Social Desirability Bias in Program Evaluation: The Case of a Childhood Nutrition Program in Sri Lanka

Bailey Peterson-Wilhelm, University of Saskatchewan, bailey27@ksu.edu
Benjamin Schwab, Kansas State University, benschwab@ksu.edu
Dilini Hemachandra, University of Peradeniya, dilinisp@gmail.com
Derek Headey, IFPRI, D.Headey@cgiar.org

Selected Lightning Paper Presentation prepared for presentation at the 2025 AAEA & WAEA Joint Annual Meeting
in Denver, CO; July 27-29, 2025

Social Desirability Bias in Program Evaluation: The Case of a Childhood Nutrition Program in Sri Lanka

Bailey Peterson-Wilhelm

Co-Authors: Benjamin Schwab, Dilini Hemachandra, & Derek Headey

Abstract

We estimate the prevalence of social desirability bias in childhood feeding reports in a UNICEF nutrition cash-plus program in Sri Lanka. Social desirability bias occurs when respondents give the socially "correct" answer, rather than the true answer. While cash benefits were not explicitly conditioned on meeting childhood feeding targets, the training, or "plus" component, made the ideal dietary outcome explicit. We test whether participants misreport the consumption of vitamin A rich foods among young children in this context using list experiments. We find households overstate adherence to program advice by 23 percentage points. The mismeasurement of one feeding component passes through and affects aggregate measures of dietary diversity. The magnitude of the findings suggests that social desirability bias could serve as a potential explanation for the persistent gap between recalled dietary intake and anthropometric outcomes in cash-plus program evaluations. The findings of this study bring together the broader measurement error and program evaluation areas of literature.

1 Introduction

Self-reported data are often subject to bias, particularly when the subject matter is sensitive or morally charged. Social pressures influencing self-reported responses are referred to as social desirability bias, which occurs when a respondent gives what they deem to be the socially "correct" answer rather than the honest answer to a sensitive question. Social desirability bias has been widely investigated in the literature. Social pressures have been shown to bias measures of child labor use (Jouvin, 2023), intimate partner violence (Cullen, 2023), sensitive health behaviors (Lépine et al., 2020; Lépine, Treibich, and D'Exelle, 2020), and voting behavior (Rosenfeld, Imai, and Shapiro, 2016). The implications of such social pressure in the context of program evaluation are comparatively understudied.

Cash plus programs, which incorporate a cash transfer and behavior change communication (BCC), have become popular in the last decade. The programs typically include a cash transfer that is then paired with training aimed at improving a specific outcome domain, which is childhood nutrition in the context of this study. Evaluations of these programs find generally positive effects on measures based on recalled consumption, but evidence for consistent impacts on anthropometric outcomes is mixed (de Groot et al., 2017; Little et al., 2021; Olney et al., 2022). The inconsistency between outcomes raises concerns about the reliability of self-reported data.

We hypothesize that social desirability bias influences key self-reported metrics in nutrition programs that provide a combination of cash transfers and training. Though such transfers are unconditional, training components create a social expectation that the provided cash should be spent in service of the program's objective, such as improving the healthfulness of children's diets. The implied, or explicitly stated, correct use of the cash transfer creates a soft conditionality.[1] While the training element of cash-plus programs intends to foment behavioral change through imparting knowledge, skills, and an improved social environment, the soft conditions may also incentivize respondents to augment their responses to evaluation surveys to align with program objectives and expectations.

The possibility that social desirability bias may contaminate evaluations of social programs is well known. Prior literature has drawn from psychology methods to identify people inherently prone to social desirability bias based on personality (Rawat et al., 2017; Reynolds, 1982). We argue that typically benign topics in the context of a cash-plus program evaluation can become sensitive and therefore subject to social desirability bias. Our hypothesis does not fit within the psychology method parameters, as the circumstance, not the person, determines sensitivity. Instead, our hypothesis more closely fits with the framework proposed by Blair, Coppock, and Moor (2020), which relies on criteria related to the context.

[1]Conditional cash transfers often rely on administrative data to check adherence, which mitigates potential social desirability bias concerns.

Blair, Coppock, and Moor (2020) suggest that sensitivity bias[2] occurs only when four elements are present. The first element is a social referent, a person that the respondent has in mind when responding to a question. For in-person surveys, this social referent is generally the enumerator but could also be the person or group analyzing the survey responses. The second is a perceived risk that the referent will know the respondent's answer to a sensitive question. Third, the respondent assumes to know the referent's preferential answer to the sensitive question. Fourth, there is a perception that the respondent will face social, monetary, or physical repercussions for failing to give the preferred response.

All the criteria for social desirability bias, outlined by Blair, Coppock, and Moor (2020), seem to be present in the context of a program evaluation. The referent is the enumerator or a representative from the implementing agency that will observe the responses in the future. When asked directly, the respondent must blatantly reveal whether they have followed the program advice or not, meeting the second criteria. Programs with a training or information campaign component make explicit the "correct" answer from the perspective of the implementing agency, satisfying criteria three. Respondents face social repercussions, such as shame or embarrassment, and potentially fear being required to return the money not used for food or reduced future benefits—representing financial risks. These factors fulfill criterion four.

Since the typical psychology methods proposed by Reynolds (1982) are unlikely to accurately detect social desirability bias in our setting, we instead consider which of the criteria Blair, Coppock, and Moor (2020) outline can be disrupted.[3] Given the nature of the evaluation survey and the program implementation characteristics, anonymizing responses is the ideal option. If the question can be answered indirectly, the second criterion (the referent must know the response) is not met. Therefore, social desirability bias should not influence the response when sensitive questions are asked indirectly. We leverage a common indirect questioning method, a list experiment, to estimate a measure of child feeding behavior in a childhood nutrition cash-plus program in Sri Lanka. Comparing the indirect and direct measures provides an estimate of the level of social desirability bias.

We find significant levels of social desirability bias in reports of child feeding in a UNICEF cash-plus program in Sri Lanka. Respondents overstated their adherence to nutrition advice given as a part of the BCC component. Using a list experiment to examine one of the specific messages provided during the training, our analysis finds that traditional dietary recall modules overstate feeding of vitamin A-rich foods to children by 23 percentage points. Our analysis is extended to investigate heterogeneity in levels of social desirability bias. Further, we demonstrate downstream implications of using biased

[2]Sensitivity bias is an alternative definition of social desirability bias used by Blair, Coppock, and Moor (2020) that includes monetary and safety threats. In the broader literature, the two terms are often used interchangeably.

[3]Relying on objective measures (like anthropometric health measures) would also prevent social desirability bias. While collecting such measures was not practical for our study, we discuss this method further in Section 5.

data in the calculation of dietary diversity. We contribute to existing literature by bringing context to existing cash plus program literature; highlighting the potential role that measurement error plays in findings.

1.1 Background

Sri Lanka experienced a major economic crisis in 2022. In the five years prior, economic growth had slowed as a result of political uncertainty and the COVID-19 pandemic. Wage labor and the tourism industry were negatively impacted. Significant public debt accumulated, and as a result, Sri Lanka lost access to international financial markets. Due to foreign exchange limitations, imports slowed to a crawl. The slowing economy, paired with restrictive trade policies and low foreign exchange reserves, brought on shortages of day-to-day essentials, high inflation, and negative GDP growth (FAO, 2023; World Bank, 2024).

The macroeconomic conditions had significant effects on individual households. Rising inflation increased the price of a sufficient food basket by as much as 66.4%, with the largest impacts concentrated among poor households (World Bank, 2022). The poverty rate is estimated to have doubled from 2021 to 2022 (World Bank, 2022), leading to a sharp increase in the use of coping strategies that could damage long-term human capital accumulation (FAO, 2023; World Bank, 2024). Sri Lanka currently lacks data systems necessary for a functional social registry that can be used to target social programs (World Bank, 2022). Therefore, assistance efforts were unable to effectively target the most vulnerable households, especially those who were newly poor as a result of the financial crisis.

Among those most impacted were children. As of October 2022, 43.4% of children under five faced some degree of nutritional challenge, the vast majority of which came from undernutrition. The prevalence of wasting increased from 8.2% to 10.1% nationally from 2021 to 2022 (Ministry of Health, 2022). Undernutrition trends hold true for all age groups and sectors (urban, rural, and estate). The annual report prepared by the Sri Lanka Ministry of Health in 2022 recognized a need for intervention and the importance of a healthful diet, especially among young children (Ministry of Health, 2022). In 2023, UNICEF began a childhood nutrition program that involved training and cash transfers in vulnerable districts.

1.2 UNICEF Program

The UNICEF childhood nutrition program provided five monthly cash transfers to eligible households in districts classified as vulnerable. The selected districts have high rates of severe child wasting. Payments were distributed in two waves (Figure 1). The first wave, which began in March 2023 and ended in July 2023, included Anuradhapura, Kegalle, Kilinochchi, Monaragala, Mullaitivu, and

Puttalam. The second wave began in July 2023 and ended in November 2023 and included the Ratnapura, Nuwara Eliya, and Vavuniya districts.

Households with at least one child born between May 1, 2021 and December 31, 2022, were eligible for the program. All households with a young child in the selected districts received the payments uniformly, with no other needs-based criteria. For each eligible child, the household was given 6,750 rupees every month for five months via bank transfer. For reference, in 2023 the average per capita monthly food expenditure was 10,000 rupees (FAO, 2023). Meaning, the cash transfer was a non-trivial sum, making up over half of the average monthly food expenditure for the beneficiary child.

Dietary knowledge, or lack thereof, is another potential determinant of a child's nutritional status. In addition to the cash transfers, a BCC training component was incorporated into the program. The training implementing partner was Sarvodaya, a large community development non-governmental agency (NGO) working in Sri Lanka. Each month households were visited by local Sarvodaya animators. During the visit, caretakers were given a calendar with a monthly nutrition message (Figure 2) along with other relevant Ministry of Health information. The objective of these visits was to improve dietary diversity, increase awareness of available health and nutrition services, and improve the quality of diets for both the child and the mother.

The cash transfer was not conditional on participation in the training or implementation of advice. All eligible households received the cash transfer, regardless of involvement in the trainings. Due to implementation challenges, the transfer of cash and the Sarvodaya training elements were not uniformly synchronized. The delays further reduce the link between the cash transfer and the programmatic advice, potentially weakening any implied soft conditionality. Though the second wave more closely linked the cash transfers with the Sarvodaya training components.

In addition to the Sarvodaya training, Sri Lanka has a robust public health system, compared to other low- and middle-income countries, through which childhood nutrition and health information is disseminated. Therefore, the priming to report specific childhood nutrition outcomes would come not only from the program trainings but also from generally available resources. We discuss the distinction of social desirability bias specific to program interventions and general social desirability bias further in the appendix. Given the universal nature of the program, we refrain from drawing conclusions about the difference between program specific and general social desirability bias. Instead, the appendix provides a discussion useful for future work aimed at disentangling the two.

The criteria for social desirability bias are likely to be met in the context of the UNICEF cash plus program. The social referent is the enumerator and the UNICEF and Sarvodaya representatives that may see the data. When answering questions directly, as is common in surveys, the response is revealed, meeting criteria two. The trainings made the ideal behavior from the perspective of the implementing agencies explicit, so that criteria three holds. Finally, the repercussions for not following program advice are shame and potentially the fear of diminished future benefits.

There are two important ways the repercussions from criterion four manifest. Respondents are likely to want to appear as willing and engaged participants in the event the program had (1) unofficial requirements or (2) potential extensions. Registration for eligible households was publicized with a pamphlet detailing that the cash transfers would be given unconditionally for four months. Data was not collected based on respondents understanding of the unconditional nature of the program. So it is not clear if beneficiaries were aware that the payments were **not** conditional on behavior changes. Another potential motivation for overstating behavior change is that payments were extended once. Initially, the program advertised four monthly payments, which were then extended to five. The final evaluation survey reflects that 44% of households reported expecting four payments while only 16% expected five payments. The program extension may prompt households to alter responses with the belief that program success may lead to another extension. In total, this indicates that the criteria for social desirability bias to influence responses are met.

2 Research Design

2.1 Data

A survey was conducted to evaluate the UNICEF program (Headey, Hemachandra, and Ranucci, 2024). Measurement experiments were embedded within the final survey. The follow-up survey was given to 498 randomly selected households in the second wave of the UNICEF transfers, of which 490 were complete and usable responses. Characteristics of the surveyed households are summarized in Table 1.

The sample was drawn from the three districts in the second wave. Rathnapura has 198 observations, Nuwara Eliya includes 181 observations, and Vavuniya has 111 observations. Sampling was limited to the second wave of the program implementation to reduce the required recall length. The broader survey focused on program involvement and use of the funds, and not necessarily metrics with a long time horizon. Therefore, it was natural to focus on households with more recent program interaction. The shorter recall window may also make social expectations more salient than would likely be present for a survey with a longer time horizon. However, all beneficiaries were a part of the second wave, so the recall window should be equivalent for the sample, eliminating concerns about variation in the recency of program interaction.

It is important to note this study is not a randomized control trial. All eligible households received the cash and had the option to have some training. Any variability in program access did not occur randomly. The universal nature of the cash and training means that we cannot disentangle the impact of program intensity on social desirability bias. We leave this to future work, though a discussion of identifying the impact of program intensity is available in the appendix.

The survey was conducted by district from February 2024 to April 2024 by a team of 18 enumerators. Caregivers, most commonly the mother, were surveyed as they were the most knowledgeable party. Households were selected based on two levels of stratification. First, the sample was stratified based on rural, urban, and estate areas. Then, four Divisional Secretariat (DS) divisions were selected from each district. The DS division selection was random for Rathnapura and Nuwara Eliya; however, Vavuniya only has four DS divisions, so all were selected. Finally, within each DS division, four Grama Niladhari (GN) divisions were selected, representing the second level of stratification. Vavuniya and the urban populations were slightly oversampled relative to the population to have an adequate sample size.

2.2 Direct Questions

Direct, self-reported behavioral questions are the standard approach to measuring unobservable characteristics. Typical food frequency modules involve recall over a set period of time. Respondents were asked to report if their child had consumed a list of foods in the past 24 hours in snacks, meals, or mixed with other foods. The food module included 34 food and drink options, all asked in reference to the same 24-hour time frame. The child food frequency module 24-hour recall window matches the 24-hour recall window used for all items, sensitive and non-sensitive, in the list experiment question, discussed in the next section. Matching the recall window length between the direct and indirect questions ensures comparability of questions across question formats.

Within the food frequency module, the specific food items we are interested in are dark green and leafy vegetables or orange fruits and vegetables, which are typically grouped as vitamin A-rich fruits and vegetables. The dark green leafy vegetables and orange vegetables and fruits were asked in three separate, non-consecutive questions embedded in the food frequency module. First, respondents are asked if their child had consumed *pumpkin, carrots, squash, or sweet potatoes that are yellow or orange inside.* Then, two questions later, they were asked if their child had consumed *any dark green, leafy vegetables, such as spinach, okra, lettuce, gotukola, mukunuwanna, Kathurumurunga or other dark green, leafy vegetables.* Finally, two questions later they were asked if their child had consumed *ripe mangoes, ripe papayas, guava, or passion fruit.* The potential responses for all questions were yes, no, or do not know.

The nutritional advice, in the form of a calendar (Figure 2), recommended parents feed their children leafy green vegetables or orange vegetables or fruits every day, which again matches the 24-hour recall window. Respondents were considered to have followed the programmatic advice by the direct measure, if they answered yes to any of the three questions.

2.3 List Experiments

List experiments, also referred to the item count technique, are commonly used to measure sensitive topics that respondents may be averse to answering honestly. By design, the response to the sensitive item in the list experiment is anonymous even to the enumerator giving the survey. The respondent is alleviated of external pressure to provide what is socially the "correct" response, as they are able to answer indirectly. Meaning the second criteria is unlikely to hold, so that social desirability bias is not present.

In a standard single list experiment, individuals are randomly assigned to two groups, which we refer to as Group A and Group B.[4] Respondents in Group B are read a list of items and asked to report how many are true for them or their household. Individuals in Group A are read the same list of items, with the addition of the sensitive item of interest, and are similarly asked to report how many are true. Importantly, respondents never reveal which statements are true, instead only reporting in aggregate how many are true. The difference in mean affirmative responses between Group A and Group B is the estimated aggregate prevalence of the sensitive item. Therefore, we can uncover the prevalence of a sensitive item indirectly, which ensures the privacy of each individual respondent.

While list experiments mitigate the impact of social desirability bias, it comes with a loss in efficiency. List experiments add random noise to the estimation of the prevalence of an item, especially when compared to the low-variance direct measure. The efficiency of a list experiment can be improved by using a dual list experiment, rather than a single list experiment (Droitcour et al., 1991). In the dual list experiment, respondents are asked two list questions, one with the sensitive item and one without. The order in which the respondents saw the list with the sensitive item is different for Group A and Group B. The non-sensitive items differ between list questions 1 and 2; however, the sensitive item remains the same. The dual list uncovers the indirect response to the sensitive item for all respondents, rather than half, greatly improving efficiency. The double list requires two questions to estimate the prevalence of one metric, which has the potential to increase fatigue and cognitive burden to respondents (Glynn, 2013). Given our relatively limited sample size, the improved efficiency from a dual list is necessary.

Our primary objective is to detect social desirability bias in key programmatic evaluation measures, and so the selection of the sensitive item in the list experiment is key. As noted in Section 1.2, in addition to the unconditional cash transfer, households received nutritional training with a monthly theme. While all of the themes were informational, only one monthly theme had clear, measurable instructions. Parents were given advice to feed their child a dark green leafy or orange vegetable or

[4] Other studies typically refer to these groups as the treatment group and control group. We have elected to use more general language to not confuse the measurement group assignment with program assignment, since this study is not a randomized control trial.

fruit every day, with a corresponding list of health benefits of doing so. The reference to "daily" is a clear boundary that matched the recall window for the direct question. "The beneficiary child in my household ate dark green leafy vegetables or orange colored vegetables or fruit in the past 24 hours" was therefore selected as the key item.

Clear instructions are an important tool to reduce the cognitive burden of answering the complex list experiment questions. To account for differing levels of literacy and numeracy, techniques like passing beans between hands (Tadesse, Abate, and Zewdie, 2020) or counting fingers (Jouvin, 2023) have been used. We opted to use the counting fingers strategy as it did not require additional materials. Respondents were instructed to keep track of affirmative statements on their fingers behind their back, out of sight of the enumerator. Doing so lowers the cognitive effort required for the question while still keeping the response to each item anonymous. The full list experiment instructions and questions is detailed in Table 2.

3 Estimation Strategy

Following notation from Tsai (2019), let T_i be the group indicator. If respondent i is in Group A then $T_i = 1$ and for a respondent in Group B, $T_i = 0$. Respondent i has potential answer S_i to the sensitive item and $R^1_{i,j}$ ($R^2_{i,k}$) to non-sensitive item j (k) with a total of J (K) non-sensitive items in List 1 (List 2). In the context of our study, $S_i = 1$ indicates the respondent i's young beneficiary child was fed a green leafy or orange vegetable or orange fruit in the past 24 hours. For List 1, a response of $R^1_{i,1} = 1$ would indicate a person in respondent i's household had consumed rice in the past 24 hours. By design, the responses to S_i, $R_{i,j}\ \forall\ j$, and $R_{i,k}\ \forall\ k$ are not observable. Instead, we observe $Y^1_i = T_iS_i + \sum_{j=1}^{J} R_{i,j}$, where Y^1_i is the number of affirmative statements in List 1 for respondent i. Similarly, we observe $Y^2_i = (1 - T_i)S_i + \sum_{k=1}^{K} R_{i,k}$, where Y^2_i is the number of affirmative statements in List 2 for respondent i.

To estimate the prevalence of the sensitive item ($P(S_i = 1)$), we use a difference-in-means estimator. Using a double list experiment, we estimate the prevalence of the sensitive item separately for List 1 and List 2 then take the arithmetic mean.

$$P(S_i = 1) = \left[\left(\frac{\sum_{i=1}^{n} Y^1_i T_i}{\sum_{i=1}^{n} T_i} - \frac{\sum_{i=1}^{n} Y^1_i (1 - T_i)}{\sum_{i=1}^{n}(1 - T_i)}\right) + \left(\frac{\sum_{i=1}^{n} Y^2_i (1 - T_i)}{\sum_{i=1}^{n}(1 - T_i)} - \frac{\sum_{i=1}^{n} Y^2_i T_i}{\sum_{i=1}^{n} T_i}\right)\right] /2 \qquad (1)$$

We extend our analysis to include a multivariate analysis to explore heterogeneous levels of social desirability bias. The difference-in-means approach cannot accommodate multivariate analysis without further splitting the sample, so an alternative approach must be used. A linear probability model would allow for a multivariate analysis; however, estimated probabilities may be outside the allowable zero to one interval. Therefore, a nonlinear least-squares regression is most appropriate.[5] Imai (2011) proposes the following:

$$Y_i = J * logit^{-1}(1 + e^{-\gamma_\alpha - \gamma_\beta X_i}) + T_i * logit^{-1}(1 + e^{-\delta_\alpha - \delta_\beta X_i}) + \epsilon_i \tag{2}$$

Where X_i is a vector of covariates. In our specification, we estimate with a series of binary variables separately, which are discussed in Section 4.1. In addition, we include an asset index with a mean of zero to account for household wealth, as the program was not targeted. The asset index was created using responses from the asset module and the *swindex* package in Stata (Schwab et al., 2021). A two-step, nonlinear squares estimation is used to estimate $\hat{\gamma}$ and $\hat{\delta}$. First, $\hat{\gamma}$ is estimated using the non-sensitive list group, where $T_i = 0$. Second, $\hat{\delta}$ is estimated with the $T_i = 1$ group, using $\hat{\gamma}$ from step one for γ. When x_i is a binary variable[6], the prevalence of the sensitive item can be uncovered as follows:

$$P(S_i = 1|x_i = 0) = \frac{e^{\hat{\delta_\alpha}}}{e^{1+\hat{\delta_\alpha}}} \tag{3}$$

$$P(S_i = 1|x_i = 1) = \frac{e^{\hat{\delta_\alpha}+\hat{\delta_\beta}}}{e^{1+\hat{\delta_\alpha}+\hat{\delta_\beta}}} \tag{4}$$

3.1 List Experiment Assumptions

The validity of list experiments hinges on three key assumptions: randomization, no design effects, and no liars. When these assumptions hold, the difference between Group A and Group B for each of the lists captures only the prevalence of the sensitive item.

The randomization assumption requires that Group A and Group B be similar so that the average response to the non-sensitive items is balanced. In Table 1, we see that for nearly all observable characteristics Group A and Group B are statistically similar. The electricity bill is the only characteristic where the groups are statistically different. The difference occurs at the 10% significance level and other measures of household wealth do not vary by group. Therefore, the randomization assumption appears to hold.

For the no design effects assumption to hold, the presence of the sensitive item cannot influence the response to the non-sensitive items. The likelihood of design effects can be proactively reduced by selecting non-sensitive items that are similar in nature to the sensitive item (Blair and Imai, 2012). When all of the items in the list are similar, the sensitive item is less likely to stand out and is therefore unlikely to influence the response to other items. In our list experiments, all of the items were in some

[5] Recent work by Ahlquist (2018) indicates that item count technique regressions models are more sensitive to measurement error compared to the difference in means approach, which is why we did not apply this methodology to our base results.

[6] Since the asset index has a mean zero, it is not necessary to include it in the mean prevalence estimate.

way related to food, from market to consumption.

Post-survey implementation, we use the Blair and Imai (2012) test for design effects, results in Table 3. The test is only suited for a single list experiment, so we have conducted the tests on list 1 and list 2 separately. Panel A tests the joint probability of each response. Negative coefficients indicate that the design effects assumption is violated. In practice, adding one sensitive item to a set list of non-sensitive items should not decrease the probability of that response occurring. None of the probabilities are negative, so we fail to reject the initial no-design effects assumption test. Panel B includes an additional test to determine if any negative coefficients occurred randomly. There are two tests of stochastic dominance relationships conducted separately with the results combined using the Bonferroni correction. The Bonferroni-adjusted p-values near 1 indicate that we fail to reject the null hypothesis of no design effects. In total, we have enough evidence to suggest that the no design effects assumption holds.

The final assumption is that no one is purposefully giving a false answer to the list question, called the no liars assumption. The primary reason that someone would *strategically* give a non-truthful answer to the list experiment is if, in doing so, they would reveal their answer to the sensitive item. In that case, the list question is functionally similar to the direct question. Respondents would inadvertently reveal the answer to the sensitive item if they gave an answer of zero or four when they had the list that includes the sensitive item. These are commonly referred to as floor and ceiling effects, respectively.

To avoid floor and ceiling effects, it is common to include a non-sensitive item with a low prevalence and an item with a high prevalence (Glynn, 2013). For list 1, we included rice consumption to avoid floor effects, as rice is widely consumed. Similarly, purchasing livestock in the past 24 hours would be rare and should reduce the chance of ceiling effects. In list 2, consuming tea was included as a high-prevalence item and preparing food with kerosene is a low-prevalence item. In Table 4 we see that zero and three responses among the groups with three item lists are rare. Answers outside of the allowable range were also uncommon. The prospectively designed questions effectively reduce floor and ceiling effects, making it likely that the no liars assumption holds.

The necessary assumptions for a valid list experiment appear to hold. Therefore, the dual list difference-in-means estimation strategy in Eq. 1 is valid. Our indirect estimate gives the share of households where the beneficiary child was fed a dark leafy green or orange vegetable or fruit in the past 24 hours.

3.2 Social Desirability Bias

We define social desirability bias as the difference between the direct and indirect measures of the prevalence of vitamin A-rich fruit and vegetable consumption in the past 24 hours. Assuming the anonymous nature of the indirect question relieves social pressures, we expect the direct prevalence $P_{direct} > P_{indirect}$. This gives,

$$\text{Social Desirability Bias} = P_{direct} - P_{indirect} \tag{5}$$

Identifying social desirability bias requires that the only difference between the two questions used to estimate prevalence is only the level of social influence. Put another way, both methods would uncover the same prevalence rate in the absence of social pressure. There are three potential threats to our identification strategy, though we argue they are not significant concerns. They are (1) the difference in the cognitive difficulty of the questions, (2) asking all respondents both questions, and (3) the difference in the indirect and direct questioning.

Addressing the first concern, the two questions certainly require different levels of mental effort. However, inaccurate responses to the list experiment due to the cognitive burden add noise to the estimate but should not systematically bias the results. Group A and Group B alike should require similar levels of effort, so inaccuracies decrease efficiency but do not introduce bias, as long as the inaccurate responses are random. Strategically giving a false response would violate the no liars assumption and would have been identified when testing validity assumptions. Instead, any impact of the difference in question difficulty may decrease efficiency, but should not threaten identification.

The second concern is that the response to the first question (list question) about the sensitive topic will influence the response to the second question (direct question) on the same topic. While possible, the list experiment questions were unlikely to influence the responses to the direct questions for the following reasons. First, the questions were in different modules, with a nutritional knowledge module separating them, and therefore would not have been asked consecutively. Second, the direct questions were embedded within a food frequency module, and so the questions related to the dark green leafy vegetables or orange-colored vegetables or fruits would not have stood out in relation to the other food items. In fact, even the items that make up the direct measure were asked non-consecutively. After balancing our maximum potential sample size against these concerns, we maintained that asking all respondents both questions was the appropriate choice.

Finally, the recall window length and the unit of measurement match for both the indirect and direct questions. The key difference is the listing of potential foods in the direct question, while the indirect question only states the more general dark green leafy vegetables or orange-colored vegetables or fruits food groups. This difference could influence the social desirability bias measure in two different ways. Listing the potential food items, rather than giving generic categories, may artificially reduce the number of affirmative answers to the direct question because a household may not include other items outside of the listed food items. If this is the case, then the direct measurement would be an underestimate, and the level of social desirability bias would be understated. In this case, the results of this paper are a lower bound. On the other hand, the list of items may be more helpful in recalling the foods given to their child. If true, the indirect measure may be an understatement so that the social desirability bias would also be overstated. However, given the recall window of 24 hours, recall should not be a significant concern, making overstatement of social desirability bias less likely.

4 Results

When asked directly, 86% of caretakers report that their youngest child ate at least one vitamin A-rich fruit or vegetable in the past 24 hours (Table 5). When asked indirectly, and therefore anonymously, only an estimated 64% of caretakers said they fed their youngest child a vitamin A-rich fruit or vegetable in the past 24 hours (Table 5). We do not observe the true prevalence of vitamin A consumption in young children, as the monitoring would be invasive and infeasible. However, social pressures are more likely to lead to the overstatement of such behavior rather than understatement. Feeding your child a well-balanced diet, especially vitamin A-rich foods as instructed by the Sarvodaya training, is perceived to be the socially "correct" response. When asked directly, households that did not implement this advice may feel pressure to give the "optimal" response. Therefore, the anonymity provided by the indirect question should uncover an estimated prevalence closer to the unobserved true prevalence of vitamin A-rich fruit and vegetable consumption in the recall window.

Since the estimates are statistically different from one another (at the 1% significance level) we find positive, non-trivial levels of overstatement, assuming the indirect measure is more accurate. The results in Table 5 and Figure 3 suggest that 23 percentage points, or 30.5%, of caretakers are overstating their child's consumption of vitamin A-rich foods in the past 24 hours. Since the primary difference between the direct and indirect measures is the anonymity of the question, we argue the mismeasurement in the direct measure can be attributed in large part to social desirability bias. The overstatement reflects inflation of compliance with program advice, which aligned with our expectations of social desirability bias.

In Section 3.1 we provided evidence that the necessary list experiment assumptions held. When using a dual list experiment, we can conduct an additional check. Chuang et al. (2021) suggest comparing the prevalence of the key item for each list question used in the dual list estimation independently.

If the assumptions hold, we would expect statistically equivalent estimates of prevalence for each of the lists estimated separately. In Figure 4 we can see that the estimations of the prevalence of the key item are statistically equivalent. Using either list 1 or list 2 independently, we would still find positive levels of social desirability bias even with the less precise estimates. Figure 4 also visually highlights the efficiency gains from using a dual list experiment. Together, Figure 4 demonstrates the benefits of using a dual list experiment and also gives further confidence about the validity of the list experiments.

4.1 Heterogeneous Effects

We have demonstrated there is an overstatement of adherence to nutritional advice. We now estimate heterogeneous effects to uncover characteristics associated with the level of social desirability bias. Tables 6 & 7 consider how the prevalence of vitamin A-rich fruit and vegetable consumption and social desirability bias vary by key characteristics. Table 6 includes variables related to program interaction and need characteristics. Table 7 has household characteristics. Our limited sample size means many of our heterogeneity results lack statistical power, so much of our discussion is in terms of sign and relative magnitude. Still, in nearly all cases, there is statistically detectable social desirability bias, indicating our core finding is robust to specification and estimation strategy.

Beginning with the first column of Table 6, we look at the impact of receiving nutrition training of any kind, not necessarily tied to the UNICEF program. For both direct and indirect questions, receiving nutrition training was associated with a higher rate of vitamin A fruit and vegetable consumption. Though the difference is not statistically significant, those with nutrition training have lower rates of social desirability bias when compared to those without training. In part, this could be a function of the high rate of consumption, meaning when more people are consuming, there are simply fewer people to overstate consumption.

Column two estimates the relationship between receiving the nutrition calendar, described in Section 1.2 and Figure 2, and social desirability bias. The binary variable is equal to one if the respondent reported that they have received a calendar, regardless of if they were able to show the enumerator the calendar. The share of households that received the calendar (66.5%) was nearly identical to the share of households that reported having Sarvodaya visits, which is when calendars were given, giving confidence in our definition. Here, both direct and indirect measures indicate that receiving the nutritional calendar from the Sarvodaya visits was associated with lower prevalence of vitamin A food consumption. The Sarvodaya visits were not conducted randomly, and the difference in magnitude is small, so the results are unusual but not unexpected. Both those with and without the calendar indicate similar levels of social desirability bias, providing some evidence that general social desirability bias is a major contributor, though as we note in the appendix, this study is not well suited to disentangle the general social desirability bias from that specific to the program intervention.

The third column captures if the respondent was able to correctly identify vitamin A foods. Respondents were given a list of foods and asked to identify those that are a good source of vitamin A. A correct response required selecting all of the vitamin A foods and no additional foods, so getting the question correct by chance is unlikely. Here, there are two key findings. First, those that are able to correctly identify vitamin A foods feed their child more vitamin A foods than those that did not identify vitamin A foods, by both direct and indirect measures. The relationship is statistically significant at the 5% significance level. The higher levels of vitamin A food consumption among those that can identify vitamin A-rich foods provide some support for the UNICEF program hypothesis that knowledge is a barrier to a diverse diet among children. While there is some marginal social desirability bias among those that can identify vitamin A-rich foods, their levels of social desirability bias are lower than those that did not correctly identify vitamin A-rich foods (at the 10% significance level). Since this question required participants to select vitamin A foods from a list, the question is not subject to social desirability bias. As such, this measure demonstrates the difference between self-reported, unobserved behaviors and knowledge-based reports.

The next three columns include indicators of need for assistance at varying degrees of food insecurity. We hypothesize that for those that struggle to afford an adequate and healthful diet, the risk of not meeting the referent's expectations and facing repercussions (like not receiving further assistance) is more salient. Therefore, the incentive to appear in line with social expectations in stronger. Importantly, the binary indicators are also variables likely subject to social desirability bias themselves. Tadesse, Abate, and Zewdie (2020) find significant levels of overstatement of food security. The direct food security questions do not account for any social desirability bias, while the indirect questions only eliminate social desirability bias in vitamin A food consumption, not measures of food security.

Respondents were asked if they were able to afford food on a five item scale; those that indicated less than "adequate" (the third ordered item) ability to afford food were designated as unable to afford food. The fourth column looks at the food security question that asks if households had skipped a meal because there were not resources to get more food in the past four weeks. The final column includes a more severe measure of food security, in which households are asked, if during the past four weeks, the household had run out of food of any kind in the household.

In all three needs-based measures, the direct measure estimates of prevalence indicate that those who show some signs of food insecurity consume vitamin A-rich foods at a higher rate than those who do not indicate food insecurity. This finding is unusual. While the opposite is true for the indirect measure estimates. By the indirect estimates, food insecure households consume vitamin A-rich foods at a lower rate than those that are food secure, which is more in line with expectation. While we have no way to remove the potential social desirability bias in the food security measures, it seems that when the social desirability bias is reduced in the dependent variable, the results are more sensible. Our findings suggest that correcting for social desirability bias in consumption measures could be an important corrective step.

There are statistically higher levels of social desirability bias among those that indicate food insecurity compared to those that do not (10% significance level). The differing level of social desirability bias is in line with our previously stated expectation. Those with a higher degree of need have a higher risk if they do not report the socially acceptable answer. The difference in social desirability bias between the food secure and food insecure is roughly equivalent across the three different measures of food affordability and security. Indicating that social desirability bias does not change significantly based on the depth of need, simply the presence of any need. In addition, these findings suggest that social desirability bias could be disguising relevant inequalities in program effectiveness. The direct results indicate that the program was most successful for households with higher need, though this finding appears to be a function of measurement error rather than realized outcomes.

We included household characteristics, in Table 7, to see if social desirability bias is related to any characteristics. The first characteristic we consider is if the household has above the median asset index. We find similar prevalence of vitamin A food consumption and rates of social desirability bias among those above and below the median asset index. The same is true of the estate sector compared to non-estate and Sinhala compared to Tamil as the primary language. In all cases, the level of social desirability bias is similar between groups. Together, this indicates that social desirability bias is pervasive and seemingly unrelated to cultural or household characteristics.

There are larger differences in the prevalence and social desirability bias among differing levels of education. We use completion of secondary education as our education cut-off point. Someone with an A level or O level education has had a much different educational experience from someone who only completed secondary education. So despite the uneven sample split when using secondary education as the dividing point, it does represent the largest true difference in education experience. For both direct and indirect measures, those with more education have a higher prevalence of vitamin A consumption. Generally, those with higher education have higher dietary diversity scores (Sirasa, Mitchell, and Harris, 2020), and so this is the expected result. We also find higher rates of social desirability bias among the group with less formal education. Respondents with a secondary education or less have similar prevalence of vitamin A food consumption and rates of social desirability bias as those with some degree of food insecurity. In reality, education, income, and food security are all tightly linked, making disentangling the differential impact of each impossible. Still, we find that those with more education still have non-zero levels of social desirability bias. Meaning more education does not necessarily reduce the driving social pressures to overreport vitamin A fruit and vegetable consumption.

Finally, we consider the food production behavior of the household. Households that produce their own vegetables predictably report consuming more vitamin A-rich fruits and vegetables, by both direct and indirect estimates. Interestingly, those that produce their own vegetables have the lowest level of social desirability bias in terms of magnitude of any group in the heterogeneity analysis. In fact, the social desirability bias is only detectable at the 10% significance level. While households that do not produce their own vegetables have higher levels of social desirability bias. It could be that households that are growing vegetables may feel like the production of vegetables is readily observable and therefore they are both more likely to consume the vegetable and less likely to exaggerate their consumption. Meanwhile, when the definition is expanded to include the production of any food, which would still include vegetables, the story flips. The non-producing households report higher prevalence of vitamin A-rich food consumption and have lower rates of social desirability bias when compared to the producing households. The discrepancy could be due to social desirability bias in the reporting of one's own production. Since home gardening was a part of the Sarvodaya training and recommendation, when asked broadly, households may feel obligated to report gardening. However, when asked about producing a specific, and therefore verifiable, food product, like vegetables, the incentive to overstate is diminished. Another potential explanation is that many of those that are producing food in part for their own consumption are producing paddy, which is a major source of calories and therefore precludes the consumption of vitamin A foods. Our data does not allow for disentangling the two, and so the relationship between production, food consumption, and mismeasurement patterns remains an open question for future research.

4.2 Dietary Diversity

In addition to measuring the prevalence of vitamin A-rich foods, consumption data may also be used to calculate other metrics of interest. Dietary diversity is commonly measured in childhood nutrition programs. Using the direct measures and an eight-food group dietary diversity measure[7], 85% of households have adequate dietary diversity.[8] The average number of food groups consumed is 6.05.

The share of households with adequate dietary diversity is quite high. In part, this is because the program is not targeted to exclusively low-income households. Since the only eligibility requirements were residing in a designated district and having a young child, there were beneficiaries and respondents of all income levels. Households with higher income would naturally have more diverse diets, even

[7]The eight food groups are: (1) breastmilk, (2) grains, root, and tubers, (3) legumes and nuts, (4) dairy, (5) flesh foods, (6) eggs, (7) vitamin A rich fruits and vegetables, and (8) other fruits and vegetables. Seven food groups, which excludes breastmilk, is also commonly used. However, 76% of our sample reported that they were currently breastfeeding, so we selected the eight group definition.

[8]Adequate dietary diversity is defined as consuming greater than four food groups.

before the cash transfer or training. In part, the high dietary diversity could also be a function of social desirability bias in the consumption measures passed through to the dietary diversity metric.

Since list experiments only uncover aggregate prevalence, we have no way to uncover which households were overstating vitamin A-rich fruit and vegetable consumption, only that 23 percentage points were doing so. So, we use simulation to demonstrate the impact of social desirability bias on the share of households with adequate dietary diversity and the average number of food groups consumed by the youngest child in the household. To allow for uncertainty, we randomly selected both a level of social desirability bias and which households were adjusted. First, a random level of social desirability bias (B) was selected from a normal distribution with a mean 0.23, our estimated level of social desirability bias, and a standard deviation, derived from the standard error of 0.037 and sample size of 490. Then, a randomly selected B% of vitamin A food consuming households had their response to the vitamin A food group changed from one to zero. This process was repeated 1000 times with the mean and confidence interval reported in Figure 5. When only the vitamin A food group is adjusted for social desirability bias, the share of households with adequate dietary diversity dips to 74%. The average number of food groups consumed also decreased, reaching 5.39 food groups. These simmulated measures of dietary diversity are statistically different from the direct measures at the 5% significance level. Meaning, the consumption mismeasurement passes through to mismeasurement in dietary diversity, though to a slightly lesser degree (11 percentage points overstatement compared to 23 percentage points in consumption measures).

We selected vitamin A-rich fruits and vegetables as our key item because of the bounded and measurable advice given in the monthly calendar. However, other calendars with different nutrition messages were given, though the advice was more difficult to measure. From an evaluation perspective, that distinction is relevant; however, the bounding is not likely to be salient to a respondent. Further, in column 3 of Table 6, we showed that there was higher social desirability bias among respondents that did not correctly identify vitamin A-rich foods. Without being able to pinpoint vitamin A-rich foods, respondents were likely overstating other food products as well. For these reasons, social desirability bias could influence other foods in the dietary diversity measure as well.

Commonly consumed staples are unlikely to be influenced by social desirability bias, so no adjustments were made to these food groups. To account for other social desirability bias, we do another simulation, this time allowing non-staple foods to have some level of social desirability bias between zero and 23%, drawn from a uniform distribution. The zero to 23 interval was selected based on the assumption that the targeted vitamin A messaging would lead to the maximum level of social desirability bias. The selected range and uniform distribution were used to avoid making any assumptions about the social desirability bias of foods we have no indirect measure for. The non-staple foods we allowed to vary are dairy, flesh foods, eggs, and other fruits and vegetables. Vitamin A-rich fruits and vegetables followed the same distributional assumptions and selection process made in the prior scenario. We assumed that the social desirability bias for one food group was independent of another for the same household. In practice, this is unlikely to hold, and so our simulated results represent an upper bound of dietary diversity and a lower bound on social desirability bias.

When non-staple foods are simulated to reduce hypothetical social desirability bias, 66% of households have adequate dietary diversity. Which is different from the direct measure at the 5% significant level but not different from the vitamin A food only estimate at the 5% significance level. The number of food groups consumed is 5.06, which is statistically lower than the directly measured number at the 5% significance level. The social desirability bias in consumption measurement is not likely to be limited to one food group, so the downstream effects on dietary diversity metrics are potentially meaningful.

5 Discussion

A large number of program evaluations find that cash-plus programs have a positive impact on measures of childhood nutrition, like dietary diversity. However, the evidence on anthropometric measures is mixed. Several review articles find some positive effects of cash transfers (Manley et al., 2020) and BCC (Manley, Alderman, and Gentilini, 2022; Olney et al., 2022). While others find null effects of cash transfers (de Groot et al., 2017) and BCC (Little et al., 2021). Even programs with clear positive outcomes may have muddled interpretations due to potential mismeasurement (Maffioli, Tint Zaw, and Field, 2024).[9] The connection between improved diet quality and health has been well established, holding confounding wealth factors constant (Arimond and Ruel, 2004). So the lack of consistent pass-through to improved physical health outcomes is surprising. de Groot et al. (2017) suggest several potential reasons for the inconsistent results, which include supply-side market failures, poor targeting, length of program, and age of children.

We propose mismeasurement, in the form of social desirability bias, as an additional cause. Consumption measures are self-reported, while anthropometric measures are collected through physical measurement. We hypothesize that the positive results using self-reported measures could, in part, be due to overstatement to appear in line with program objectives. In doing so, respondents are reporting behavior change without making any significant or consistent changes in behavior. Without meaningful changes in behavior, there will be no impact on physical outcomes. The longstanding lack of physical evidence of program success, which is not affected by social desirability bias, provides further

[9]Maffioli, Tint Zaw, and Field (2024) find reduced stunting in the cash+BCC group compared to cash only group, yet the control group indicates limited upward potential for BCC interventions.

evidence of this hypothesis.

For example, Premand and Barry (2022) conduct a similar cash transfer and BCC program in Niger. The key difference is that the Premand and Barry (2022) study is an RCT, where the study design allows them to disentangle the differential impact of the cash transfer and the behavior change intervention. Improving dietary diversity is one of the stated goals of the BCC. They find significant improvements in children's dietary diversity (0.24 standard deviation increase), both from a magnitude and statistical perspective. Though the cash transfers alone do not lead to the positive change in dietary diversity among children. The targeted impact of the BCC is unsurprising, as 92% of households in the BCC group attended the training. Despite promising improvements in dietary diversity, there are no meaningful impacts on anthropometric measures, including on wasting (weight-for-height), which generally changes more quickly than stunting (height-for-age).

Premand and Barry (2022) suggest the reallocation of consumption among the household as a potential explanation for the discrepancy. The cash transfer alone, with no BCC intervention, improves the household dietary diversity but not child-level dietary diversity. While the cash transfer paired with the BCC improves the child's dietary diversity, but it has no statistical impact on the household's dietary diversity (though the coefficient is negative). Together, this could indicate an increase in dietary diversity among the household when budgets increase but more targeted investment in the children when BCC aimed at childhood nutrition is provided. While that certainly may explain some of the unrealized anthropometric gap, social desirability bias may also be an important factor.

We compare the impact of social desirability bias in our study context and the RCT in the Premand and Barry (2022) Niger study on dietary diversity. Since different measures of dietary diversity were used, we compare the effect of the program on changes in terms of standard deviation, the unit of the results in the original study. Premand and Barry (2022) estimate that the cash plus BCC increased dietary diversity by 0.24 standard deviations. In our study, the simulated results discussed in Section 4.2 revealed that the food groups consumed decreased by 0.66 groups if the measurement error in vitamin A foods is removed. The direct measure of dietary diversity has a standard deviation of 1.50. Therefore, reducing social desirability bias reduced dietary diversity by 0.44 standard deviations. Meaning the change in dietary diversity is larger for social desirability bias than the childhood nutrition program.

Notably, the context of the two studies is quite different in terms of cultural, geographical, and socioeconomic characteristics. We make no effort to generalize our results to other contexts. We have no way to extrapolate the magnitude of our measures of social desirability bias to Niger or any other context. Our aim is to demonstrate that the conditions for social desirability bias exist in many RCT self-reported data collection efforts and that the impact of the bias may be meaningful in estimation.

For a study like the one conducted by Premand and Barry (2022), mismeasurement could in part account for the gap between self-reported and anthropometric impacts.

6 Conclusions

In this study we estimate the prevalence of social desirability bias in self-reported measures of vitamin A fruit and vegetable consumption among UNICEF program beneficiaries, who received a cash transfer and childhood nutrition training. The nutrition training specifically advocated for the daily consumption of vitamin A rich fruits and vegetables, meaning the necessary conditions for social desirability bias are likely to exist. We find that 23 percentage points of respondents falsely report the daily consumption of vitamin A fruit and vegetable consumption among their youngest child. The overstatement indicates non-trivial levels of social desirability bias.

The level of social desirability bias is consistent and robust to the inclusion of other covariates. We do not find evidence that social desirability bias varies by most household characteristics. We do, however, find higher rates of social desirability bias among those who demonstrate higher levels of need. Those who report facing food insecurity have higher rates of social desirability bias. The higher social desirability bias associated with a higher degree of need indicates overstatement is likely related to the severity of the repercussions faced in the social desirability bias framework proposed by Blair, Coppock, and Moor (2020).

The social desirability bias in consumption passes through, leading to mismeasurement in dietary diversity. Our simulated results indicate that social desirability bias could lead to overstating dietary diversity. While we did not estimate mismeasurement in other food groups, there were other four other calendars with differing themes. So, social desirability bias might also cause overstatement of other food items with similar training themes. Dietary diversity is a common evaluation metric for nutrition programs, especially those aimed at increasing the nutritional status of children. Mismeasurement has the potential to lead to fundamentally skewed interpretations of program effectiveness.

Our study contributes to existing measurement error literature by suggesting social desirability bias as an additional reason for the regularly observed gap between self-reported and anthropometric measures in nutrition program evaluation. The UNICEF program in Sri Lanka was focused on childhood nutrition, and so much of our discussion has been focused on such interventions. However, the conditions necessary for social desirability bias to influence data quality would be present in a wide variety of programs outside of the realm of nutrition. We leave it to future research to investigate the presence of social desirability bias in other programmatic interventions across topics.

Future research should attempt to differentiate generally present social desirability bias from program specific social desirability bias. In the case where a program may amplify existing normative behavior, disentangling the source of social desirability bias (program or broader society) is useful. The natural next step for future research is to build tests for social desirability bias into randomized control trials that incorporate BCC into treatment arms. A study with multiple treatment arms with differing levels of BCC intervention could begin to disentangle general and specific social desirability bias. In addition, incorporating checks of social desirability bias could uncover mechanisms, test alternative measurement strategies, and boost confidence in findings. Future work should also investigate ways to reduce social desirability bias with a lighter touch so that more efficient estimation strategies can be employed.

Figure 1: UNICEF Program Waves

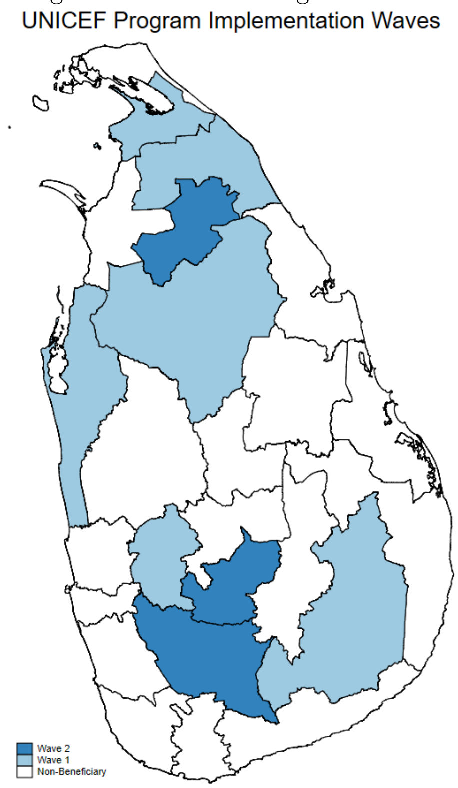

Figure 2: Sarvodaya Vitamin A Foods Themed Calendar

English translation: Dark green leafy vegetables and dark yellow/orange vegetables and fruits help maintain healthy eyesight and protect against disease. So add these to your child's diet daily.

Table 1: Randsomization Assumption Balance Table

	Total	**Group A**	**Group B**	**p-value**
Household Characteristics				
Respondent Age	31.37	31.46	31.28	0.72
Household Size	4.63	4.54	4.73	0.13
Children 4 and under	1.21	1.20	1.23	0.53
Language[a]				
Sinhala	303	152	151	0.34
Tamil	187	102	85	0.34
Education[a]				
Primary or less	7	2	5	0.21
Secondary Grade 6-9	58	28	30	0.56
Secondary O/L level	232	126	106	0.30
Secondary A/L level	136	68	68	0.62
Vocational stream	13	8	5	0.48
Higher education	44	22	22	0.80
Household Wealth				
Has Debt	0.54	0.53	0.56	0.54
Electricity Bill (LRK)	2984.28	3233.25	2719.54	0.08
Owns Refrigerator	0.53	0.51	0.55	0.44
Owns Motorcycle or Car	0.48	0.47	0.49	0.55
Owns Land	0.19	0.18	0.19	0.79
Owns Livestock	0.13	0.15	0.12	0.45
Food Consumption Behavior				
Produces own food	0.76	0.74	0.78	0.42
Child ate orange vegetables in last 24 hours	0.65	0.63	0.68	0.27
Child ate orange fruits in last 24 hours	0.50	0.50	0.49	0.85
Child ate green leafy vegetables in last 24 hours	0.68	0.67	0.68	0.78
Child ate Vit A rich food in last 24 hours	0.86	0.87	0.86	0.85
N	490	254	236	

Note: [a] P-values are for tests of no difference between sample shares in Group A and Group B

Table 2: List Experiment Instructions and Questions

List Instructions

I am going to read a few statements about your household. Do not reply yes or no after I have read each statement. Your answers must remain confidential. Please put one hand behind your back. If the statement is true for your household, lift a finger and keep it raised. If the statement is not true, and does not apply to your household, do not lift a finger. After I have read all the statements, tell me the number of fingers you have raised. Do not tell me which statements are true; only the total number of statements that are true. Let's begin:

List 1

Group A

- Someone in the household has consumed rice in the past 24 hours
- We have purchased livestock in the past 24 hours
- **The beneficiary child in my household ate dark green leafy vegetables or orange colored vegetables or fruit in the past 24 hours**
- Food our household ate in the past 24 hours had been stored in a refrigerator in our house

Group B

- Someone in the household has consumed rice in the past 24 hours
- We have purchased livestock in the past 24 hours
- Food our household ate in the past 24 hours had been stored in a refrigerator in our house

List 2

Group A

- Someone in the household has consumed tea in the past 24 hours
- Someone in our household travelled by car or tuk tuk to purchase food in the past 24 hours
- Food was prepared using Kerosene cooking in the past 24 hours

Group B

- Someone in the household has consumed tea in the past 24 hours
- Someone in our household travelled by car or tuk tuk to purchase food in the past 24 hour
- **The beneficiary child in my household ate dark green leafy vegetables or orange colored vegetables or fruit in the past 24 hours**
- Food was prepared using Kerosene cooking in the past 24 hours

Table 3: No Design Effect Assumption Tests

Panel A : Joint distributions of the key and non-key items

	List 1				List 2			
	Coef	Robust SE	z	P>z	Coef	Robust SE	z	P>z
Pr(R=0,S=1)	0.00	0.01	0.05	0.52	0.00	0.01	-0.10	0.46
Pr(R=0,S=0)	0.00	0.00	1.00	0.84	0.01	0.01	1.74	0.96
Pr(R=1,S=1)	0.51	0.04	13.15	1.00	0.47	0.04	11.99	1.00
Pr(R=1,S=0)	0.20	0.03	7.81	1.00	0.30	0.03	9.85	1.00
Pr(R=2,S=1)	0.17	0.03	6.77	1.00	0.11	0.02	4.97	1.00
Pr(R=2,S=0)	0.10	0.04	2.69	1.00	0.10	0.03	2.94	1.00
Pr(R=3,S=1)	0.00	0.00	1.00	0.84	0.00	0.00	1.00	0.84
Pr(R=3,S=0)	0.01	0.01	1.06	0.86	0.00	0.01	0.52	0.70
Panel B: Test for design effects (with GMS)								
Ha: Pr<0	K^a	Lambda	P>Lambda	#P>Lambda	K^a	Lambda	P>Lambda	#P>Lambda
Pr(R ,S=0)	1.00	0.00	0.50	1.00	0.00	0.00	1.00	1.00
Pr(R ,S=1)	1.00	0.00	0.50	1.00	1.00	0.01	0.46	0.92

Note: Use the Tsai (2019) package to implement Blair and Imai (2012) design effect test

[a] indicates the number of tests to be conducted

Table 4: No Liars Assumption: Frequency of of Floor & Ceiling Responses by Group

	List 1				**List 2**			
	Group A		Group B		Group A		Group B	
Response	Frequency	Percent	Frequency	Percent	Frequency	Percent	Frequency	Percent
0	1	0.39	1	0.42	3	1.15	3	1.26
1	51	20	168	70.29	200	76.92	72	30.13
2	156	61.18	65	27.2	52	20	137	57.32
3	46	18.04	3	1.26	2	0.77	26	10.88
4	1	0.39	-	-	-	-	1	0.42
Out of range	-	-	2	0.84	3	1.15	-	-

Table 5: Vitamin A Foods Consumption by Question Type & Social Desirability Bias

	Direct	**Indirect**	**Difference**
Prevalence	86.38%	63.50%	23.03%
Se	0.015	0.037	0.037

Figure 3: Vitamin A Food Consumption Prevalence & Social Desirability Bias

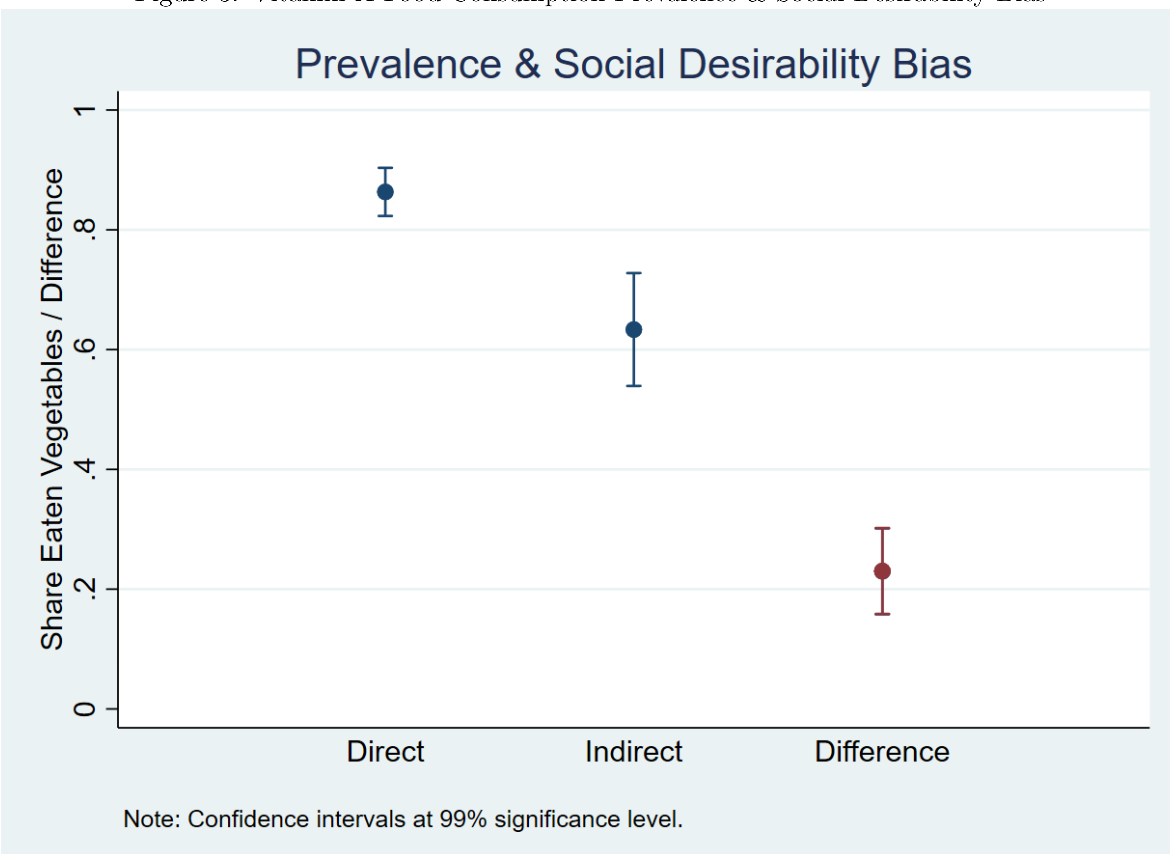

Figure 4: Single List Comparison Validity Check

Figure 5: Simulated Impact of Social Desirability Bias on Childhood Dietary Diversity Adequacy and Food Groups Consumed

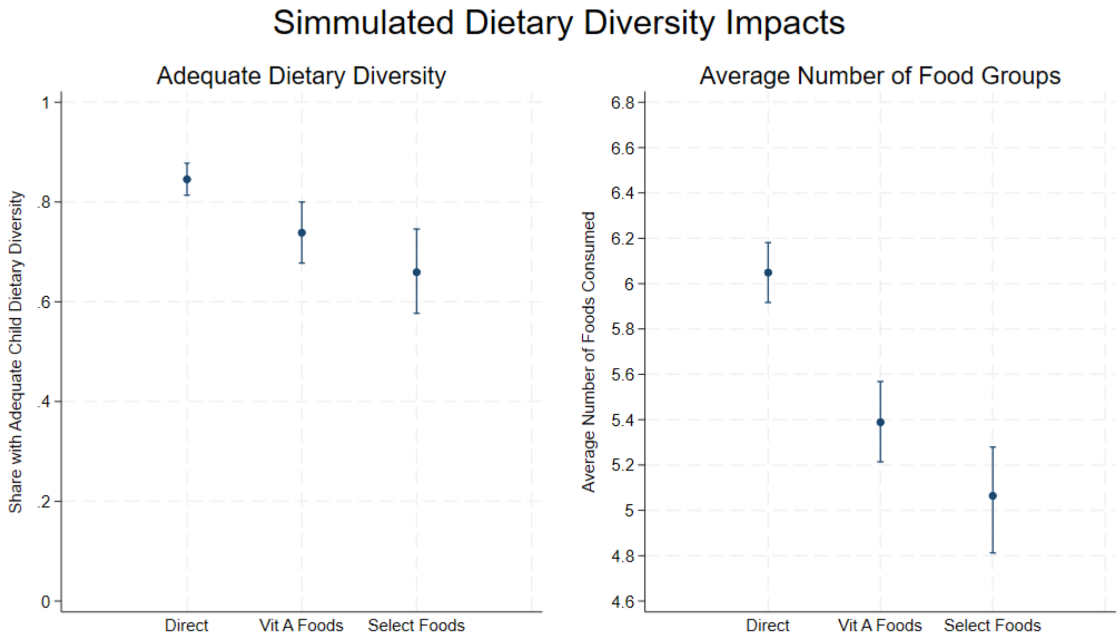

Table 6: Heterogeneity Analysis: Program & Need Characteristics

	Received Any Nutrition Training		Received Calendar		Identified Vitamin A Foods		Cannot Afford Food		Skipped Meal		Ran Out of Food	
	Direct	Indirect	Direct	Indirect	Direct	Indirect	Direct	Indirect	Direct	Indirect	Direct	Indirect
B1=0	86.3%	58.1%	87.3%	64.7%	84.9%	58.4%	86.2%	69.0%	86.1%	67.5%	85.9%	64.7%
SDB		28.2%		22.6%		26.5%		17.2%		18.6%		21.2%
pval		0.00		0.00		0.00		0.00		0.00		0.00
B1=1	87.2%	68.5%	85.9%	63.5%	91.0%*	78.1%**	86.7%	56.28%*	87.2%	53.24%*	91.9%	57.4%
SDB		18.7%		22.4%		12.8%		30.4%		34.0%		34.5%
pval		0.00		0.00		0.06		0.00		0.00		0.00
Diff in SDB		9.5%		0.2%		13.6%		-13.2%		-15.4%		-13.3%
pval		0.21		0.98		0.10		0.07		0.07		0.22
Share=1	50.0%		66.5%		25.0%		39.2%		25.0%		8.1%	

Note: * (**) Indicates B1=1 different from B1=0 at 10% (5%) significant level of the same measure

Table 7: Heterogeneity Analysis: Household Characteristics

	Asset Index above median		Language Sinhala=1, Tamil=0		Estate		Education above Secondary		Produces Own Vegetables		Produces Any Own Food	
	Direct	Indirect	Direct	Indirect	Direct	Indirect	Direct	Indirect	Direct	Indirect	Direct	Indirect
B1=0	87.9%	63.0%	83.6%	61.1%	87.9%	65.7%	82.0%	51.3%	85.3%	60.2%	89.2%	72.6%
SDB		24.9%		22.5%		22.1%		30.7		25.1%		16.6%
pval		0.00		0.00		0.00		0.00		0.00		0.03
B1=1	84.2%	63.3%	88.1%	65.9%	83.7%	60.5%	87.0%	65.9%	90.5%	77.6%	85.5%	61.1%
SDB		21.0%		22.2%		23.2%		21.1%		12.9%		24.4%
pval		0.00		0.00		0.00		0.00		0.07		0.00
Diff in SDB		4.0%		0.3%		-1.1%		9.6%		12.2%		-7.8%
pval		0.60		0.99		0.88		0.38		0.14		0.36
Share =1	50.0%		61.8%		35.4%		86.8%		21.3%		75.8%	

References

Ahlquist, J.S. 2018. “List Experiment Design, Non-Strategic Respondent Error, and Item Count Technique Estimators.” *Political Analysis* 26:34–53.

Arimond, M., and M.T. Ruel. 2004. “Dietary diversity is associated with child nutritional status: evidence from 11 demographic and health surveys.” *The Journal of Nutrition* 134:2579–2585.

Blair, G., A. Coppock, and M. Moor. 2020. “When to Worry about Sensitivity Bias: A Social Reference Theory and Evidence from 30 Years of List Experiments.” *American Political Science Review* 114:1297–1315.

Blair, G., and K. Imai. 2012. “Statistical Analysis of List Experiments.” *Political Analysis* 20:47–77.

Chuang, E., P. Dupas, E. Huillery, and J. Seban. 2021. “Sex, lies, and measurement: Consistency tests for indirect response survey methods.” *Journal of Development Economics* 148:102582.

Cullen, C. 2023. “Method Matters: The Underreporting of Intimate Partner Violence.” *The World Bank Economic Review* 37:49–73.

de Groot, R., T. Palermo, S. Handa, L.P. Ragno, and A. Peterman. 2017. “Cash Transfers and Child Nutrition: Pathways and Impacts.” *Development Policy Review: The Journal of the Overseas Development Institute* 35:621–643.

Droitcour, J., R.A. Caspar, M. Hubbard, T.L. Parsley, W. Visscher, and T.M. Ezzati. 1991. “The item count technique as a method of indirect questioning: A review of its development and a case study application.”, Jan., pp. .

FAO. 2023. *Special Report – FAO/WFP Crop and Food Security Assessment Mission (CFSAM) to the Democratic Socialist Republic of Sri Lanka: 25 May 2023*. No. No. 03/2023 in CFSAMs Special Reports, Rome, Italy: FAO.

Glynn, A.N. 2013. “What Can We Learn with Statistical Truth Serum?: Design and Analysis of the List Experiment.” *Public Opinion Quarterly* 77:159–172.

Headey, D., D. Hemachandra, and I. Ranucci. 2024. “An empirical investigation of UNICEF’s nutrition-sensitive short-term cash transfer programme in Sri Lanka.”, Aug., pp. .

Imai, K. 2011. “Multivariate Regression Analysis for the Item Count Technique.” *Journal of the American Statistical Association* 106:407–416, Publisher: Taylor & Francis _eprint: https://doi.org/10.1198/jasa.2011.ap10415.

Jouvin, M. 2023. "Addressing Social Desirability Bias When Measuring Child Labor Use: An Application to Cocoa Farms in Côte d'Ivoire." *The World Bank Economic Review*, Sep., pp. lhad030.

Little, M.T., K. Roelen, B.C.L. Lange, J.I. Steinert, A.R. Yakubovich, L. Cluver, and D.K. Humphreys. 2021. "Effectiveness of cash-plus programmes on early childhood outcomes compared to cash transfers alone: A systematic review and meta-analysis in low- and middle-income countries." *PLOS Medicine* 18:e1003698, Publisher: Public Library of Science.

Lépine, A., C. Treibich, and B. D'Exelle. 2020. "Nothing but the truth: Consistency and efficiency of the list experiment method for the measurement of sensitive health behaviours." *Social Science & Medicine (1982)* 266:113326.

Lépine, A., C. Treibich, C.T. Ndour, K. Gueye, and P. Vickerman. 2020. "HIV infection risk and condom use among sex workers in Senegal: evidence from the list experiment method." *Health Policy and Planning* 35:408–415.

Maffioli, E.M., N. Tint Zaw, and E. Field. 2024. "A comparison between different models of delivering maternal cash transfers in Myanmar." *Health Policy and Planning* 39:674–682.

Manley, J., H. Alderman, and U. Gentilini. 2022. "More evidence on cash transfers and child nutritional outcomes: a systematic review and meta-analysis." *BMJ Global Health* 7:e008233.

Manley, J., Y. Balarajan, S. Malm, L. Harman, J. Owens, S. Murthy, D. Stewart, N.E. Winder-Rossi, and A. Khurshid. 2020. "Cash transfers and child nutritional outcomes: a systematic review and meta-analysis." *BMJ global health* 5:e003621.

Ministry of Health, . 2022. "Nutrition Month 2022." Working paper, Family Health Bureau, Sri Lanka, Oct.

Olney, D.K., A. Gelli, N. Kumar, H. Alderman, A. Go, and A. Raza. 2022. "Social assistance programme impacts on women's and children's diets and nutritional status." *Maternal & Child Nutrition* 18:e13378, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/mcn.13378.

Premand, P., and O. Barry. 2022. "Behavioral change promotion, cash transfers and early childhood development: Experimental evidence from a government program in a low-income setting." *Journal of Development Economics* 158:102921.

Rawat, R., P.H. Nguyen, L.M. Tran, N. Hajeebhoy, H.V. Nguyen, J. Baker, E.A. Frongillo, M.T. Ruel, and P. Menon. 2017. "Social Franchising and a Nationwide Mass Media Campaign Increased the Prevalence of Adequate Complementary Feeding in Vietnam: A Cluster-Randomized Program Evaluation1." *The Journal of Nutrition* 147:670–679.

Reynolds, W.M. 1982. "Development of reliable and valid short forms of the marlowe-crowne social desirability scale." *Journal of Clinical Psychology* 38:119–125, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/1097-4679%28198201%2938%3A1%3C119%3A%3AAID-JCLP2270380118%3E3.0.CO%3B2-I.

Rosenfeld, B., K. Imai, and J.N. Shapiro. 2016. "An Empirical Validation Study of Popular Survey Methodologies for Sensitive Questions." *American Journal of Political Science* 60:783–802.

Schwab, B., S. Janzen, N.P. Magnan, and W.M. Thompson. 2021. "SWINDEX: Stata module to create a standardized weighted index of multiple indicator variables." *Statistical Software Components*, Feb., pp. , Publisher: Boston College Department of Economics.

Sirasa, F., L. Mitchell, and N. Harris. 2020. "Dietary diversity and food intake of urban preschool children in North-Western Sri Lanka." *Maternal & Child Nutrition* 16:e13006, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/mcn.13006.

Tadesse, G., G.T. Abate, and T. Zewdie. 2020. "Biases in self-reported food insecurity measurement: A list experiment approach." *Food Policy* 92:101862.

Tsai, C.l. 2019. "Statistical analysis of the item-count technique using Stata." *The Stata Journal: Promoting communications on statistics and Stata* 19:390–434.

World Bank. 2022. *Sri Lanka Development Update 2022: Protecting the Poor and Vulnerable in a Time of Crisis*. World Bank.

—. 2024. *Sri Lanka Development Update 2024: Bridge to Recovery*. World Bank.

Appendix

A General & Specific Social Desirability Bias

There are two potential ways social desirability bias can influence the observed response to self-reported questions for BCC programs. The first is what we have termed general social desirability bias, which is present for any measure of a socially charged issue. Regardless of the context of the survey, a well-balanced, healthful diet is a more socially appealing response. Therefore, even in a survey unaffiliated with a program, a survey respondent may overstate their dietary diversity or healthfulness of diet. Secondly, there is likely some degree of what we call specific social desirability bias. In the context of this program, a non-trivial monetary transfer was given along with nutritional training. Beneficiaries understand that the money is intended to be used for enhancing the diets of young children. The training itself may induce social desirability bias that is specific to the program and would not be present otherwise.

Since all respondents received the cash transfers and training "PLUS" components, disaggregating general and specific social desirability bias is not possible in this study. Still, due to varying levels of training availability and perceptions of training, we can shed some light on general versus specific social desirability bias. Since the variation is non-random, we are limited in our ability to disentangle general and specific social desirability bias. Instead, we use the following exercise as a framework for future research to further expand this discussion.

Every respondent was asked if they had received any nutrition training or counselling ever. Another question asked if they had been visited by a Sarvodaya NGO worker in the past three months. With two binary questions, we have four potential groups. The first group received no nutrition training in any form, answering no to both questions. The second group answered yes to the generic nutrition training question but no to the Sarvodaya specific question. The third group said yes to receiving the Sarvodaya visit but no to having a nutrition visit. This response is unusual, considering the Sarvodaya visits included receiving a calendar with nutritional information on it. There are two potential explanations: (1) reporting being visited by the implementing NGO is itself overstated due to social desirability bias or (2) respondents did not perceive the visit to be nutritional counselling. We have no way to disentangle the two retroactively. Finally, the last group said yes to both questions. This indicates they either perceived the Sarvodaya visit as a nutritional visit or another nutritional training happened in addition to the Sarvodaya visit. Again, there is no way to retroactively disentangle the two possibilities.

The difference between no visit and a nutrition only visit would capture general social desirability bias because these groups did not interact with the training component of the program. Still, in our study these individuals still received the cash transfer, and so they are not completely free from program-related social pressures. On the other hand, the difference in social desirability between the nutrition only and both groups would capture the program specific social desirability bias since both groups received nutrition training but only one group received it from the implementing partner. We find some evidence that there is program specific social desirability bias, though we lack the statistical power to estimate precisely (Table 1). Finally, the difference between the no visit and both visits should capture the total level of social desirability bias. As previously noted, we do not have the experimental design or statistical power to fully disentangle specific versus general social desirability bias. We do, however, find some evidence that both forms of social desirability bias exist in our study context, highlighting the distinction between specific and general social desirability bias as an area for future research.

Table 1: Nutrition Training

	Direct	Indirect	Difference	P-val	N
No Nutrition or Sarvodaya Visit	85.6%	52.2%	33.3%	0.00	90
Nutrition Visit Only	89.0%	79.3%*	9.7%	0.36	63
Sarvodaya Visit Only	80.7%	55.8%	24.9%	0.00	160
Both	90.7%	68.2%	22.5%	0.00	173
			Diff-in-Diff		
No Visit- Nutrition Only			23.6%	0.06	
No Visit - Sarvodaya Only			8.5%	0.39	
No Visit- Both			10.8%	0.24	
Nutrition Only - Sarvodaya Only			-15.1%	0.23	
Nutrition Only - Both			-12.8%	0.29	
Sarvodaya Only -Both			2.4%	0.79	

Note: * indicates statistical difference from no nutrition or Sarvodaya visit at the 10% significance level