best practices in the planned treatment villages. Unfortunately, these behavioral and input changes did not translate into changes in the prevalence of diseases such as diarrhea that count among the main reasons for the low nutritional status of children in Senegal.

As discussed above, the planned treatment status can not only be used as an instrument but can also improve the overlap of the support for the propensity score for villages receiving the treatment and those that do not benefit from the intervention in conjunction with difference in difference estimates. Table 8 reports such an analysis following the approach in Hirano, Imbens, and Ridder (2003) and can be interpreted as the impact of actual receipt of the intervention for villages similar in their propensity score.[9] The results indicate coefficients that are statistically significant at the 5% level and sizeable in magnitude, with about a .27 standard deviation increase in weight-for-age score. Somewhat surprisingly, we do not find that younger children benefit relatively more from the intervention than their older peers. However, given that we find an average positive effect on weight-for-age for all age groups, the extra effect for young children may be incorporated in the average impact in the villages.

6. Conclusion

The aim of the current study is to evaluate the success of a pilot program forming part of the *Programme de Renforcement de la Nutrition*, a nutrition intervention program targeted at young children in Senegal that introduces the program components to three poor rural regions. Identification of the treatment effects is based on the random assignment of the treatment status among 212 villages in April 2004 before receiving the intervention and being re-surveyed in June 2006. However, given substantial deviation from the assigned treatment status, we compare these results with approaches based on the actual receipt of the intervention. The planned treatment status is used both as an instrument that is plausibly exogenous by construction, but can also be used as an input into the propensity score in a matching approach.

We find significant changes in health care practices in the villages assigned to the treatment status. But using this assignment as an indicator of treatment, we do not find

[9] Following common procedure in the literature, we constrict our sample to observations with a propensity score lying in the interval [.05; .95].

an average overall impact on weight-for-age of children. We do, however, observe that those children whose mothers benefit from the intervention during their pregnancy display a significantly improved nutritional status than their older peers who were likely weaned before the program began. These observations can guide the allocation of resources in similar programs.

However, while these core results give an indication of the project's success, the magnitude of the impact is biased downwards due to cross over effects. Thus, we also report results using differences in differences based on the actual treatment status as of June 2006 instead of the planned one from 2004 as well as an instrumental variables approach based on the planned treatment assignment. Both approaches to evaluating the study tend to results in larger estimates of the treatment effects compared to the results based on planned treatment status alone. These results shed more light on the impact of a large-scale nutrition intervention in a situation where adherence to the assigned treatment status is less than perfect.

References

Alderman H. Improving Nutrition through Community Growth Promotion: Longitudinal Study of the Nutrition and Early Child Development Program in Uganda. World Development 2007;35; 1376-1389.

Alderman H, Hoddinott J, Kinsey B. Long term consequences of early childhood malnutrition. Oxford Economic Papers 2006;58; 450-474.

Alderman H et al. Effectiveness of a community-based intervention to improve nutrition in young children in Senegal: a difference in difference analysis. Public Health Nutrition 2008, forthcoming.

Allen L, Gillespie S. 2001. What Works? A Review of the Efficacy and Effectiveness of Nutrition Interventions. ACC/SCN Nutrition Policy Paper No. 19 2001; United Nations Administrative Committee on Coordination Sub-Committee on Nutrition.

Angrist J, Bettinger E, Bloom E, King E, Kremer M. Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment. American Economic Review 2002;92; 1535-1558.

Angrist J, Krueger A. Instrumental variables and the search for identification: from supply and demand to natural experiments. Journal of Economic Perspectives 2001;15; 69-85.

- 1999. Empirical Strategies in Labor Economics. In: Ashenfelter O, Card D (Eds) Handbook of Labor Economics, vol. 3A. North Holland: Amsterdam; 1999., 1277-1366.

Behrman J, Rosenzweig M. The Returns to Birth Weight. Review of Economics and Statistics 2004;86; 586-601.

Behrman J, Skoufias E. Correlates and determinants of child anthropometrics in Latin America: background and overview of the symposium.Economics and Human Biology 2004;2; 335-51.

Behrman J, Todd P. Randomness in the Experimental Samples of PROGRESA (Education, Health, and Nutrition Program). International Food Policy Research Institute Research Report, March 26, 1999.

Bhutta ZA, et al. What works. Interventions for maternal and child undernutrition and survival. Lancet 2008;371; 417-440.

Chay K, McEwan P, Urquiola M. The Central Role of Noise in Evaluating Interventions that use Test Scores to Rank Schools. American Economic Review 2005;95; 1237-1258.

De Onis M, Frongillo E, Blössner M. Is malnutrition declining? An analysis of changes in levels of child malnutrition since 1980. Bulletin of the World Health Organization 2000;78; 1222-1233.

Deaton A. The Analysis of Household Surveys. World Bank Publications: Washington DC; 1997.

Duflo E. Field Experiments in Development Economics. Paper prepared for the World Congress of the Econometric Society, January 2006.

Duflo E, Glennerster R, Kremer M. Using Randomization in Development Economics Research: A Toolkit. Centre for Economic Policy Research Discussion Paper No. 6059. 2006.

Glewwe P, Kremer M, Moulin S, Zitzewitz E. Retrospective vs. Prospective Analyses of School Inputs: the Case of Flip Charts in Kenya. Journal of Development Economics 2004;74; 251-268.

Hirano, K., Imbens, G., Ridder, R. Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score. Econometrica 2003:71; 1161-1189.

Lalonde R. Evaluating the Econometric Evaluations of Training Programs with Experimental Data. American Economic Review 1986;76; 604-620.

Ravallion, M. Evaluating Anti-Poverty Programs. Chapter 59 in Handbook of Development Economics. Elsevier Press. 2008:4.

Rubin D. Estimating Causal Effects of Treatments in Randomized and Non-randomized Studies. Journal of Educational Psychology 1974;66; 688-701.

Shrimpton R, Victora C, de Onis M, Costa Lima R, Blössner M, Clugston G. Worldwide Timing of Growth Faltering: Implications for Nutritional Interventions. Pediatrics 2001;107; 75-81.

Smith J, Todd P. Does Matching Overcome Lalonde's Critique of Nonexperimental Estimators?. Journal of Econometrics 2005;125:305-353.

Strauss J. Households, communities and preschool children's nutrition outcomes: Evidence from rural Cote d'Ivoire. Economic Development and Cultural Change 1990;36; 231-262.

Strauss R. Adult Functional Outcome of Those Born Small for Gestational Age. Journal of the American Medical Association 2000;283; 625-632.

Svedberg P. Undernutrition in Sub-Saharan Africa: is there a gender bias?. Journal of Development Studies 1990;26: 469-486.

Victora C, Adair L, Fall C et al. Maternal and child undernutrition: consequences for adult health and human capital. Lancet 2008;371; 340-357.

Table 1: Summary Statistics of socioeconomic variables in 2004 and 2006

	2004		2006	
Continuous variables	Mean	Standard Deviation	Mean	Standard Deviation
Weight-for-age	-1.317	1.417	-1.210	1.415
Categorical variables				
Male Dummy	.511	.500	.511	.500
Age 0-5 months	.195	.395	.178	.383
Age 6-11 months	.170	.376	.172	.378
Age 12-17 months	.190	.392	.196	.397
Age 18-23 months	.145	.352	.148	.355
Age 24-29 months	.187	.390	.153	.360
Age 30-35 months	.113	.317	.121	.326
Mother primary schooling	.143	.351	.173	.378
Mother secondary schooling	.023	.149	.033	.180
Husband primary schooling	.121	.326	.132	.339
Husband secondary schooling	.072	.259	.063	.243
Household size	14.889	8.483	14.294	7.195
Access to tap water	.372	.483	.215	.411
Water Closet	.121	.326	.064	.244
NGO in village	.673	.470	.810	.394
Healthpost in village	.313	.465	.286	.453
# of observations	4296		6144	

Table 2: Comparison of control and treatment villages along key dimensions at baseline

Village status	Planned Treatment Group	Planned Control Group	p-value
# of villages (number of children in sample)	111 (2321)	100 (1975)	
Weight-for-age in 2004	-1.352	-1.276	.265
Took iron supplements	.845	.846	.971
Took malaria medication	.828	.830	.931
Early introduction of liquids	.782	.791	.772
Took vitamin A during pregnancy	.617	.593	.423
Child had diarrhea in last two weeks	.333	.337	.849
Child received oral rehydration solution	.056	.042	.090
Child received deworming medicine	.073	.073	.990
Early introduction of solid foods	.162	.167	.807
Household has bednets	.390	.406	.693

Table 3: Planned versus actual treatment status of villages

		Realised Status 0	1	Total
Planned	**0**	91 (92%)	8 (8%)	**99**
Status	**1**	31 (28%)	80 (72%)	**111**
		122	**88**	

Table 4: Comparison of planned treatment villages along actual intervention status at baseline

Village status	Planned Treatment, treatment received	Planned Treatment, no treatment received	p-value
# of villages (number of children in sample)	80 (1733)	31 (588)	
Weight-for-age in 2004	-1.397	-1 .220	.102
% of children under -2SD wfa	.330	.267	.038
% villages with a market	.175	.323	.092
Road impassable	.263	.290	.770
NGO active in 2004	.788	.581	.028
Health post in 2004	.363	.290	.477

Table 5: Analysis of weight-for-age using planned and actual treatment status

	(1) Planned treatment status	(2) Planned treatment status, age interaction	(3) Actual treatment status	(4) Actual treatment status, age interaction
Second round * planned intervention	-0.02	-0.06	-	-
	(0.061)	(0.065)		
Second round * actual intervention	-	-	0.112*	0.074
			(0.062)	(0.066)
Second round	0.126***	0.113**	0.066	0.058
	(0.046)	(0.047)	(0.040)	(0.041)
Full exposure * planned intervention	-	0.242***	-	-
		(0.078)		
Full exposure * actual intervention	-	-	-	0.177**
				(0.081)
Full exposure	-	1.292***	-	1.321***
		(0.05)		(0.048)
Male Child	-0.019	-0.014	-0.019	-0.014
	(0.028)	(0.027)	(0.028)	(0.027)
Twin	-0.724***	-0.708***	-0.726***	-0.713***
	(0.106)	(0.103)	(0.106)	(0.104)
Primary education female	0.078*	0.088**	0.079**	0.088**
	(0.040)	(0.041)	(0.040)	(0.041)
Secondary education female	0.068	0.075	0.066	0.071
	(0.091)	(0.089)	(0.092)	(0.089)
Primary education male	0.011	0.008	0.011	0.007
	(0.042)	(0.043)	(0.042)	(0.043)
Secondary education male	0.159***	0.153***	0.161***	0.155***
	(0.057)	(0.058)	(0.057)	(0.058)
Husband education missing	0.018	0.027	0.018	0.028
	(0.042)	(0.041)	(0.042)	(0.041)
Tapwater	0.027	0.034	0.028	0.035
	(0.038)	(0.038)	(0.038)	(0.038)
Watercloset	0.047	0.05	0.046	0.05
	(0.052)	(0.051)	(0.052)	(0.051)
Wealth index	0.011	0.008	0.01	0.007
	(0.012)	(0.012)	(0.012)	(0.012)
Constant	-1.385***	-1.573***	-1.382***	-1.577***
	(0.046)	(0.031)	(0.045)	(0.031)
Observations	10127	10127	10127	10127
Number of villages	211	211	211	211
R^2	0.19	0.16	0.19	0.16

Notes: Absolute value of standard errors below the coefficient estimates. * indicates significance at 10% level; ** at 5% level and *** significant at 1% level of confidence. Standard errors corrected for clustering at the village level.

Table 6: Two stage least squares for weight-for-age using planned treatment status as an instrument

	(1) Pooled sample	(2) Pooled sample, age interaction	(3) Bad road	(4) Good road	(5) Low average wealth	(6) High average wealth
Second round * actual intervention (instrumented)	0.024	-0.054	0.274**	-0.103	0.006	0.177
	(0.073)	(0.091)	(0.112)	(0.093)	(0.086)	(0.114)
Second round	0.106**	0.106**	-0.025	0.184***	0.065	0.100
	(0.041)	(0.050)	(0.063)	(0.054)	(0.052)	(0.062)
Full exposure * actual intervention (instrumented)	-	0.310***	-	-	-	-
		(0.098)				
Full exposure	-	1.290***	-	-	-	-
		(0.055)				
Male Child	-0.023	-0.018	0.030	-0.049	-0.030	-0.016
	(0.025)	(0.026)	(0.043)	(0.031)	(0.035)	(0.036)
Twin	-0.724***	-0.707***	-0.553***	-0.827***	-0.639***	-0.810***
	(0.081)	(0.099)	(0.134)	(0.102)	(0.118)	(0.113)
Primary education female	0.077**	0.087*	0.087	0.067	0.135**	0.030
	(0.037)	(0.045)	(0.069)	(0.045)	(0.057)	(0.050)
Secondary education female	0.073	0.083	-0.068	0.125	-0.028	0.101
	(0.079)	(0.091)	(0.150)	(0.093)	(0.143)	(0.096)
Primary education male	0.004	0.001	0.018	0.002	-0.081	0.062
	(0.042)	(0.047)	(0.075)	(0.050)	(0.064)	(0.055)
Secondary education male	0.155***	0.148**	0.065	0.193***	0.244**	0.136**
	(0.056)	(0.062)	(0.107)	(0.065)	(0.102)	(0.068)
Husband education missing	0.014	0.024	0.024	0.008	0.009	0.025
	(0.042)	(0.038)	(0.078)	(0.050)	(0.066)	(0.056)
Tapwater	0.029	0.035	-0.085	0.085*	0.012	0.066
	(0.037)	(0.043)	(0.064)	(0.045)	(0.050)	(0.054)
Watercloset	0.055	0.060	0.069	0.047	0.052	0.065
	(0.050)	(0.051)	(0.091)	(0.060)	(0.072)	(0.069)
Wealth index	0.010	0.008	-0.001	0.016	0.013	-0.002
	(0.010)	(0.011)	(0.016)	(0.013)	(0.012)	(0.017)
Constant	-1.383***	-1.571***	-1.383***	-1.382***	-1.395***	-1.386***
	(0.045)	(0.044)	-0.076	(0.056)	(0.062)	(0.066)
Observations	10043	10043	3436	6607	5065	4978
Number of villages	211	211	69	142	97	114
R^2	0.19	0.16	0.17	0.20	0.19	0.19

Notes: Absolute value of standard errors below the coefficient estimates. * indicates significance at 10% level; ** at 5% level and *** significant at 1% level of confidence. Standard errors corrected for clustering at the village level. The standard errors in column 2 were derived using the bootstrap method.

Table 7: Health inputs: logit estimation based on planned treatment status

		Coefficient	S.E.
Behavioral Change			
Early introd. of liquids	Second round	-0.859 ***	(0.083)
	Planned Treatment	-0.412 ***	(0.108)
	# of obs.	10318	
	p-value	0.000	
Should give colostrum	Second round	0.548 ***	(0.068)
	Planned Treatment	0.508 ***	(0.094)
	# of obs.	10283	
	p-value	0.000	
Physical Health Inputs			
Worm drugs	Second round	0.858 ***	(0.108)
	Planned Treatment	0.804 ***	(0.143)
	# of obs.	9987	
	p-value	0.000	
Bednets	Second round	1.310 ***	(0.074)
	Planned Treatment	0.190 *	(0.099)
	# of obs.	10297	
	p-value	0.000	
Malaria pills	Second round	-0.109	(0.084)
	Planned Treatment	0.369 ***	(0.115)
	# of obs.	10098	
	p-value	0.000	
Vitamin A	Second round	-1.114 ***	(0.066)
	Planned Treatment	0.182 **	(0.088)
	# of obs.	10328	
	p-value	0.000	
Iron supplement	Second round	0.323 ***	(0.094)
	Planned Treatment	0.310 **	(0.127)
	# of obs.	9958	
	p-value	0.000	
Disease Incidence			
Diarrhea	Second round	-0.159 **	(0.067)
	Planned Treatment	-0.122	(0.090)
	# of obs.	10328	
	p-value	0.000	
Cough	Second round	-0.266 ***	(0.063)
	Planned Treatment	-0.104	(0.085)
	# of obs.	10328	
	p-value	0.000	

Notes: The results were derived using the same control variables as in Table 5 that are not presented for space reasons.

Table 8: Weight-for-age: propensity score matching combined with difference-in-differences

	(1)	(2)	(3)
	PS including planned treatment status	as in (1) but excluding extreme 5%	as in (2) but with age interaction
Second round * actual intervention	0.263**	0.275**	0.266**
	(0.102)	(0.107)	(0.123)
Second round	0.059	0.060	0.058
	(0.053)	(0.060)	(0.063)
Full exposure * actual intervention	-	-	0.023
			(0.130)
Full exposure	-	-	1.334***
			(0.063)
Male Child	-0.020	-0.021	-0.027
	(0.033)	(0.035)	(0.036)
Twin	-0.672***	-0.657***	-0.651***
	(0.132)	(0.141)	(0.127)
Primary education female	0.157***	0.151**	0.163**
	(0.058)	(0.063)	(0.064)
Secondary education female	-0.004	0.003	0.066
	(0.128)	(0.141)	(0.144)
Primary education male	0.011	0.003	0.002
	(0.044)	(0.047)	(0.049)
Secondary education male	0.110**	0.086	0.075
	(0.052)	(0.055)	(0.055)
Husband education missing	0.048	0.050	0.052
	(0.053)	(0.057)	(0.057)
Tapwater	0.012	0.011	0.024
	(0.055)	(0.059)	(0.059)
Watercloset	0.050	0.060	0.054
	(0.090)	(0.099)	(0.099)
Wealth index	-0.002	-0.001	-0.004
	(0.014)	(0.016)	(0.016)
Constant	-1.487***	-1.519***	-1.621***
	(0.085)	(0.093)	(0.056)
Observations	10127	8481	8481
Number of villages	211	175	175
R^2	0.19	0.19	0.16

Notes: Absolute value of standard errors below the coefficient estimates. * indicates significance at 10% level; ** at 5% level and *** significant at 1% level of confidence. Standard errors corrected for clustering at the village level.

Appendix

Table A.1: Variable definitions for regressions in Table 7

Variable name in regression	Question from survey instrument
Early liquid introduction	In the first three days after the birth of your child, did (s)he receive any other liquids than your breastmilk?
Colostrum	Do you think that one should give the baby the yellow liquid coming out of the breast before the normal milk arrives?
Worm drugs	Has your child (*name*) received drugs against worms in the last six months?
Bednets	Do you have malaria bednets in your household?
Malaria pills	During your pregnancy, have you taken any medication against malaria?
Vitamin A	Has your child in the last six months received a dose of vitamin A such as this one (*show the container*)?
Took iron during pregnancy	During your pregnancy, have you been given iron capsules or syrup containing iron?
Diarrhea	Has your child had diarrhea in the last two weeks?
Cough	Has your child suffered from a cough, at any moment, over the last two weeks?

Source: Translation of the survey instruments by the author.