Essays on behavioral economics: Applications to agricultural surveys and decision making

by

Bailey Peterson-Wilhelm

B.S., Kansas State University, 2019

M.S., University of Arkansas, 2021

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Agricultural Economics
College of Agriculture

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2025

Abstract

Chapter 1

I 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.

Chapter 2

List experiments utilize indirect survey questions to reduce social desirability bias in measures of sensitive behaviors and sentiments. While often used to assess retrospective behavior or opinions of respondents, list experiments have not been widely applied to assessing "deep" parameters of economic models, such as willingness to pay. Common stated preference methods of estimating willingness to pay may be impacted by social desirability bias, particularly when a product has been provided to survey recipients for free. List experiments can uncover the share of respondents willing to pay a given price while reducing social desirability bias. Repeating the method at a variety of prices recovers a partial demand curve. This study discusses the conditions required to satisfy the list experiment validity assumptions and demonstrates the method in an e-extension platform randomized control trial in Sri Lanka. We show that the "no design effect" assumption for list experiments requires that the budget constraint for a household be nonbinding. Under conditions where that assumption is likely to hold, we find direct estimates overstate willingness to pay at low prices. Our findings suggest list experiments may provide a cheap method of more accurately assessing the typically large share of respondents unwilling to pay any non-zero-sum (extensive margin), but are less effective at reducing bias from exaggerated demand (intensive margin).

Chapter 3

Recent events influence decision-making in many domains, including insurance. Pasture, Rangeland, and Forage (PRF) insurance is a novel product that is relatively underused compared to other federal crop insurance programs. As such, we may expect producers to use recent events as a heuristic in decision making. We estimate the impact of recent events on PRF insurance decision making on enrollment decisions and interval selection. We use a county-level two-way fixed effects model to identify the impact of a recent PRF experience, a Livestock Forage Disaster Program (LFP) payment, and weather events on PRF enrollment and the share of acres enrolled in the growing season. A recent positive experience with PRF (defined as an increase in indemnity payments relative to premiums) is associated with increased participation in PRF and but has no relationship with interval selection. LFP payments on the other hand, are consistently positively related to PRF enrollment and selection of growing season intervals. These findings suggest that producers may view PRF and LFP as complements rather than substitutes.

Essays on behavioral economics: Applications to agricultural surveys and decision making

by

Bailey Peterson-Wilhelm

B.S., Kansas State University, 2019

M.S., University of Arkansas, 2021

A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Agricultural Economics
College of Agriculture

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2025

Approved by:

Major Professor
Benjamin Schwab

Abstract

Chapter 1

I 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.

Chapter 2

List experiments utilize indirect survey questions to reduce social desirability bias in measures of sensitive behaviors and sentiments. While often used to assess retrospective behavior or opinions of respondents, list experiments have not been widely applied to assessing "deep" parameters of economic models, such as willingness to pay. Common stated preference methods of estimating willingness to pay may be impacted by social desirability bias, particularly when a product has been provided to survey recipients for free. List experiments can uncover the share of respondents willing to pay a given price while reducing social desirability bias. Repeating the method at a variety of prices recovers a partial demand curve. This study discusses the conditions required to satisfy the list experiment validity assumptions and demonstrates the method in an e-extension platform randomized control trial in Sri Lanka. We show that the "no design effect" assumption for list experiments requires that the budget constraint for a household be nonbinding. Under conditions where that assumption is likely to hold, we find direct estimates overstate willingness to pay at low prices. Our findings suggest list experiments may provide a cheap method of more accurately assessing the typically large share of respondents unwilling to pay any non-zero-sum (extensive margin), but are less effective at reducing bias from exaggerated demand (intensive margin).

Chapter 3

Recent events influence decision-making in many domains, including insurance. Pasture, Rangeland, and Forage (PRF) insurance is a novel product that is relatively underused compared to other federal crop insurance programs. As such, we may expect producers to use recent events as a heuristic in decision making. We estimate the impact of recent events on PRF insurance decision making on enrollment decisions and interval selection. We use a county-level two-way fixed effects model to identify the impact of a recent PRF experience, a Livestock Forage Disaster Program (LFP) payment, and weather events on PRF enrollment and the share of acres enrolled in the growing season. A recent positive experience with PRF (defined as an increase in indemnity payments relative to premiums) is associated with increased participation in PRF and but has no relationship with interval selection. LFP payments on the other hand, are consistently positively related to PRF enrollment and selection of growing season intervals. These findings suggest that producers may view PRF and LFP as complements rather than substitutes.

Table of Contents

List of Figures

List of Tables

Acknowledgments

Thank you to my major advisor Dr. Ben Schwab, and my committee members Drs. Jisang Yu, Tim Dalton, and Mandy Gaulke for your willingness to engage in my development as an economist. Thank you to you and many others along the way that have provided mentorship, input, and encouragement.

I am grateful for the many collaborators who were willing to include my research ideas in ongoing projects. Thank you to Dilini Hemachandra and Derek Headey for your efforts in survey implementation and thoughtful feedback. Thank you to Sara Burrone and Antonio Scognamillo for your willingness to engage in a novel research area and manage data collection efforts. In addition, thank you to many enumerator trainers and enumerators that were vital to the detailed data collection process required for my first and second chapters.

Thank you to everyone in the Department of Agricultural Economics at Kansas State for fostering a collaborative and supportive environment for such a large part of my formal education. I am also grateful for my experience with the National Science Foundation Natural Resource Trainee (NRT) Fellowship, which was key to developing my interdisciplinary research philosophy.

To our graduate cohort, learning from and alongside each of you has been a joy. Our community has been a tremendous source of support and friendship. I look forward to collaborating in the future.

To my friends and family, thank you for your consistent encouragement, compassion, and love. I am so grateful I have so many people worth thanking. Thank you especially to my parents, for being my lifelong supporters. No matter the obscurity of the dream, you have had an unwavering belief in me that is foundational to my belief in myself.

Finally, thank you to Karl. I could not possibly fit an appropriate thank you for a decade of support, positivity, and partnership in this document. So for the sake of brevity, I will leave it at: thank you for being you.

Preface

This dissertation is comprised of three essays linked by behavioral economics themes. The sub-field of behavioral economics allows for the possibility that, when studying human beings, realized behaviors may diverge from the self-interested, rational optimizing agents assumed in broader economics. I apply two behavioral economics concepts, (1) normative expectations and (2) availability heuristic, to agricultural applications in this dissertation.

Essays one and two consider the influence of normative expectations on self-reported behaviors and willingness to pay. These essays argue that in contexts with clear social expectations, agents can gain utility from reporting pro-social behavior, even if untrue. Essay three proposes that the availability heuristic is used in current agricultural insurance decision making as a proxy for a complex optimization problem.

Essay one measures the impact of normative expectations on survey data quality, considering the case of a childhood nutrition program in Sri Lanka. In this essay, I argue that the cash benefits and training associated with a program make explicit the socially ideal response to survey questions posed to respondents. I find that the combination of social expectations and the human desire to reciprocate biases program evaluation metrics collected using standard survey techniques. The findings of this essay are relevant for future program evaluation and contributes to a growing literature on measurement error.

The second essay hypothesizes that normative expectations bias measures of willingness to pay for promoted agricultural technologies. Further, this essay proposes an alternative measurement strategy that mitigates social desirability bias. The proposed method is effective at uncovering unbiased estimates of willingness to pay at low prices, accurately capturing the discontinuous nature of the demand curve at zero. However, the method faces practical implementation challenges at higher prices. The novel strategy developed in this essay makes a methodological contribution to program and policy evaluation literature.

In the third essay, I estimate the relationship between recent weather, insurance, and government program payments, and current Pasture, Rangeland, and Forage (PRF) insurance decision making. The availability heuristic suggests that when decision making is complex, mental shortcuts may be used to ease the cognitive load. In this essay, I find evidence that there is a statistically relevant relationship between recent events and PRF index insurance decisions. The findings of this essay bring behavioral elements to a growing field of PRF literature and are relevant to policy evaluation and design.

This collection of essays brings behavioral concepts to agricultural economics. The unifying theme of this dissertation is accounting for the human element of economics. Each essay covers a different subject matter, yet finds evidence of behavior that would diverge from a standard rational agent assumption. In total, these findings identify mismeasurement in common evaluation methods, propose new methods that account for normative pressures, and discuss the potential for heuristics in decision making.

Chapter 1

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

1.1 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, Treibich, and D'Exelle, 2020; Lépine et al., 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.

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

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

[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.

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

[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 2.6.

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 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.

1.1.1 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.

1.1.2 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 1.1). 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 1.2) 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 <u>four</u> 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.

1.2 Research Design

1.2.1 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 1.1.

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.

1.2.2 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 1.2), 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.

1.2.3 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

[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.

Section 1.1.2, 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 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 2.1.

1.3 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 $Y^1_i = T_i S_i + \sum_{j=1}^{J} R_{i,j}$, where Y_i^1 is the number of affirmative statements in List 1 for respondent i. Similarly, we observe $Y_i^2 = (1 - T_i)S_i + \sum_{k=1}^{K} R_{i,k}$, where Y_i^2 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.

$$P(S_i = 1) = \left[\left(\frac{\sum_{i=1}^{n} Y_i^1 T_i}{\sum_{i=1}^{n} T_i} - \frac{\sum_{i=1}^{n} Y_i^1 (1 - T_i)}{\sum_{i=1}^{n} (1 - T_i)} \right) + \left(\frac{\sum_{i=1}^{n} Y_i^2 (1 - T_i)}{\sum_{i=1}^{n} (1 - T_i)} - \frac{\sum_{i=1}^{n} Y_i^2 T_i}{\sum_{i=1}^{n} T_i} \right) \right] /2 \tag{1.1}$$

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:

$$Y_i = J * logit^{-1}(1 + e^{-\gamma_\alpha - \gamma_\beta X_i}) + T_i * logit^{-1}(1 + e^{-\delta_\alpha - \delta_\beta X_i}) + \epsilon_i \tag{1.2}$$

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 1.4.1. 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

[5] Ahlquist (2018) indicates that item count technique regression 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.

is a binary variable[6], the prevalence of the sensitive item can be uncovered as follows:

$$P(S_i = 1|x_i = 0) = \frac{e^{\hat{\delta_\alpha}}}{e^{1+\hat{\delta_\alpha}}} \tag{1.3}$$

$$P(S_i = 1|x_i = 1) = \frac{e^{\hat{\delta_\alpha}+\hat{\delta_\beta}}}{e^{1+\hat{\delta_\alpha}+\hat{\delta_\beta}}} \tag{1.4}$$

1.3.1 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 1.1, 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 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 1.3. The test is only suited for a single list experiment, so we have con-

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

ducted 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 1.4 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. 2.1 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.

1.3.2 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,

$$\text{Social Desirability Bias} = P_{direct} - P_{indirect} \tag{1.5}$$

Identifying social desirability bias requires that the only difference between the two questions used to estimate prevalence is 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.

1.4 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 1.5). 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 1.5). 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 1.5 and Figure 1.3 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 1.3.1 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 1.4 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 1.4 also visually highlights the efficiency gains from using a dual list experiment. Together, Figure 1.4 demonstrates the benefits of using a dual list experiment and also gives further confidence about the validity of the list experiments.

1.4.1 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 1.6 and 1.7 consider how the prevalence of vitamin A-rich fruit and vegetable consumption and social desirability bias vary by key characteristics. Table 1.6 includes variables related to program interaction and need characteristics. Table 1.7 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 1.6, 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 1.1.2 and Figure 1.2, 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 1.7, 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 O level or A 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.

1.4.2 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 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,

[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.

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 1.5. 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 1.6, 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.

1.5 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) sug-

[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.

gest 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 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. Now, to contextualize the magnitude of of the social desirability effect, we discuss our findings in similar units. In our study, the simulated results discussed in Section 1.4.2 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 magnitude of the change in dietary diversity due to social desirability bias is large, even when compared to changes in dietary diversity due to an intervention.

Notably, the context of the two studies is quite different in terms of cultural, geographical, and socioeconomic characteristics. The social desirability bias in vitamin A foods in Sri Lanka may not generalize to other programs or countries. We have no way to extrapolate the magnitude of our measures of social desirability bias to Niger or any other context. The aim of this discussion 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.

1.6 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. We may expect to find differing levels of social desirability bias between treatment arms. Any non-zero level social desirability bias observed in the control group could be attributed to generally present social desirability bias. Differences in social desirability bias between the control group and treatment groups would identify social desirability bias specific to the program. Further, differences in social desirability bias between treatment arms, for example cash only and cash plus training groups, would uncover the impact of program intensity on social desirability bias. The universal nature of the UNICEF program in Sri Lanka precludes such an analysis, and so the we leave disentangling social desirability bias further for future research.

In addition, incorporating checks of social desirability bias could uncover mechanisms, test alternative measurement strategies, and boost confidence in findings. Unocvering characteristics associated with higher or lower levels of social desirability bias could help identify useful mitigation strategies. Future work should also investigate ways to reduce social desirability bias with a lighter touch so that more efficient estimation strategies can be employed. A possible extension would be to test the impact of increasing the perceived verifiability of questions, so that the incentive to overstate is balanced with the incentive to be honest.

Alternatively, using honesty scripts, following the cheap talk literature, could be tested as a low cost solution to mitigate social desirability bias.

Figure 1.1: *UNICEF Program Waves*

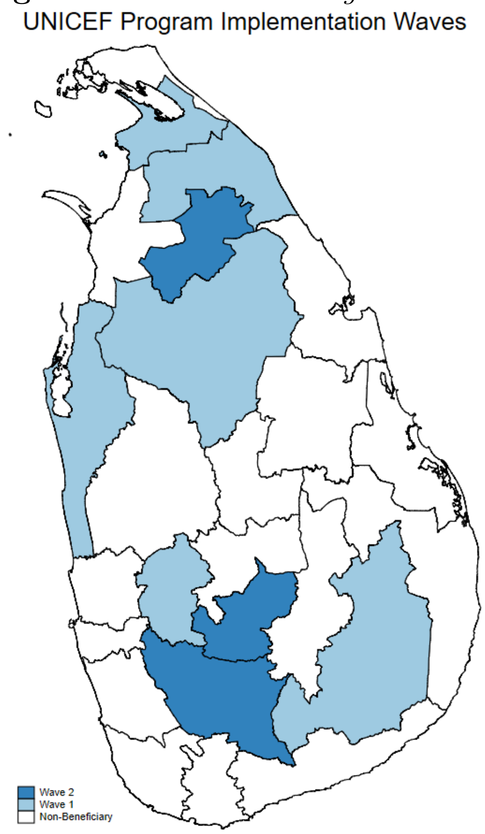


Figure 1.2: *Sarvodaya Vitamin A Foods Themed Calendar*

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 1.1: *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 1.2: *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	**Group B**
• Someone in the household has consumed rice in the past 24 hours	• Someone in the household has consumed rice in the past 24 hours
• We have purchased livestock 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
• Food our household ate in the past 24 hours had been stored in a refrigerator in our house	

List 2

Group A	**Group B**
• Someone in the household has consumed tea in the past 24 hours	• 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	• Someone in our household travelled by car or tuk tuk to purchase food in the past 24 hour
• Food was prepared using Kerosene cooking 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 was prepared using Kerosene cooking in the past 24 hours

Table 1.3: *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 1.4: *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 1.5: *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 1.3: *Vitamin A Food Consumption Prevalence & Social Desirability Bias*

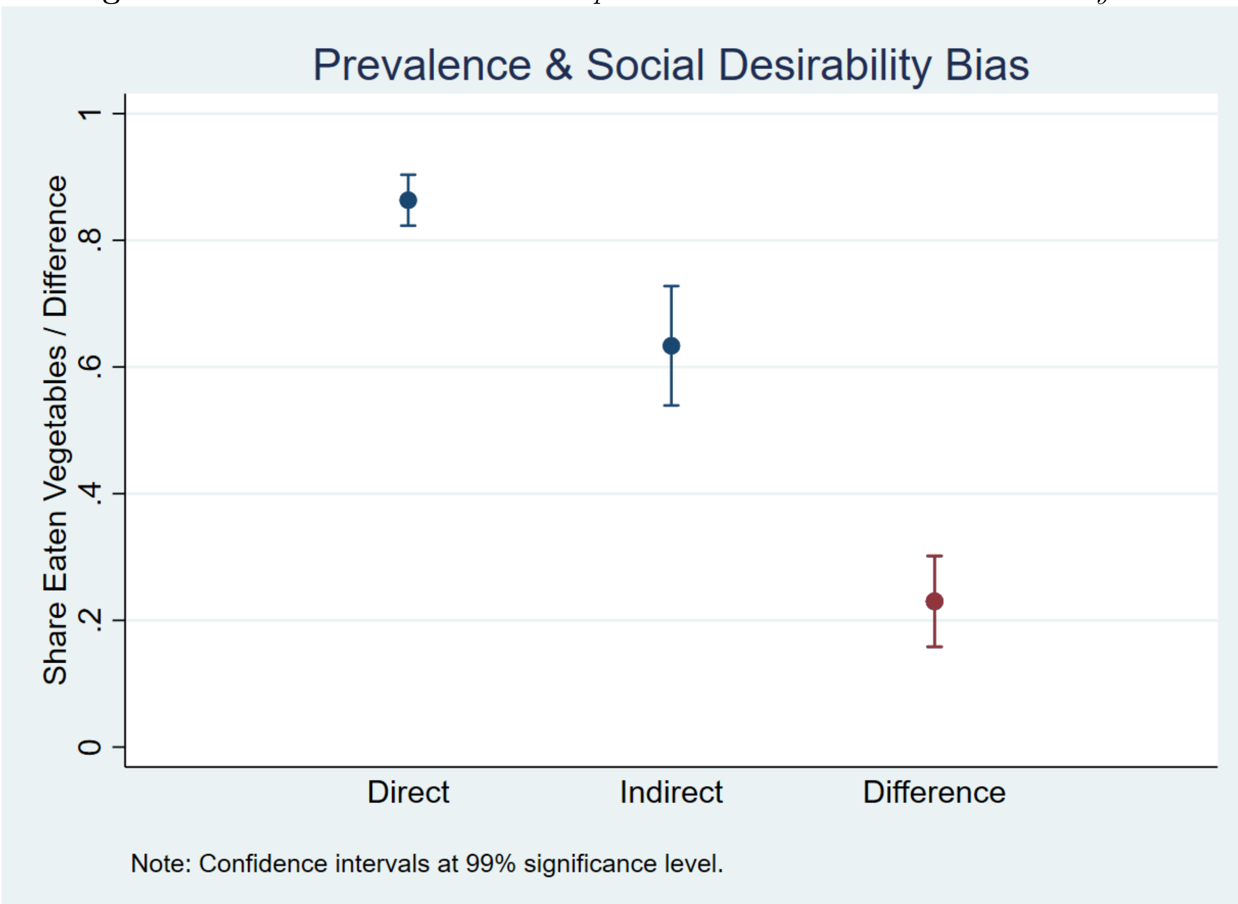


Figure 1.4: *Single List Comparison Validity Check*

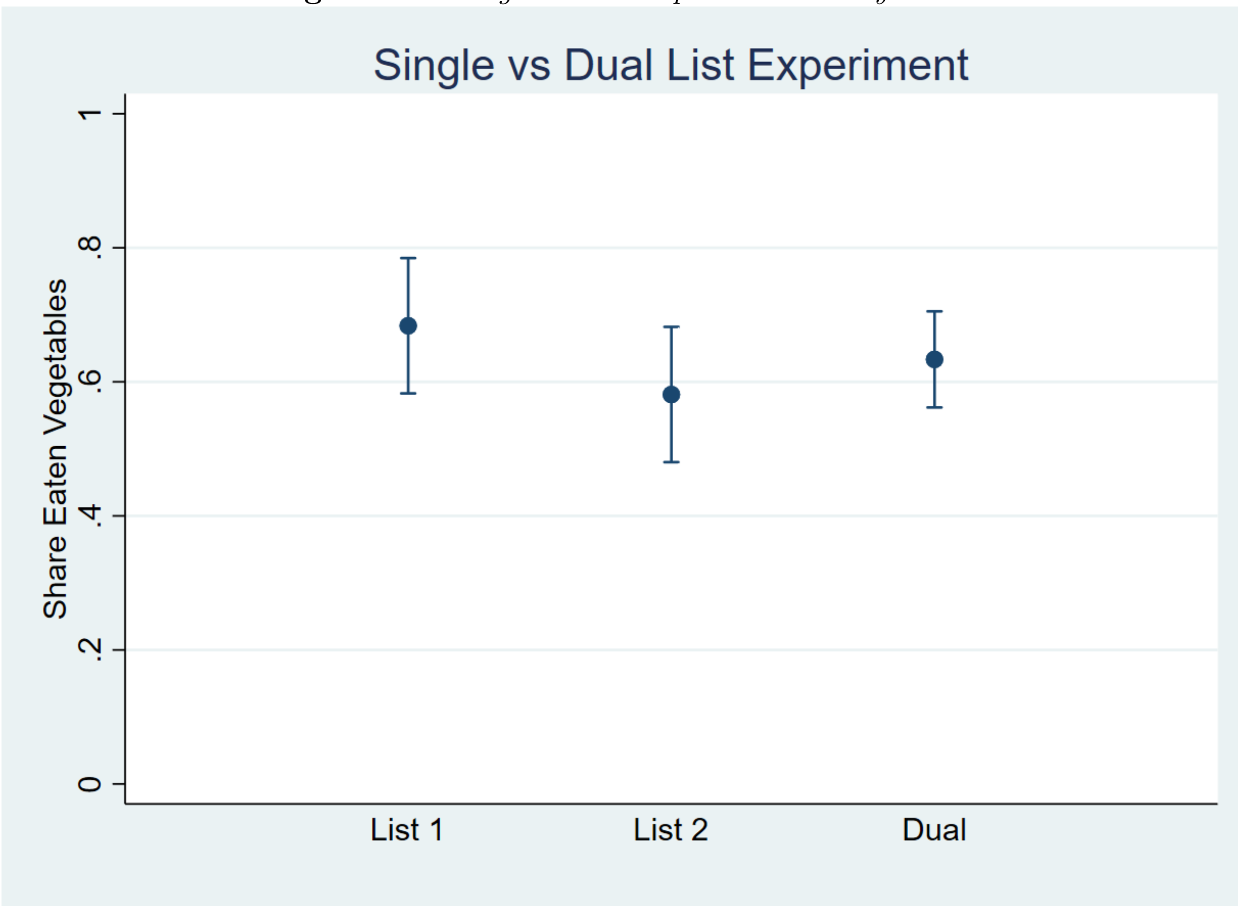


Table 1.6: *Heterogeneity Analysis: Program & Need Characteristics*

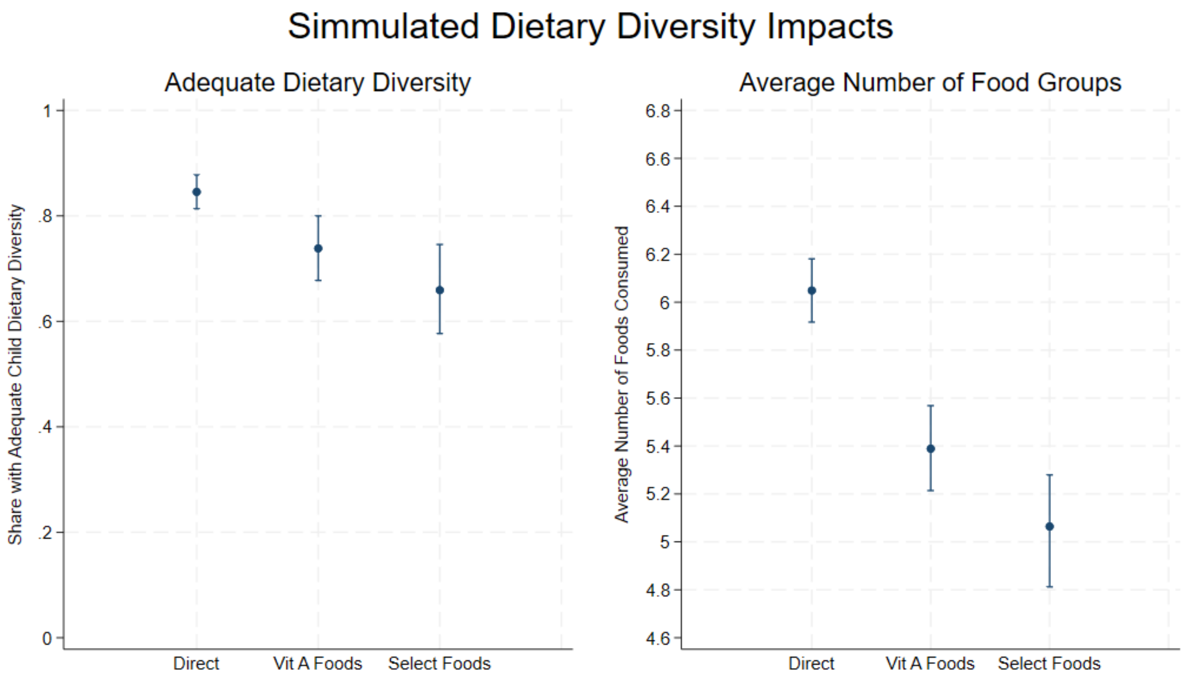
	Received Any Nutrition Training		Received Calendar		Identified Vitamin A Foods		Cannot Afford Food		Skipped Meal		Ran Out of Food	
	Direct	Indirect	Direct	Indirect	Direct	Indirect	Direct	Indirect	Direct	Indirect	Direct	Indirect
B1=0	86.3%	58.1%	87.3%	64.7%	84.9%	58.4%	86.2%	69.0%	86.1%	67.5%	85.9%	64.7%
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 1.7: *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%	

Figure 1.5: *Simulated Impact of Social Desirability Bias on Childhood Dietary Diversity Adequacy and Food Groups Consumed*



Chapter 2

Willingness to Pay for an Agricultural Technology: An Economic Application of List Experiments

2.1 Introduction

Producers' willingness-to-pay (WTP) for an agricultural technology in low- and middle-income countries is often of interest to academics, companies, governments, and nongovernmental organizations (NGOs). When a product has previously been given at a subsidized or zero price as part of a government or NGO intervention, WTP is relevant for decisions about the scale and longevity of the program. However, estimating the WTP may be a secondary or tertiary objective of a survey, with evaluating program objectives and outcomes taking precedence. In these events, as is often the case (Breidert, Hahsler, and Reutterer, 2006), the method to obtain WTP is largely dictated by the remaining time and resources.

Researchers have evaluated many strategies for measuring WTP (Breidert, Hahsler, and Reutterer, 2006; Miller et al., 2011). These methods can be broadly classified into revealed and stated preference methods (Breidert, Hahsler, and Reutterer, 2006). Revealed preference methods require actual purchase decisions in response to prices, and can be applied to market data or experimental designs conducted in the field or a laboratory setting. Stated preference methods, on the other hand, estimate WTP from consumer responses to hypothetical purchase or valuation decisions. These methods span a variety of techniques, including open-ended question formats, inferred valuation, and discrete choice analysis. Both types of WTP elicitation methods carry important trade-offs.

Revealed preferences methods, such as a Becker-DeGroot-Marschak (BDM) auctions, have been used to uncover WTP in the agricultural technology space (Berry, Fischer, and Guiteras, 2020; Channa et al., 2019; Fuller and Ricker-Gilbert, 2021; Schwab and Yu, 2024). Because these auctions require money to be provided in exchange for the good, the participants' bids should reflect their true demand. Thus, estimated WTP in such incentive compatible auctions should closely approximate true WTP. However, auction methods are intensive to deploy in surveys. They require additional time, training, and cognitive effort. If WTP is not the primary objective of data collection efforts, then revealed preference methods may not be feasible.

Stated preference questions are less intensive. They can be asked in a few questions appended to an existing evaluation survey. However, without the transfer of funds, stated WTP tends to yield inflated WTP estimates, which is commonly referred to as hypothetical bias (Champ, Moore, and Bishop, 2009; Murphy et al., 2005; Schmidt and Bijmolt, 2020). There are many causes of hypothetical bias that fall into two categories termed survey bias and social desirability bias by Entem et al. (2019). Survey bias includes elements that are related to the survey implementation itself. It can be circumvented by increasing the consequentiality of the question by, for example, adding a cheap talk script or certainty follow-up questions (Carlsson, Kataria, and Lampi, 2018; Champ, Moore, and Bishop, 2009; Entem et al., 2019).

Social desirability bias is driven by social pressures when the item of interest has normative expectations. Unlike survey bias, social desirability bias can influence responses in both stated and revealed preference contexts (List et al., 2004; Menapace and Raffaelli, 2020). The impact of social desirability bias on revealed preference methods is more severe, due to the increased scrutiny in an experimental setting (Levitt and List, 2007; Lusk and Norwood, 2009a), the costless nature of the signal (Carlsson, Daruvala, and Jaldell, 2010; Menapace and Raffaelli, 2020; Norwood and Lusk, 2011), and in some instances, the presence of the enumerator or experimenter (Leggett et al., 2003; List et al., 2004).

Most of the literature that examines social desirability bias in WTP is estimating the value of a public or environmental good, such as species at risk of extinction (Entem et al., 2019), animal welfare (Lusk and Norwood, 2010), water quality (Carlsson, Kataria, and Lampi, 2018), natural landscapes (Yadav, Van Rensburg, and Kelley, 2013), national parks (Leggett et al., 2003), and other related topics. There has been considerably less focus on social desirability bias influencing estimates of WTP for private goods, with Menapace and Raffaelli (2020) being a notable exception. Still, there are normative social pressures in the context of new, promoted agricultural technologies, not yet available in a market. The existence of the survey itself is a signal the product should be valued (Carlsson, Kataria, and Lampi, 2018; Zizzo, 2010). If a product has been promoted through training or subsidies, the signal is clear - this item should be valued at a positive price.

Outside of WTP, social desirability bias has been examined in social science literature, primarily examining behavior related to sensitive subjects (Cullen, 2023; Jouvin, 2023; Lépine, Treibich, and D'Exelle, 2020; Lépine et al., 2020; Rosenfeld, Imai, and Shapiro, 2016; Tadesse, Abate, and Zewdie, 2020). Respondents may be unwilling to honestly discuss sensitive subjects, and therefore may not honestly respond to direct survey questions on such topics. Their resulting responses may be misrepresented based on their perception of the socially correct answer. Four necessary conditions must be present for social desirability bias to influence responses: (1) a social referent, (2) the respondent assumes to know the referent's preferred response (3) the respondent must respond directly, and (4) some repercussion for not giving the preferred response (Blair, Coppock, and Moor, 2020).

Those four criteria are likely to be met when evaluating WTP for a promoted agricultural technology. An enumerator or the implementing agency would serve as the social referent; the person the respondent has in mind when responding. As noted above, reporting being unwilling to pay their own money for something they have been receiving free of charge may not be perceived as the ideal response on behalf of the referent. Directly responding to questions about WTP means the referent will certainly know the response. Finally, respondents may feel shame for being unwilling to enter the market or fear the service being discontinued if they reveal they do not want to pay for the product themselves.

The same criteria provide insight into mitigation strategies. If one criterion is disrupted, then social desirability bias should be eliminated. In the context of a survey, and especially a survey administered in person by an enumerator, a social referent cannot be removed. Similarly, the social expectation surrounding normative attitudes or actions can not be easily disrupted, as the term "norm" suggests. Removing personal feelings of shame, embarrassment, or fear of future actions is not feasible in a survey setting. In fact, including scripts aimed at decreasing social desirability bias, may inadvertently amplify it by drawing attention to the sensitivity of the item. Therefore, avoiding a direct response is the ideal mode of mitigating social desirability bias. Using indirect questioning should reduce the impact of social desirability bias.

Inferred valuation, a method proposed by Lusk and Norwood (2009a,b), is a commonly used indirect questioning method. People gain satisfaction from simply reporting a positive WTP for a good with normative characteristics, and so have an incentive to overreport WTP. However, a person does not gain satisfaction from overstating another person's WTP. Put another way, another's altruism does not improve my own self-image.[1] Therefore, inferred valuation asks a respondent to report their perception of the WTP of the average person.[2] The inherent assumption is that people use their own, unbiased preferences as a proxy for the preferences of others (Carlsson, Daruvala, and Jaldell, 2010). Inferred valuation has been shown to consistently uncover WTP and voting behavior closer to true outcomes when compared to traditional direct questioning methods (Entem et al., 2019; Lusk and Norwood, 2010, 2009b; Menapace and Raffaelli, 2020; Norwood and Lusk, 2011; Yadav, Van Rensburg,

[1] In fact, there is some evidence that there is satisfaction to be gained by understating another's WTP to elevate my own image in comparison (Menapace and Raffaelli, 2020; Yadav, Van Rensburg, and Kelley, 2013)

[2] Or the share willing to support a ballot measure.

and Kelley, 2013). Even so, there are two instances where inferred valuation may be less effective. First, when the value placed on an item (or service) is highly heterogeneous, identifying the value placed by a random other person is challenging. Such a task would require a respondent to assume another person's values, which can often be biased (Menapace and Raffaelli, 2020). Second, inferred valuation is less effective when there are significant spill-over effects, also referred to as free-riding (Yadav, Van Rensburg, and Kelley, 2013).

These concerns will likely be particularly salient in a promoted agricultural technology program. Agricultural technologies will have variable returns related to farmers' production characteristics, so WTP is likely to be heterogeneous. More pressing is the nature of the spill-over effects. In the context of a program, the availability of a promoted technology is often randomized by geographic area, like farm organization. Since the randomization occurs for the full area, the threat of the technology being diminished is not just to the individual. Instead, the reduction in technology availability would influence the entire community. So, all individuals alike have the incentive to overstate WTP, even when responding from the perception of their neighbors. In the context of a community-based program evaluation, inferred valuation should not eliminate the utility gained from signaling. Therefore, there is value in proposing a new method that does not hinge on the perception of others.

We look to the social science literature for an alternative method. Prior literature has used psychometric scales in surveys to identify *people* inherently prone to social desirability bias (Rawat et al., 2017; Reynolds, 1982). Norwood and Lusk (2011) cast doubt on this method. Alternatively, list experiments are an indirect question method that measures social desirability bias in normative behaviors. Yet, list experiments have not been widely applied to assessing "deep" parameters of economic models, such as WTP. When applied in an economic context, list experiments can uncover the share of respondents willing to pay a given price. Implementing this strategy at a variety of prices traces a partial demand curve, giving insight into consumers' WTP.

We are particularly interested in the ability to measure near-zero prices using stated preference methods. Prior studies that utilize auction methods have shown that demand for promoted health products is strongly discontinuous at zero. In Kenya, Cohen and Dupas (2010) demonstrate a 60% reduction in demand for insecticide-treated bednets when prices increase from zero to \$0.60. An experiment in Zambia testing demand for a new water purification product over a traditional domestic alternative at different subsidy levels found a similar phenomenon. Among those receiving the information treatment, take-up of the new product over the traditional fell from 50% to zero when the new price increased \$0.03 relative to the previous price (Ashraf, Jack, and Kamenica, 2013). In Japan, Iizuka and Shigeoka (2022) show that extremely small co-pays sharply reduce child visits to a physician. These studies have fueled debate on whether cost-sharing methods are appropriate to use when promoting beneficial products in low-income settings.

A recent line of literature also suggests that the initial provision of a good or service for a zero price depresses future demand (Fischer, van den Berg, and Mutengwa, 2015; Shukla, Pullabhotla, and Baylis, 2022). That phenomenon may arise because items or services given at a zero price are viewed fundamentally differently than those at small, but positive prices (Shampanier, Mazar, and Ariely, 2007).

Taken together, the literature suggests that price dynamics across the zero-price threshold comprise a key element for understanding demand for a novel product. However, understanding whether consumers value a product at any price is complicated by measurement concerns. The previous studies rely on incentive-compatible, revealed preference methods. Without a transfer of money, respondents may overstate WTP on the extensive margin to avoid the potential social repercussions of revealing that they place no value on the program's product. As a result, the risk of social desirability bias from stated preference methods may be especially high at trivially low prices.

Our primary objective is to evaluate the effectiveness of list experiments at uncovering WTP at a variety of prices. Using a list experiment as an alternative WTP estimation strategy capitalizes on the experimental benefits of revealed preference methods and the brevity benefits of stated preference methods. In addition, we discuss the considerations and limitations of using list experiments in an economic application. We demonstrate the effectiveness of this theoretical strategy in an empirical application to an e-extension platform in Sri Lanka.

The Food and Agriculture Organization of the United Nations (FAO) has implemented a randomized control trial (RCT) to assess the impact of the Smart Extension and Efficient Decision-making (SEED) Hub, which provides farmers in the treatment groups with free-of-charge, geo-localized, timely, and integrated advisories. This digital tool is expected to enhance farmers' risk management practices and agricultural productivity and strengthen resilience in the face of climate and market risks. Understanding the value participants place on the application is relevant for future program scaling and longevity.

While list experiments effectively reduce social desirability bias, the applicability of the method to demand elicitation is not universal. We show that this method is only suited for products and settings where the respondent's budget constraint is unlikely to bind. The necessary conditions for a valid list experiment are violated when the respondent's ability to purchase at a given price is limited by their income level. Consequently, list experiments generate more reliable estimates of demand at lower price levels.

In the Sri Lankan sample, where the product was provided free prior to the survey, we find stated willingness to pay at low price levels is overstated when elicited by direct questions. Compared to direct elicitation, the share of respondents willing to pay the lowest price level declines by 30% when measured using the list experiment. At higher prices, the list experiment appears to offer either no advantage, and may even perform worse at very high price levels.

Our study offers a novel application of the widely used list experiment method: estimating a partial demand curve. A list experiment can provide a light touch measurement option when the necessary conditions hold and strategic design choices are made. The remainder of this paper will be structured as follows. We begin by describing the list experiment approach.

Then we propose a conceptual framework before introducing the RCT study background and data. Finally, we will discuss our results before closing with our discussion of the potential use cases.

2.2 List Experiments

List experiments allow respondents to answer indirectly, and therefore anonymously. Instead of individually answering yes or no to each item in a list of statements, respondents provide one single aggregated count of their affirmative answers. List experiments can uncover the sample-level prevalence of the behavior or sentiment of interest (or WTP as we will demonstrate) while allowing each respondent to keep individual answers private.

In a standard list experiment, respondents are randomly assigned to either Group A or Group B, so that the groups are similar.[3] Both groups are read a list of items and asked how many are true for their household.[4] Vital to the success of the list experiment, respondents do not reveal which specific items are true, only the aggregate number. For Group B, the list includes items that are not sensitive but are topically similar to the sensitive item. Group A has a list of the same items as Group B, but with the addition of the sensitive item to be measured. The difference between the average number of affirmative responses in Group A and Group B is the sample prevalence of the sensitive behavior.

In our list experiment, each respondent is asked whether they are willing to pay for each of a list of items at price p. Adapting notation used by Tsai (2019), the share of respondents willing to pay the price p, is $P(S_p = 1)$. The observable individual response $Y_{i,p}$ to the list question is a function of the unobservable responses and the individual's group status T_i. For Group A $T_i = 1$ and $T_i = 0$ for Group B. The response for individual i at price p is $Y_{i,p} = S_{i,p}T_i + \sum_{j=1}^{J} R_{i,j,p}$. Here, $S_{i,p}$ is the unobservable response to the sensitive item

[3]Many studies refer to these as treatment and control groups. Since this experiment is nested within a randomized control trial, we use alternative language to avoid confusion. Group A refers to the group that answers a list of question that includes the sensitive item (i.e. the 'treatment group'), and group B refers to the group that does not see a sensitive item (i.e. the 'control group')

[4]Or in sentiment based list experiments, "how many do you agree with?".

at price p and $R_{i,j,p}$ is the unobservable response to the non-sensitive item j in the list of J items at price p. Using the observable responses, we estimate the following difference in means at each p in the set of prices $p \in \{150, 600, 1050, 1500\}$.

$$P(S_p = 1) = \left(\frac{\sum_{i=1}^{n} Y_{i,p} T_i}{\sum_{i=1}^{n} T_i} - \frac{\sum_{i=1}^{n} Y_{i,p}(1 - T_i)}{\sum_{i=1}^{n}(1 - T_i)} \right) \tag{2.1}$$

2.2.1 Assumptions

List experiment validity rests on three key assumptions: randomization, no liars, and no design effects. The literature contains ample discussion of best practices in the design and implementation of list experiments in order to satisfy these assumptions. However, that guidance focuses on identifying the sample prevalence of retroactive behaviors or opinions. As such, the standard recommendations may need to be reexamined for economic applications like assessing demand.

Assumption 1: Randomization

The randomization assumption requires that Group A and Group B are similar. The difference in means framework hinges on Group A and Group B having on average equivalent responses to the non-sensitive items in the list. Standard practice is to check that observable characteristics are balanced across groups. The randomization assumption does not need to be adjusted in an economic context. While no formal test for randomization exists, a standard balance table gives assurances that the two groups are similar.

Assumption 2: No Liars

The second assumption necessitates that respondents do not falsely answer the question. Respondents may strategically give a false answer when reporting the true answer will reveal a socially undesirable response. This can occur when the true answer lies at the extremes of the list length. When that occurs, the list experiment fails to guarantee anonymity and thus becomes functionally the same as a direct question. A response at the high (ceiling effect) or low point (floor effect) reveals the sensitive response, which may prompt a respondent to give a strategically altered response. To avoid floor and ceiling effects, common practice is to include an item that is very likely to have an affirmative response and an item that is very unlikely to have an affirmative response in the list (Glynn, 2013).

The same practice can be applied to the WTP context by including both an item highly likely to be purchased and an item highly unlikely to be purchased at all proposed prices. Prevailing market prices can facilitate the choice of items to satisfy the criteria. To reduce the chance of floor effects, include an item with a prevailing value well above all of the proposed prices. Given the discounted price of the item, even those with small budgets or weak preferences would be likely to select the item. Similarly, including an item that is valued lower than the stated price reduces ceiling effects. Even those with large budgets and high idiosyncratic preferences are unlikely to report a desire to overpay for an item. Importantly, the non-sensitive items in the list should have well-established market values so that respondents have a clear reference for price. The sensitive item, on the other hand, may not have a clear market price so respondents must evaluate what they are willing to pay in the absence of a market reference.

As with the randomization assumption, no formal test for the no liars assumption exists. However, checking for floor and ceiling effects (i.e. responses of 0 and the maximum list length J) in Group B is a good indicator of the potential that the assumption may be violated. Even with the addition of prices, this practice remains a valid and important diagnostic check in an economic context.

Assumption 3: No Design Effects

Assuming no design effects requires that the inclusion of the sensitive item does not influence the responses to the non-sensitive items when the items are privately counted. If adding the sensitive item changes a respondent's answer to a non-sensitive item, then Group A and

Group B are no longer comparable. If the inclusion of the sensitive item increases the non-sensitive response total, then the share willing to pay price p for the sensitive item is overstated. The share willing to pay price p for the sensitive item is understated if the sensitive item decreases the non-sensitive response total.

In a standard list experiment, common practice is to include items of a similar nature in the list so that the sensitive item does not stand out. While this is good practice in an economic context, it is not sufficient to ensure that the no design effects assumption holds. Two factors may also cause the no design effect assumption to be violated: (1) the sensitive item and any non-sensitive item are complements or substitutes and (2) the inclusion of the sensitive item changes the optimal purchase decision for non-sensitive items via a binding budget constraint.

When a good is a substitute or complement for the sensitive item, then the addition of the sensitive item may cause a respondent to change their response to non-sensitive items.[5] The direction of the bias would depend on whether the goods are substitutes or complements. When substitutes, the desire for the sensitive item would be underestimated as the switching between items would not be captured in aggregate. If the goods are complements, the share willing to pay would be overstated since increased demand for the non-sensitive complement item would be falsely attributed to the sensitive item. The impact of complements and substitutes can be mitigated with careful selection of the items in the list.

The binding budget constraint problem occurs when the likelihood that a respondent's hypothetical budget for the category of goods under study is exceeded differs based on group assignment. To motivate, in the scenario considered here, respondents aggregate the number of agricultural inputs they are willing to purchase at a given price. To simplify, assume a household has a predefined separable budget for agricultural input purchase. Respondents in the experiment are asked to consider their current budget constraint for such goods.[6]

[5]Non-sensitive items can be compliments or substitutes with each other, since both items appear in the list for Group A and Group B.

[6]Reminding participants of their budget constraint was originally recommended by Cummings and Taylor (1999). While the method was shown to be less effective in contingent valuation settings, we find evidence

The budget creates a two-step decision process. First, respondents must consider the total number of items at the given price that could be purchased within their budget. If the separable budget exceeds the total to purchase all items in the list, then the budget constraint is non-binding. In this case, in the second step, respondents report the number of items in the list where their own WTP exceeds the listed price. This scenario represents an interior solution. On the other hand, if the separable budget does not exceed the total cost of all items at the given price, then respondents must compare the items in the list relative to one another rather than individually against the given price. Importantly, this comparison of goods relative to one another is partially a function of the list length and, therefore, a function of the group assignment. This point is demonstrated by expanding on the list experiment notation used in Section 2.2.

$$Y_{i,p} = \begin{cases} T_i S_{i,p} + \sum_{j=1}^{J} R_{i,j,p} | B_i \geq (J + T_i)p \\ \lfloor \frac{B_i}{p} \rfloor | B_i < (J + T_i)p \end{cases} \tag{2.2}$$

The observable response $Y_{i,p}$ is now determined in part by the separable budget B_i. If the budget constraint is not binding, then the observable responses will continue to be a function of the unobservable responses to the sensitive and non-sensitive items. However, when the budget constraint is binding, the response is now a function of the budget, rather than the unobservable preferences. The design effects assumption is violated because the switching point between those two unobservable processes is potentially a function of the group assignment.

For example, if an individual could purchase two items at price p with their predetermined budget, then a respondent in Group B could select two of the three items while a respondent in Group A can still only select two but their list has four items. If both answered two, it would look like there was zero demand for the sensitive item. However, it could be that individuals in Group A were selecting the sensitive item instead of one of the non-sensitive

that in a list experiment context, the script is sufficient to remind participants of a binding budget constraint.

items. The comparison of goods, due to a binding budget constraint, would lead to a no design effects assumption failure.

A binding budget constraint is a key limiting factor of using list experiments in an economic context. The next section explores a conceptual framework that demonstrates implications of a binding or non-binding budget constraint in a theoretical application.

2.3 Conceptual Framework

To illustrate the application of a list experiment to demand measurement, we build on the conceptual framework of WTP elicitation from surveys proposed by Lusk and Norwood (2009b) and Menapace and Raffaelli (2020). The key feature of a list experiment is the anonymity provided by the list, which eliminates sources of hypothetical bias related to social signalling that may be present when using a single direct response survey question. However, as noted above, the list of items makes a budget constraint more salient. In the following exposition, we assume the budget for agricultural technologies is separable from other household purchasing decisions. Following Lusk and Norwood (2009a,b), utility is derived from both market and nonmarket sources in our model. Here, we propose that utility is derived from normative behaviors and consumption derived from income maximization. Assuming the components are additive, the resulting framework is such that:

$$U = wM(S, H, E) + (1 - w)V(I, G) \tag{2.3}$$

Where $M(S, H, E)$ is the utility gained from social behavior, which is a function of social signaling (S), honesty (H), and expenditure signaling (E). Utility can be derived by signaling pro-social behaviors ($\frac{\partial M}{\partial S} > 0$), giving honest responses ($\frac{\partial M}{\partial H} > 0$), and expenditure signaling ($\frac{\partial M}{\partial E} \geq 0$).While utility is strictly increasing in signaling and honesty, expenditure signaling has two distinct possibilities: $\frac{\partial M}{\partial E} = 0$ or $\frac{\partial M}{\partial E} > 0$, discussed further with Equation 2.6.[7]

$V(I, G)$ is a standard indirect utility function, where I is income and G is the level of consumption of the good of interest. In addition, w is the weight of importance for the moral and wealth maximizing components. It is worth noting that we assume the hypothetical budget is separable from other income categories, so the income I is distinct from the budget B. Since the stated choice setting is hypothetical, we set $w = 1$, so that $U = M(S, H, E)$ following Menapace and Raffaelli (2020).

We are interested in the unbiased stated WTP for the item of interest j^* (wtp_j^s). Within a list context, there are J other items of interest where $j \neq j^*$. This gives an aggregate stated WTP such that $WTP^S = \sum_{j=1}^{J} wtp_j^s$. Similarly, the true WTP is the sum of the WTP for J items, giving $WTP^T = \sum_{j=1}^{J} wtp_j^t$. We define true WTP to be equivalent to the WTP when a monetary transaction occurs. That is, we are not attempting to remove any social desirability bias from a real market interaction. When a direct question is asked about item j^*, then $WTP^S = \sum_{j=j^*} wtp_j^s = wtp_{j^*}^s$ and $WTP^T = \sum_{j=j^*} wtpts_j = wtp_{j^*}^t$.

Social signaling (S) is a function of the stated WTP ($wtp_{j^*}^s$). Direct and indirect questions have different signaling opportunities and therefore must be defined separately. In a direct question, individuals have the opportunity to boost moral behavior, and therefore utility, by overstating WTP. For simplicity, we assume $\alpha = 1$. The indirect question, in the form of a list experiment, does not allow for signaling.[8] The anonymous response prevents the respondent from revealing socially normative behaviors, so there is no incentive to alter responses. For the indirect question, the utility of signaling is zero since there is no way to link the signal to the individual. This is represented as such:

$$S(WPT^S) = \begin{cases} a + \alpha wtp_{j^*}^s \text{ if direct} \\ 0 \text{ if indirect} \end{cases} \tag{2.4}$$

[7]While not directly relevant here, there may be some circumstances where an individual gains utility from hiding wealth or appearing less well off.

[8]Recall, we argue community-level randomization in the context of program implementation allows for utility to be gained from signaling among inferred valuation indirect questions.

Honesty is a function of stated WTP (WTP^S) and true WTP (WTP^T). Honesty increases as the difference between the stated WTP and the true WTP decreases (Lusk and Norwood, 2009a). Individuals perceive honesty for each item individually. Meaning, aggregate honesty is not the total difference between stated WTP and true WTP, but the sum of the squared difference for each item. For continuity, we keep the functional form used previously by Menapace and Raffaelli (2020).

$$H(WTP^S, WTP^T) = -0.5\sum_{j=1}^{J}(wtp_j^s - wtp_j^t)^2 \tag{2.5}$$

Lastly, the expenditure signal is similarly a function of stated WTP (WTP^S) and true WTP (WTP^T). In a list context, an individual has the opportunity to improve their own self-image by signaling their separable budget for agricultural technologies. Reporting more items signals a larger budget, which increases utility. A notable difference between the budget signal and honesty is that the budget signal occurs in aggregate. Therefore, it is the difference between the total stated WTP (WTP^S) and the total true WTP (WTP^T), not the individual differences as with honesty. We maintain the same functional form for simplicity, giving:

$$E(WTP^S, WTP^T) = 0.5(\sum_{j=1}^{J} wtp_j^s - \sum_{j=1}^{J} wtp_j^t)^2 \tag{2.6}$$

As noted in Section 2.2.1, a response to the list question is determined in two steps. The first step specifies if the budget constraint is binding or not. If the budget constraint does not bind, then the budget does not influence the second step. When the second step (determining WTP for each item) is unencumbered by the budget constraint, there is an interior solution and no expenditure signal such that $\frac{\partial M}{\partial E} = 0$. In this scenario, the assumption of no design effects is satisfied. When the budget constraint binds, the respondent's budget enters the second step, a comparison that violates the design effects assumption and allows for expenditure signaling, so that $\frac{\partial M}{\partial E} > 0$.

We identify the utility maximizing stated WTP using the first order condition of the morality utility function with respect to $wtp^s_{j^*}$.

$$wtp^s_{j^*} = wtp^t_{j^*} + \frac{M_S}{M_H - M_E} + M_E \sum_{j \neq j^*}^{J} (wtp^s_j - wtp^t_j) \tag{2.7}$$

The first term captures the true WTP. The second term represents the utility trade-off between reporting normative behavior, honesty, and expenditure signaling. As expenditure signaling increases, honesty decreases. The final term is the utility gained from expenditure signaling. There are three potential stated WTP scenarios. First, a standard direct question in which case $M_S > 0$. In a hypothetical single-question, direct format, individuals are freely able to signal.[9] Individuals only consider the value of one item. Therefore, no other $j \neq j^*$ items appear and the third expenditure signaling term is zero. In this scenario stated WTP is:

$$wtp^{Direct}_{j^*} = \frac{M_S}{M_H - M_E} + wtp^t_{j^*} \tag{2.8}$$

In the second scenario, we consider an indirect question, so that $M_S = 0$, and a non-binding budget constraint. Since there is an interior solution, $M_E = 0$. For a household where the budget constraint is not binding, a household can report purchasing the items they value. Therefore, there is no utility to be gained from reporting the desire to purchase an item they do not value at the given price. This occurs because the budget is no longer salient when making decisions, so the marginal benefit of budget signaling goes to zero. In this scenario, the stated WTP collapses to the true WTP.

$$wtp^{Indirect,Non-Binding}_{j^*} = wtp^t_{j^*} \tag{2.9}$$

Finally, the third scenario is an indirect question with a binding budget constraint. The indirect question prevents signaling so that $M_S = 0$. However, there is an opportunity to increase utility by expenditure signaling ($M_E > 0$). Importantly, the budget signaling is related to the list length, which varies by group assignment as noted in 2.2.1. This again

[9] This is also likely the stated WTP for an inferred valuation method in a randomized program context.

highlights the relevance of a binding budget constraint to the no design effects assumption. This gives:

$$wtp_{j^*}^{Indirect,Binding} = wtp_{j^*}^t + M_E \sum_{j \neq j^*}^{J} (wtp_j^s - wtp_j^t) \tag{2.10}$$

Comparing the three scenarios, there are important trade-offs to consider. An indirect approach will uncover the true WTP when the budget constraint is not binding, since there is no opportunity to signal pro-social or expenditure behavior. The budget constraint is less likely to bind, revealing an interior solution, when wtp^t is low or budgets are high. However, if the anticipated budget is low or the expected WTP is high, then the trade-off between the social signaling and honesty ratio and expenditure signaling must be evaluated. As the price increases, the probability of having an interior solution decreases, holding all else constant. In this case, a list question may perform worse relative to a direct question because respondents can signal exaggerated expenditure for an item category. The trade-off depends on which signal is stronger, which can vary by context. In situations with strong social expectations, indirect measure may be preferred. However, in scenarios with limited budgets or a highly valued item, then the likelihood the budget will bind and the size of the expenditure signal may caution against using a list approach.

2.4 Background

In July 2023, the Food and Agriculture Organization of the United Nations (FAO) began a randomized control trial (RCT) aimed at improving Sri Lankan paddy farmers' access to information. Treated farmers received access to SEED Hub, an integrated digital extension service. SEED Hub was designed in conjunction with the FAO, the Sri Lankan Ministry of Agriculture, and research institutes to bring together weather, agricultural production, and market information into one comprehensive platform.

The motivation for the program was to reduce barriers to information and therefore optimize agricultural production and risk management strategies. Climate change and natural hazards jeopardize production, especially among Sri Lankan paddy farmers. Having access to more complete and reliable information may improve farmer decision-making in a way that improves production, reduces risk and loss, and ultimately improves farm household livelihoods.

A total of 220 farm organizations were selected, so that the sample was nationally representative of paddy farmers in terms of agro-ecological zones and irrigation schemes. Half of the farm organizations were designated as the control group. The remaining 110 farm organizations were then divided again into two treatment arms. The first treatment arm received the login information for the application, we refer to this group as the "Access Only" group. The second treated group received the login information and a training session on how to use the platform, which is the "Access + Training" group. For both groups, access to the SEED Hub platform was free.

In both treatment groups, farmers received the log-in information at a hub deployment meeting. The meetings took place at the farm organization before the Maha (primary) growing season. During this time, farm organizations in the "Access + Training" group were given the training. The training consisted of a lecture, demonstration, and time for questions. A typical training session took 2.5 hours.

2.4.1 Data

The data was collected from March 2024 to May 2024 in 10 regions in Sri Lanka by a team of enumerators using the digital Survey Solutions platform.[10] The measurement experiments were embedded in the endline survey. For which, respondents were randomly assigned to Group A or Group B, stratified by RCT group assignment and farmer organization. The RCT treatment groups included a total of 1,100 individuals.

We use a single list experiment difference in means strategy to estimate the share of individuals willing to pay a given price for SEED Hub. While the improved efficiency associated

[10]The regions are: Ampara, Anuradhapura, Badulla, Batticaloa, Galle, Kegalle, Kilinochchi, Kurunegala, Matara, Vavuniya

with a double list experiment may be useful, it is not feasible in our setting. A double list experiment requires two list questions to uncover one prevalence. Each group alternatively has the list with the sensitive item and the non-key items in the list change.

We opted to implement single list experiments, despite the potential efficiency gains of a double list experiment for two reasons. First, the number of questions that would be required to uncover one prevalence was not feasible given the length of the existing survey. To maintain the sample size, each individual would need to answer six cognitively taxing questions. Doing so may jeopardize the data quality. More pressingly, in a double list experiment the individual would alternate between having SEED Hub in their list. At times, the non-key items would change but SEED Hub would remain the same. Doing so would draw attention to SEED Hub and increase the likelihood that respondents augment their answers, violating the no design effects validity assumption (discussed in Section 2.2.1). So, we maintain that a single list is the most appropriate estimation strategy.

The list experiment instructions and lists are outlined in Table 2.1. Many elements of the instructions do not change with the inclusion of prices. The instructions for the list experiment emphasize that respondents should not give individual answers for each item but should instead give the total number of responses. Even without the inclusion of prices, list experiments are more cognitively burdensome than comparable direct questions. To increase the accessibility of the questions, we ask respondents to count their affirmative statements on their fingers out of sight of the enumerator. Counting props or fingers is common practice in list experiments, given the varying levels of literacy and numeracy (Jouvin, 2023; Tadesse, Abate, and Zewdie, 2020).

There are two changes to the list experiment instructions for an economic application. First, we ask respondents to consider their current budget for purchasing farm tools, assets, or services. Since there is no transfer of funds for this question, hypothetical bias is a concern. Similar to a cheap talk statement, prompting respondents to think of their budget constraint should reduce the impact of hypothetical bias.

The second change made to the list experiment is the inclusion of the price. Each individual is asked the same question three times, with only the price varying. The price for each question was randomly drawn from a set of four prices without replacement. This design ensures each individual would see three unique prices in a random order.

The prices were selected based on responses in the midline survey for the same RCT. The midline and endline surveys consider four different types of agricultural information: weather, disaster, farming, and marketing. Weather data was the most commonly used SEED Hub function, as reported in the midline survey. In addition, farmers were asked if they had paid for any information services in any of the four categories. While the vast majority did not pay for information, among those that did pay the average monthly payment for weather information was 605 LRK with a standard deviation of 454. The average payments for the other sources of information were similar, ranging from 427 to 647 per month. Using the information from the midline, we selected 600 LRK, one standard deviation below the mean (150 LRK), and one and two standard deviations above the mean (1050 and 1500 LRK). All of these prices fall within the range of reported prices in the midline, with the highest price reporting being 3000 and the lowest 30 for one month. Selecting a near-zero price was of particular interest, given our interest in censoring in response to receiving the product freely in the past. For reference, a typical daily labor wage for men's agricultural work is 2,250 LRK, so 150 LRK would reflect less than one hour of agricultural labor.

The number of question iterations and prices were selected based on a balance of survey fatigue and statistical power. Since the measurement experiments were embedded in the endline survey of an RCT, the sample size was fixed at a maximum of 1,100. The list experiments can be burdensome, causing survey fatigue, so we limited the number of potential questions to three. Therefore, the number of potential prices we could include in our set was determined so that we could maintain statistical power to detect differences between direct and indirect methods. We used simulation to determine the necessary sample size per price level to detect (1) a WTP different from zero and (2) social desirability bias at a variety of potential levels of bias. Further discussion of this process is available in the Appendix B.3. Based on this work, we selected three questions with prices drawn from a set of four, leading to a maximum potential sample size of 885 for each price.

The indirect questions were administered after the Maha season production and crop sales modules. The prior questions would prime farmers to think about their production decisions, but not SEED Hub specifically. Three modules separate the indirect questions from the SEED Hub module, where the direct WTP questions appear. The direct SEED Hub questions appeared at the end of the information services module, which included questions about all sources of information used, including SEED Hub. The following direct question appeared four times for all respondents. *Would you be willing to pay **Price** Rs. for a one-month S.E.E.D. hub subscription, a mobile application that provides information on agronomic practices, market prices and weather forecasts, and related advisories?* Respondents were asked about all four prices in our set in a random order.

All respondents were asked both the list questions and the direct questions. There may be some concern that the list question may influence the direct response among individuals in Group A. Prior literature has suggested asking direct questions only of Group B. However, the distance between questions in the survey reduces the likelihood of the indirect response affecting the direct response. Asking the question of all respondents increases statistical efficiency, and would therefore be preferable. We compare the share willing to pay each price using the direct measure by the list group and find no statistically meaningful differences (Table 2.2). Therefore, we use the full RCT treatment sample to construct both the indirect and direct estimates.

The prices for both the indirect and direct measures appeared in random order to avoid anchoring effects. However, in doing so, respondents may have given inconsistent answers. For example, indicating directly or indirectly, that they would be willing to pay 600 LRK but not 150 LRK. Such responses were removed from the estimation sample. For the indirect question, if a respondent gave a larger number of affirmative statements at a higher price than a lower price the individual was removed from all list questions. Similarly, if a respondent answered the direct question as yes for a higher price and no for a lower price, the respondent was removed from all questions. Given the cognitive burden of the indirect questions, the inconsistent response rate is much higher compared to the direct question (Table 2.3). However, notably, the number of inconsistent responses is balanced between list groups. Meaning, that the inconsistencies were not caused by the group assignment and so the randomization holds, which is relevant for the validity assumptions discussed in Section 2.2.1.

The high number of inconsistent responses raises concerns about internal validity. Households with inconsistent responses tend to have lower education, a larger total agricultural area, use more male hired labor, less household labor, and a larger harvest. This indicates that large-scale farmers may be underrepresented in our sample (Appendix Table B.1). Prior literature in the consumer choice space has similarly found significant levels of inconsistent responses and has shown that including inconsistent responses can bias WTP (Colombo, Glenk, and Rocamora-Montiel, 2016; Sælensminde, 2002). Although we do not find that inconsistent statistically affects our results (Appendix Figure B.1), we exclude them from our analysis nonetheless.

2.5 Results

2.5.1 Assumptions

As discussed in Section 2.2.1, a valid list estimation requires three assumptions to hold: (1) randomization, (2) no liars, and (3) no design effects. In this section, we discuss the validity of these assumptions in our empirical application. First, Table 2.4 checks for balance among observable household and agricultural production characteristics. The characteristics are balanced between the list groups, suggesting that the randomization assumption is likely to hold.

Next, we consider the no liars assumption. The advice outlined in Section 2.2.1 would suggest a high and low value item should be included. Urea is a valuable input, 10 kgs would be worth around 2,150 LRK at the subsidized price (3,400 LRK unsubsidized) (Centre, 2023). At any of the prices in our set, the urea would be a good deal. A half day of labor would be valued at around 1,125 LRK, so this item should have mixed selection. The low value item selected in the list (Table 2.1) is compost. Compost is generally available for free to all farmers (Centre, 2023), so any positive price would be unlikely to be accepted. Importantly, compost is widely available, consistently free, and frequently used. Even though SEED Hub is currently free, it is less widely used and does not have a clearly established market price, so it should not be viewed in the same way as compost.

Table 2.5 looks at the potential floor and ceiling effects. The first thing to note is the rate of misunderstood or declined responses. A respondent was recorded as misunderstood if they repeatedly tried to answer each individual item or if they gave a number outside of the allowable range for their group. The number of respondents that misunderstood is similar across list group assignment and price level, so that does not appear to be a threat to list experiment validity. We also see a relatively low frequency of zero and three responses in Group B, aside from price 150. In addition, the rates are comparable between Group A and Group B indicating that it is unlikely respondents are strategically changing their responses to avoid detection.

Price 150 does not seem to follow the same pattern as the other prices. Here, we see a higher rate of zero responses compared to other prices and in Group A relative to Group B. An answer of zero among Group A would reveal the socially undesirable response (not valuing SEED Hub), so the higher number of zeros among Group A compared to Group B does not suggest respondents are strategically altering their responses. However, if there were strategic altering of responses then that would lead to an overstatement of SEED Hub use by the indirect measure and an underestimate of social desirability bias, representing a lower bound. Comparing the frequency of zeros for price 150 compared to the higher prices is unexpected. However, the near-zero price could be such an undervaluation of the high value inputs that respondents are skeptical. Since this behavior seems to be relatively balanced between Group A and Group B, it should not bias the estimates of willingness to pay. Still, this suggests a smaller range of values may be preferred. In total, we do not find strong evidence that the no liars assumption is violated.

Finally, we proceed to the no design effects assumption which we anticipate being the most difficult to satisfy in an economic context. All of the items in the list are production inputs. There are no other information sources included in the list, so substitutes should not be a concern. Avoiding compliments is more difficult, as an increase in production information likely requires more of production inputs as well. However, the sources of information in available in SEED Hub (weather, production, and market prices) are more likely to influence planting, irrigation, and marketing decisions than fertilizer or labor decisions. Still, we would like to check for the possibility of design effects due to complementary goods. Since the goods remain constant across all prices, if design effects fail universally, then it is a sign that the complementary nature of the items is a problem. On the other hand, if there is a price cut off when the design effects become present, then that indicates that a binding budget constraint is more likely to be the driver.

We use the Blair and Imai (2012) test for design effects. The test estimates the probability (π_{yt}) that a response will be given, where y is the number of non-key items and t indicates the sensitive response. Any negative probability indicates the design effect assumption is violated. The results of this test are in 2.6. The second part of the Blair and Imai (2012) tests for stochastic dominance, where $\#P$ is the Bonferroni-adjusted p-value. The results of this test are in Table 2.7.

The Blair and Imai (2012) test indicates in both Table 2.6 and 2.7 that the design effects are not universally violated. Instead, at a price of 600 per item the no design effects assumption is violated, indicating this is the price where the budget constraint is binding. To purchase all of the items in the list (three or four depending on the list group assignment) would be roughly equivalent to a day of labor. The negative probabilities are closer to zero at higher prices, indicating that as the maximum potential budget increases the impact of having one additional item is smaller. If the budget constraint would only allow an individual to purchase one good at the higher price the impact of having four items compared to three is smaller than at the lower prices where an individual may have been able to afford more items with their existing budget. This supports the suggestion that the budget constraint is the limiting factor, rather than the goods being complements.

Considering all three of the necessary conditions for a valid list, the list experiment at the price of 150 is likely to provide reliable responses. We show the results for the other prices, but they should be viewed skeptically. Given the existing literature on the misestimation of WTP near zero, the results for price 150 are of particular interest.

2.5.2 Willingness-to-Pay

We find a downward sloping demand curve with a steep drop off after 150 LKR per month when using the direct WTP questions (Figure 2.1). When asked directly, 63.6% of respondents indicated they would be willing to pay 150 LKR per month for the SEED Hub platform. Putting this number in context, 77.8% of people reported being aware of SEED Hub but only 30.6% indicated they had downloaded the app. The stark difference between those who had interacted with the application and those who were willing to pay a positive price for the application indicates that direct WTP is likely overstated. Overstatement of WTP at low prices aligns with our hypothesis that individuals may overstate their WTP when a product had previously been given for free and is associated with a program.

The indirect demand curve is initially downward sloping, with a slight upward slope at the high prices.[11] As theorized in Section 2.2.1 and demonstrated in Section 2.5.1, a binding budget constraint violates the no design effects assumption for prices 600 LKR and above. The assumptions required for valid estimates at those prices are unlikely to hold, so the

[11] Despite concerns about internal validity, the results are similar with and without the inconsistent responses. See Appendix Figure B.1

results should be viewed with caution. Still, we see that at the higher prices when the budget constraint holds, the indirect estimate exceeds the direct measure estimate. This supports the theoretical conclusion that the expenditure signal has a larger impact than the morality signal, $\frac{M_S}{M_H - M_E} < M_E \sum_{j \neq j^*}^{J} (wtp_j^s - wtp_j^t)$, in this scenario. This finding suggests that, in this case, at higher prices, the direct measure would be preferred. Further supporting the theoretical model, at the low price of 150, where the budget constraint is far less likely to be binding, the indirect measure is preferred, because there is no opportunity to signal and there is no marginal utility to overstating the budget for unvalued items.

The estimate at a price of 150 should be unaffected by the design effect violations present at the prices. At this price, the demand estimate is lower than the direct measure. By the indirect estimate, only 47.4% of households would be willing to pay 150 LKR per month for SEED Hub. The 16.3 percentage point difference is marginally significant (10% significance level). The lack of precision owes to the lower efficiency of the indirect method and the decreased sample size from inconsistent responses.

In terms of magnitude, the 16.3 percentage points difference is a nearly 30% decrease in the share of farmers willing to pay 150 LKR. The difference in share willing to pay at low prices is consistent with the claim that social pressures lead to an overstatement of WTP, particularly when the product has been previously given freely. When alleviated of the pressure to directly reveal the proposed product does not have a positive economic value, respondents are more willing to reveal a WTP of zero. This finding suggests that an indirect question may be useful for uncovering more accurate measures of WTP when the budget constraint is not binding and researchers are concerned about biased WTP at the extensive margin.

We would expect those with more experience with the SEED Hub application to have a higher WTP but they may be subject to more social pressure as well. Having benefited from the application previously, downloaders of SEED Hub may feel the need to reciprocate and overstate their economic value of the product. To explore this possibility, we limit the sample to only those who reported downloading the application. Since downloading the application is potentially verifiable, respondents may be less likely to give biased responses to the download question than other self-reported use characteristics. When including only individuals who downloaded the application, we predictably see an increase in the share willing to pay 150 LKR per month in both the direct and indirect measures. The direct measure increases to 75.7% willing to pay while the indirect measure increases more marginally to 53.8% willing to pay. Among this sample, the 21.9 percentage point gap translates to a 33.8% overstatement of willingness to pay the lowest price.

Dividing the sample by those who downloaded the application raises two important points: (1) impact of program engagement and (2) subsetting the sample. First, we see that when we limit the sample to those with the most interaction with the service, and therefore the most interaction with the program, we see a slightly larger difference between measurement strategies. It seems that while those who have more experience with the program are more likely to have a positive WTP, they are also associated with more overstatement of WTP. This indicates that the level of intensity of the program or policy associated with giving a product may also be related to mismeasurement. If this is the case, we may also expect to see different levels of WTP by the RCT treatment arm.

In Figure 2.3, we see that the RCT treatment group that received access and training did have a higher WTP by both direct and indirect measures compared to just the group with access only. However, Figure 2.3 also highlights an important limitation of the indirect estimation strategy. Returning to the second point, subsetting the data greatly reduces precision. When exploring smaller subsets of the full sample, the direct measure continues to provide precise estimates, even if accuracy is in question. On the other hand, as sample size diminishes, the list experiment estimations grows increasingly noisy. These smaller subsamples make statistically distinguishing estimates difficult.

2.6 Conclusions

In this study, we discuss the potential for an economic application of a list experiment. A common use of list experiments is to estimate the behavior or sentiment of a sensitive subject. The indirect nature of the list experiments reduces the impact of social pressures on responses. The same effect is also desirable when estimating WTP. Applying list experiments to an economic context can uncover WTP without social desirability bias and with a lighter touch than revealed preference methods require.

In a theoretical context, we demonstrate the benefits of an indirect question in reducing social desirability bias in stated WTP while highlighting the pitfalls of the method when individuals have a limited budget. Based on the conceptual model, we outlined the scenarios where a list experiment would and would not likely be useful. In addition, we discuss the ways a list experiment must be adjusted to be administered in an economic context. The randomization assumption can be satisfied without adjustment. Similarly, including a high and low value item increases the likelihood that the no liars assumption holds. However, we find that the no design effects assumption is more difficult to satisfy in an economic context. The items in the list should be topically similar, yet not complements or substitutes. More pressingly, the budget constraint should not be binding at any of the potential prices in the set, a concern highlighted in the conceptual framework. The range of prices should also be selected strategically. A wide range of prices makes selecting non-sensitive items more challenging, while a tighter range of prices uncovers a smaller part of the demand curve, representing an important trade-off.

This last consideration—avoiding a binding budget constraint—limits the general applicability of list experiments for WTP estimation. However, a key area where this method holds promise is in estimating WTP for products that have previously been provided for free or at subsidized prices. The discontinuous nature of zero and small positive prices makes a low price, where participants are deciding to enter the market or not, after a program intervention the most likely part of the demand curve to be biased. The indirect list method may provide a solution. Given the near-zero prices, budget constraints are unlikely to be binding so list experiments are a viable alternative. Our empirical application demonstrates that at low prices when the no design effects assumption is more likely to hold, direct WTP elicitation overstates WTP compared to the indirect list method.

To be implemented successfully, a list experiment should be piloted. The risk of not satisfying the validity assumptions is greater in an economic context than in a behavioral context, which itself suffers from a file-drawer problem (Blair, Coppock, and Moor, 2020). Meaning, that without strategic design, pre-testing, and piloting, there is a significant risk of having unusable responses. However, when done with care in appropriate circumstances, the economic application of a list experiment holds promise.

Our work is relevant for academics, NGOs, and governments that are interested in the future valuation of a product that is currently being given for free or at a substantially subsidized price. Understanding how recipients value a product given as a part of a program may be useful in decision-making about program expansion or extension and the future price of the product. However, future economic applications of a list experiment should be limited to lower value items or contexts ready for investment and the items in the list should be selected with caution. Ultimately, our findings suggest that economic applications of list experiments have the potential to reduce upwardly biased WTP found at the extensive margin.

Table 2.1: *List Experiment Instructions and Question*

List Instructions

Think of your current budget for purchasing farm tools, assets, or services. I will read you a list of four items or services, and I would like to know how many of these items or services you would purchase one or more of if each cost *Price*[a] Rs.? Do not tell me which you would purchase or the quantity, just the number of items you would purchase given the amount of money you have. Please put your hand behind your back. Lift a finger for every item that you would like to purchase at the price *Price* Rs. After I have read all the statements show me the number of fingers you have raised. That number should be the number of items you would like to purchase. Let's begin:

Group A	Group B
• 10 kg of urea	• 10 kg of urea
• Half day of hired labor	• Half day of hired labor
• **One month access to SEED Hub**[b]	• 1 kg of compost
• 1 kg of compost	

[a] Price was randomly selected from set of prices {150, 600, 1050, 1500}

[b] **One month subscription to market and weather information app** in the RCT control group

Table 2.2: *Share Willing to Pay Each Price by List Group*

Price	Group A	Group B	p-value
150	0.605	0.628	0.298
600	0.200	0.199	0.946
1050	0.061	0.052	0.394
1500	0.041	0.033	0.371
N	997	998	

Table 2.3: *Inconsistent Responses By Question Type*

Share Inconsistent	Group A	Group B	p-value
Direct	0.012	0.007	0.253
Indirect	0.291	0.292	0.994

Table 2.4: *Randomization Balance Table*

	Group A	Group B	p-value
Household size	3.961	4.041	0.290
Share with Male Head	0.501	0.499	0.837
Maximum Education	13.453	13.626	0.144
Number of Plots	3.592	3.520	0.368
Total Plot Area	4.492	4.424	0.891
Hired Male labor	5.870	5.993	0.804
Hired Female labor	0.443	0.410	0.719
HH male labor	1.505	1.468	0.878
HH female labor	0.506	0.507	0.991
Harvest (kg)	1415.692	1617.083	0.360
Largest plot	2.700	2.576	0.758

Table 2.5: *No Liars Assumption Floor & Ceiling Frequency Check*

	Price 150				**Price 600**			
	Group A		Group B		Group A		Group B	
Levels	Freq	Percent	Freq	Percent	Freq	Percent	Freq	Percent
0	106	12.71	155	17.67	30	3.8	25	3.17
1	217	26.02	244	27.82	273	34.56	396	50.19
2	185	22.18	233	26.57	319	40.38	268	33.97
3	165	19.78	188	21.44	95	12.03	59	7.48
4	106	12.71	-	-	31	3.92	-	-
Misunderstood	33	3.96	40	4.56	33	4.18	35	4.44
Declined	22	2.64	17	1.94	9	1.14	6	0.76
	Price 1050				**Price 1500**			
	Group A		Group B		Group A		Group B	
Levels	Freq	Percent	Freq	Percent	Freq	Percent	Freq	Percent
0	30	4.88	20	3.37	50	5.3	58	6.09
1	240	39.02	345	58.08	422	44.7	521	54.67
2	265	43.09	175	29.46	300	31.78	301	31.58
3	41	6.67	25	4.21	121	12.82	36	3.78
4	6	0.98	-	-	8	0.85	-	-
Misunderstood	25	4.07	26	4.38	32	3.39	34	3.57
Declined	8	1.3	3	0.51	11	1.17	3	0.31

Table 2.6: *No Design Effect Assumption Test*

	Price 150				Price 600			
Non-Key Value	π_{y1}	**Robust SE**	π_{y0}	**Robust SE**	π_{y1}	**Robust SE**	π_{y0}	**Robust SE**
0	0.24	0.13	0.13	0.02	-0.03	0.01	0.17	0.05
1	0.05	0.04	0.24	0.03	0.16	0.04	0.41	0.03
2	0.12	0.03	0.17	0.03	0.08	0.02	0.26	0.03
3	0.14	0.02	0.07	0.03	0.03	0.01	0.03	0.02
	Price 1050				**Price 1500**			
Non-Key Value	π_{y1}	**Robust SE**	π_{y0}	**Robust SE**	π_{y1}	**Robust SE**	π_{y0}	**Robust SE**
0	-0.02	0.02	0.06	0.01	-0.002	0.02	0.06	0.01
1	0.23	0.13	0.47	0.03	0.12	0.03	0.45	0.03
2	0.03	0.02	0.27	0.03	0.10	0.02	0.23	0.03
3	0.01	0.01	0.04	0.01	0.01	0.01	0.03	0.01

Table 2.7: *No Design Effect Assumption Joint Test*

	Test for design effects (with GMS)							
	Price 150				**Price 600**			
Ha:Pr<0	**K**	**Lambda**	**P>Lambda**	**#P>Lambda**	**K**	**Lambda**	**P>Lambda**	**#P>Lambda**
Pr(R,S=0)	0	0	1	1	0	0	1	1
Pr(R,S=1)	0	0	1	1	1	6.19	0.01	0.01
	Price 1050				**Price 1500**			
Ha:Pr<0	**K**	**Lambda**	**P>Lambda**	**#P>Lambda**	**K**	**Lambda**	**P>Lambda**	**#P>Lambda**
Pr(R,S=0)	0	0	1	1	0	0	1	1
Pr(R,S=1)	1	1.25	0.13	0.26	0	0.01	0.46	0.92

Figure 2.1: *Partial Demand Curves for Seed Hub by Question Type*

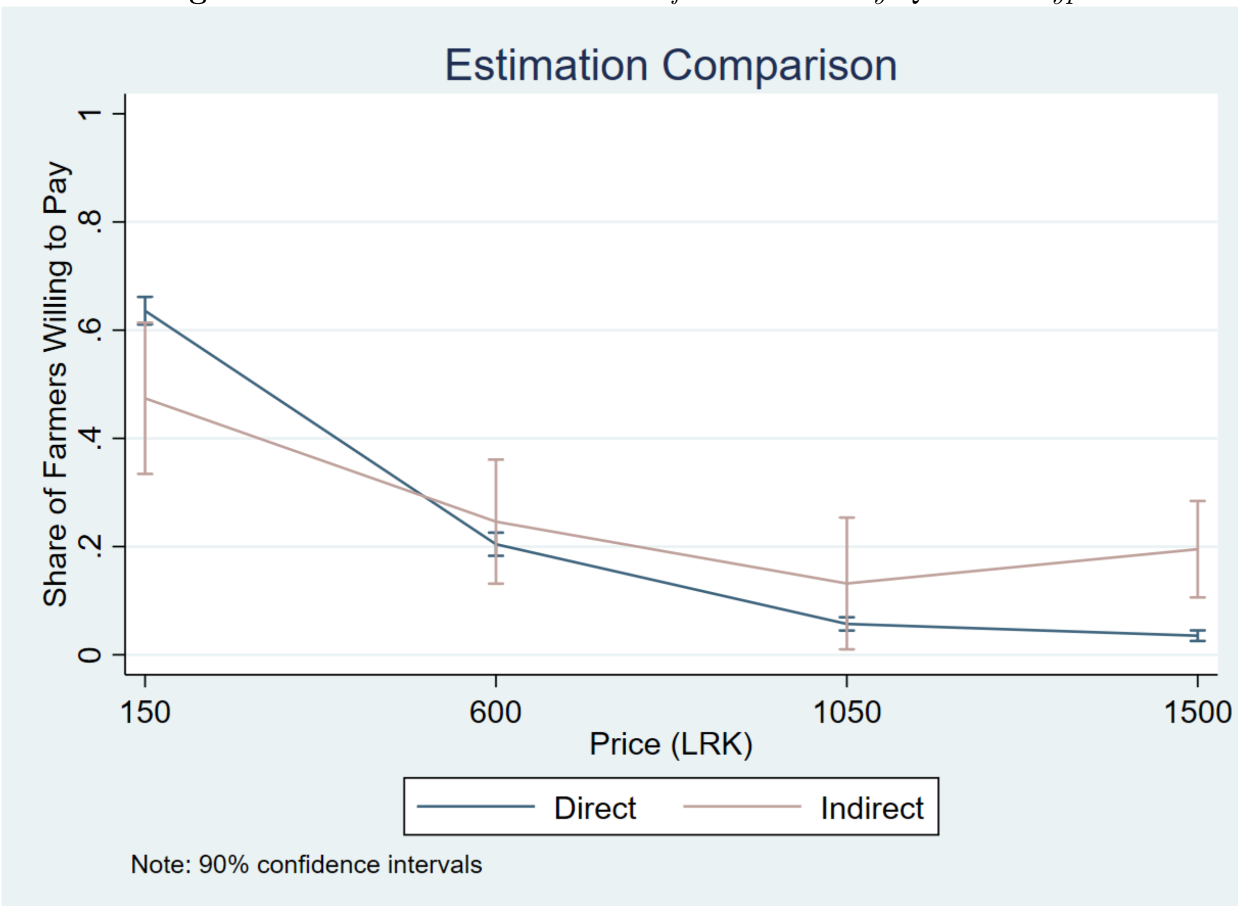


Figure 2.2: *Partial Demand Curves for Seed Hub by Question Type Among Users*

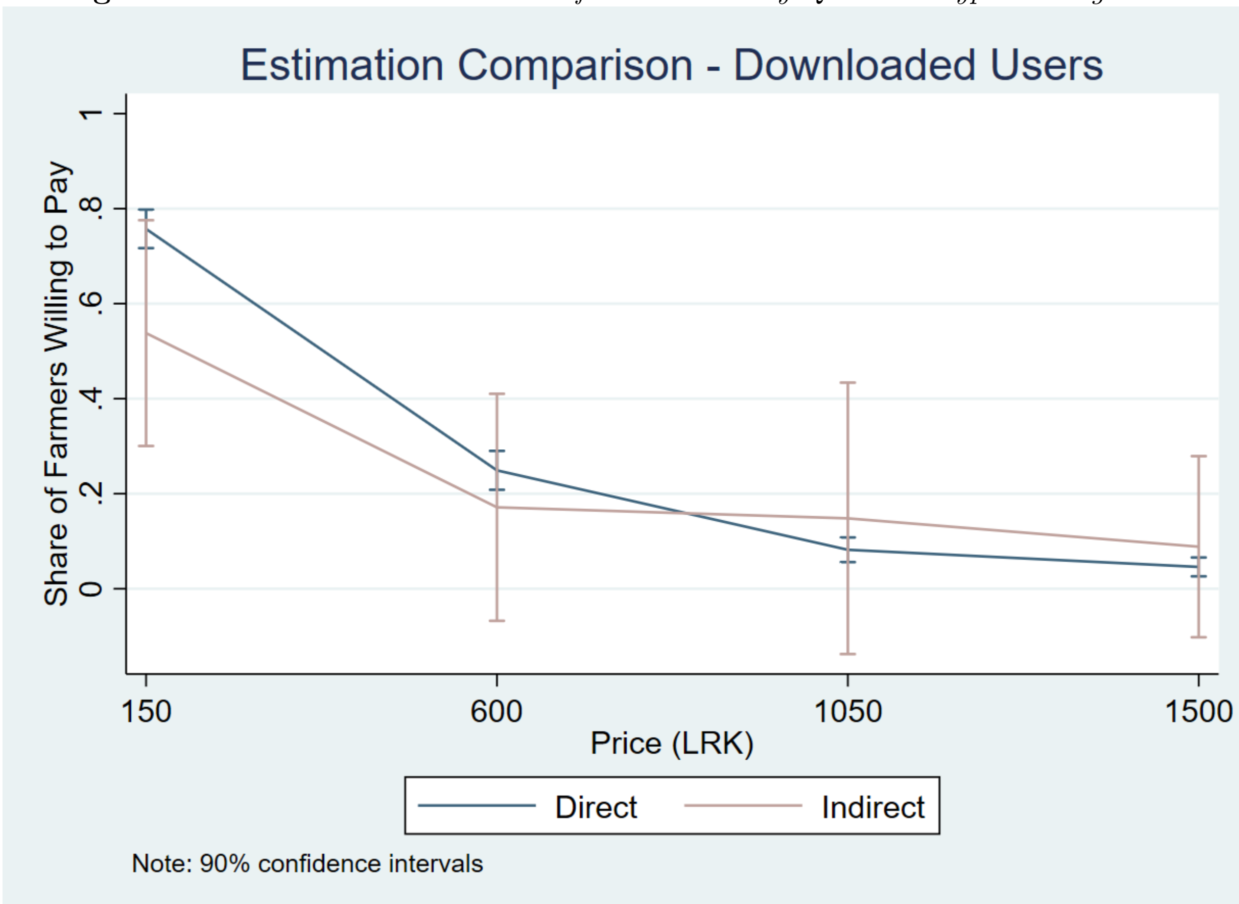
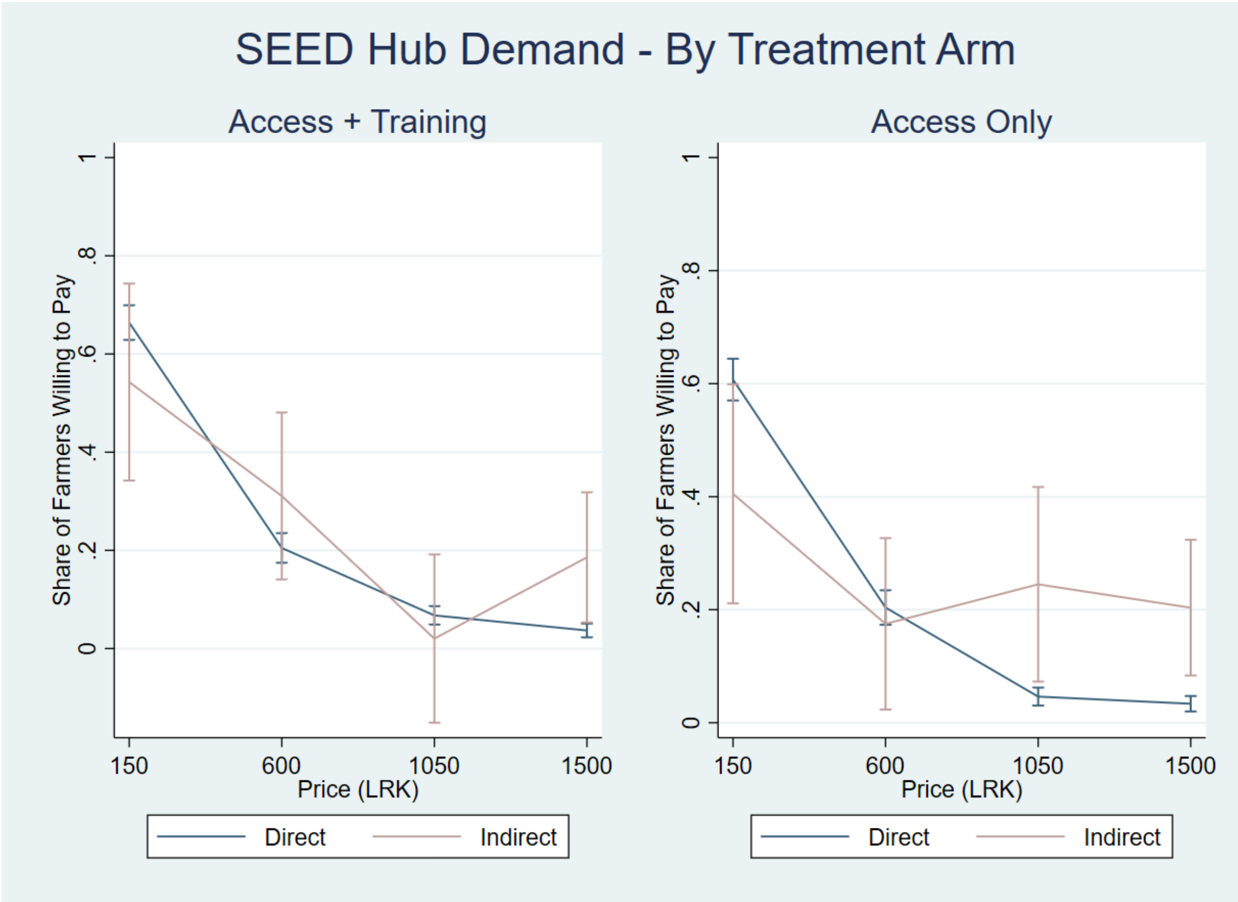


Figure 2.3: *Partial Demand Curves for Seed Hub by Question Type and RCT Treatment Arm*



Chapter 3

Recency Effects: A Heuristic for Pasture, Rangeland, and Forage Insurance Decision Making

3.1 Introduction

Decision makers often use heuristics, or mental shortcuts, to simplify complex decisions (Kahneman and Frederick, 2002; Tversky and Kahneman, 1973, 1974). Availability is the ease with which information comes to mind, and is often used as a proxy for the probability an event will occur (Tversky and Kahneman, 1973). When using the availability heuristic, individuals evaluate the likelihood of an event based on the ease with which they can recall a similar event. The ease of recall is a substitute for complex probability weighting (Kahneman and Frederick, 2002).

Prior literature has found that the availability heuristic pervades agricultural domains from weather predictions (Moore et al., 2019), to crop selection (Demnitz and Joslyn, 2020), and particularly insurance decision making (Cole, Stein, and Tobacman, 2014; Fezzi, Menapace, and Raffaelli, 2021; Karlan et al., 2014; Stein, 2016; Turner, Said, and Afzal, 2014). An optimal insurance purchase decision requires accurate estimation of the probability distribution of a wide variety of outcomes, an arduous task (Cole, Stein, and Tobacman, 2014). Instead, decision makers will often use recent events as a proxy for the likelihood of future events when making insurance enrollment and coverage decisions.

This phenomenon has been uncovered in insurance decision making all over the world. Fezzi, Menapace, and Raffaelli (2021) find significant recency effects in crop insurance decisions among apple growers in Italy. Turner, Said, and Afzal (2014) find increased uptake of flood insurance among households impacted by the 2010 flood in Pakistan, relative to non-flooded households with similar risk profiles. Receiving an insurance payout increases farmers' enrollment in that insurance product in the future in Ghana (Karlan et al., 2014). Within the United States, flood insurance enrollment spikes following a flood event, an effect that persists several years after the flood before returning to baseline (Atreya, Ferreira, and Michel-Kerjan, 2015; Gallagher, 2014).

Recently, Che, Feng, and Hennessy (2020) found that higher past indemnities encourage both enrollment and increased coverage levels in the Federal Crop Insurance Program decisions for corn and soybeans. The effect of recent indemnities is short-lived, peaking the year after a payment. We may expect to see similar, or stronger, recency impacts among Pasture, Rangeland, and Forage (PRF) index insurance.

PRF insurance is a novel product that is underused compared to other federal crop insurance programs, partly because it is an index insurance.[1] Rather than relying on farm level data, index insurance relies on weather indices to trigger payments. When rainfall, for example, falls below a historical index, a payment is given, alleviating the need for personal data or monitoring. Index insurance is commonly used in low- and middle-income countries due to its low transaction costs and relatively low data requirements (Collier, Skees, and Barnett, 2009; Miranda and Farrin, 2012). These are key benefits to administering insurance in a low- or middle-income country. In the United States, index insurance is comparatively uncommon because yield data for field crops is readily available and transaction costs are low. However, yield data for forage crops is rare.

In the United States, producers have little experience with comparable index insurance products, which adds to the complexity of optimal insurance decision making for PRF (Cho

[1]Further details about the PRF program implementation and design are available in Section 3.2.

and Brorsen, 2020). As such, we may expect producers to use recent events as a heuristic in PRF decision making. Our objective is to estimate the impact of recent events in PRF enrollment on the decision to enroll in PRF and the risk level of interval selection. The recent events of interest are (1) PRF indemnities relative to premiums, (2) Livestock Forage Disaster Program (LFP) payments, and (3) weather, all in the previous year. The risk level decision of interest is the PRF interval selection; more specifically, enrollment in the growing season compared to the non-growing season. The findings of this paper contribute to the growing field of literature examining PRF insurance.

Current PRF literature has focused on basis risk (Cho and Brorsen, 2020; Keller and Saitone, 2022; Yu et al., 2019), moral hazard (Shrum and Travis, 2022), land value (Ifft, Wu, and Kuethe, 2014), interaction between goverment programs (Goodrich and Davidson, 2024; Yu, Goodrich, and Graven, 2022), and optimal program decision making (Cho and Brorsen, 2020; Davidson and Goodrich, 2023; Goodrich, Yu, and Vandeveer, 2020; Goodrich and Davidson, 2024; Westerhold et al., 2018; Zapata and García, 2022). Since forage yields are not frequently documented, many of these studies focus on small geographic regions with available yield data, like Kansas and Nebraska (Goodrich, Yu, and Vandeveer, 2020; Westerhold et al., 2018; Yu et al., 2019), Oklahoma (Cho and Brorsen, 2020), or South Texas (Zapata and García, 2022). Other work has leveraged larger regional surveys of producers to better understand the role of awareness (Goodrich and Davidson, 2024) and informational nudges (Davidson and Goodrich, 2023). Yu, Goodrich, and Graven (2022) and Ifft, Wu, and Kuethe (2014) are notable national level studies.

Our work contributes to recent literature by bringing the behavioral economics concept of recency effect to PRF insurance. This line of literature was introduced by Tversky and Kahneman (1973) and implemented in the realm of crop insurance by Che, Feng, and Hennessy (2020). Given the novelty of PRF compared to other Federal Crop Insurance Program products, we may expect recency effects to be stronger among PRF. It is natural to substitute recent experience for complex calculations, especially when experience is low. We may also expect to see recency effects diminish as experience with the program increases. In both cases, understanding the implications of recent events among PRF decisions is relevant.

In addition, we add to the existing interval selection literature. Goodrich, Yu, and Vandeveer (2020) develop a conceptual model, and then demonstrate empirically that enrollment in the non-growing season is risk increasing and also associated with increased returns and higher subsidy levels. Unsurprisingly, the authors find increasing enrollment in the non-growing season over time. The analysis of the data in this chapter, as evidenced in Figure 3.3, supports that finding. We further add to this line of inquiry as we evaluate the impact of recent events on current decisions. The analysis here considers whether this shift to the non-growing season is in part explained by past success in risk-increasing behavior.

Finally, our study adds to the literature that identifies interactions between government programs by considering recent LFP payments in our analysis. Prior work by Yu, Goodrich, and Graven (2022) demonstrated that PRF influences other government program participation decisions, finding that PRF decreases Conservation Reserve Program (CRP) enrollment. Specifically, CRP precludes benefitting from PRF enrollment, so the negative relationship is in line with expectation. Conversely, LFP previously required PRF enrollment to be eligible for payments. While no longer required,[2] the mental connection between programs may persist. As a result, we may expect recent LFP benefits to play a role in PRF decision making. Furthermore, we investigate whether the relationship between LFP benefits and PRF decision making changes as a result of the modification of LFP eligibility requirements. When a risk management tool, like PRF, was required to receive LFP payments, the two programs are likely to be perceived as complements. However, it is unclear if this relationship persists or changes so that the programs are seen as substitutes when the LFP payment requirements change. We explore the effect of a recent LFP experience on PRF enrollment decisions to explore the nature of the PRF and LFP relationship.

Our findings suggest that a recent increase in PRF indemnities relative to premiums

[2]This is discussed further in Section 3.2.

is positively related to current enrollment in PRF. The finding is consistent across hay and grazing practices. Recent PRF indemnities have no statistically significant relationship with interval selection. LFP payments are associated with PRF enrollment and selection of growing season intervals. Our findings are consistent with recent work by Goodrich and Davidson (2024), which suggests LFP and PRF may be viewed as complements. Rainfall predictably has an inverse relationship with enrollment, while average temperature has a positive relationship with enrollment.

The remainder of this paper is structured as follows. First, we give background on the two government programs of interest (PRF and LFP). Next, we discuss the data used before outlining our empirical approach. Then we will discuss the results of our analysis, followed by our concluding thoughts.

3.2 Background

3.2.1 Pasture, Rangeland, and Forage Rainfall Insurance

PRF insurance is a rainfall index insurance that protects livestock producers from drought-induced shocks to hay and grazing feed supplies. The program was piloted as a part of the Federal Crop Insurance Program in 2007. In the years that followed, PRF gradually expanded on the county level. As of 2016, PRF is universally available in the contiguous 48 states. PRF enrollment has been increasing over the past decade. Since PRF became universally available, enrollment has increased from 43.8 million acres in 2016 to over 250 million in 2023 (Figure 3.2).

As discussed in Section 3.1, PRF is distinct from other popular federal crop insurance products because it is an index insurance. Rather than relying on own historical production data, PRF uses a historical index. Indemnities are triggered based on deviations from the historical rainfall index. The index is constructed based on precipitation within a 0.25 degree in latitude by 0.25 degree in longitude grid, which is 17 by 17 miles at the equator, for each time interval (RMA, 2024b).

With index insurance, there is potential for basis risk, or the risk that the payment determining index does not match the realized outcome for an individual. While not the focus of this paper, recent literature has identified basis risk in PRF insurance (Cho and Brorsen, 2020; Keller and Saitone, 2022; Yu et al., 2019). Yu et al. (2019) find that using more granular rainfall measures only marginally decreases basis risk. Similarly, Cho and Brorsen (2020) find a close link between the index and realized rainfall. On the other hand, Keller and Saitone (2022) find that basis risk drives shortcomings in payments.

When enrolling in PRF, a producer must decide their coverage level, which intervals to insure, and the productivity factor. The coverage level varies from 70% to 90% in five percent increments. The coverage level determines how far the rainfall index must fall below the historical index for a payment to be triggered. The productivity factor ranges from 60% to 150%. The productivity factor adjusts the level of productivity for a producer, using the productivity of the county under normal weather conditions as a base. Higher productivity factors result in higher levels of protection, while also increasing premiums paid by producers. Cho and Brorsen (2020) discusses the optimal productivity factor selection in detail.

The subsidy rate on the premium paid by producers varies by coverage level, with higher coverage levels receiving a lower subsidy rate. There are eleven potential two-month intervals to select.[3] A producer must choose at least two intervals but cannot select overlapping intervals (ie, January-February and February-March), meaning a maximum of six intervals may be selected. The subsidy rate is constant across intervals, but is a function of the premium paid. Premiums vary across intervals, so the dollar value of the subsidy varies by interval. Rainfall is more variable in the non-growing season winter months. As a result, non-growing season premiums are higher, and therefore so is the dollar value of subsidies (Goodrich, Yu, and Vandeveer, 2020).

[3]The intervals are: January-February, February-March, March-April, April-May, May-June, June-July, July-August, August-September, September-October, October-November, and November-December.

3.2.2 Livestock Forage Disaster Program

The LFP was authorized through the 2008 Farm Act, but ad hoc payments were previously given, dating back to 1997. The program provides payments to livestock producers after forage loss due to drought. Unlike PRF, which requires enrollment, LFP is issued to those who are eligible. Qualification is determined based on production characteristics and drought conditions. Livestock must be primarily grazing, and there are farm and off-farm income maximums as well.

Payments are given based on a combination of depth and length of drought, which uses the U.S. Drought Monitor's drought designation. The amount of the LFP payment is determined by annual feed costs. The goal of the payment is to provide 60% of a monthly feed cost for an animal for the months affected by drought.[4] Defraying feed costs is meant to relieve pressure to liquidate herds in times of drought. Hrozencik (2024) find that among beef cattle, stocking rates of counties receiving LFP payments are similar to counties with less severe drought, which provides suggestive evidence that LFP payments are serving their intended purpose.

From 2008 to 2014, producers must have purchased a risk management product to receive LFP payments. Enrollment in PRF satisfied the requirement.[5] However, this restriction was lifted in 2014. The change in eligibility requirements may fundamentally change the relationship between PRF and LFP, a possibility explored previously in Section 3.1.

3.3 Data

Multiple data sources are used in this analysis. The data is available from all sources from 2010 to 2023, representing our full data time series. We further subset the data to the 2016 to 2023 timeframe as a secondary analysis. The subset represents the timeline where PRF

[4]LFP payments are not limited to grazing cattle; other livestock, including buffalo, sheep, goats, elk, etc., are also eligible.

[5]The alternative risk management product that satisfied the requirement is the Noninsured Disaster Assistance Program (NAP). NAP provides coverage for products not covered by federal crop insurance. Products covered by NAP are fruits, honey, mushrooms, turfgrass, among others.

is universally available so that the data is not subject to changing geographic make-up and PRF availability.

The PRF data comes from the Risk Managagement Agency (RMA) summary of business at the type, practice, unit level (RMA, 2024b). The PRF data includes county-level data on indemnities, premiums, subsidies, and acres enrolled. The variables of interest were summed across interval and coverage level selections, giving crop, county, year level data. Grazing and hay, the two potential forage categories, were considered separately. The summary statistics in Table 3.1 and Figures 3.2 and 3.3 suggest the producers employ PRF differently across the two forage practices. As such, the impact of recent events is likely to vary between hay and grazing practices. The state, county, crop, and coverage level summary of business from RMA was used as a robustness check. The aggregated data precludes interval selection analysis, but does include the number of policies sold and indemnified, metrics not available at the more granular level. In addition, irrigated forage acres were removed from our sample.

To calculate the share of acres enrolled in PRF by county, we use the pastureland acres in 2002 for each county as the base.[6] The expansion and increased uptake of PRF since it began in 2007 may have influenced land allocated to pastureland. To omit any changes in land use as a result of PRF, we use a constant denominator for our share enrolled variable. Therefore, we use 2002 pastureland acreage as the base, since that was the last census before PRF implementation. As a robustness check, we repeat the base analysis with 2007 pastureland acres as well. Since 2007 was the first year the program was available, the impact on total pastureland in a county is likely to be minimal. The county-level data were drawn from the United States Department of Agriculture National Agricultural Statistics Service census (USDA, 2002). Any county with unavailable pastureland in 2002 was dropped from our sample.[7]

The LFP data is available from the Farm Service Agency (FSA) Payment Files Informa-

[6]Specifically, the metric used is: AG LAND, PASTURELAND, (EXCL CROPLAND & WOODLAND) - ACRES

[7]This included ten counties with non-zero PRF enrollment. Alternative imputation strategies could be considered in future expansion of this work.

tion (FSA, 2020). This data source includes LFP payment amount and is available at the individual level for each year. The LFP payment amount was aggregated to the county level for each year to remain comparable with the PRF data. When aggregated to the county level, some counties with large payment totals, while other counties have zero payments. The distribution of the data has implications for model selection, which will be discussed in Section 3.4. The average annual LFP payments by county and the average share of counties receiving a payment in a given year are summarized in Table 3.1

Finally, we include weather information. The weather data was originally from PRISM (PRISM, 2024). The Lee (2024) download function in R Studio was used to retrieve the county-level monthly data. The monthly average temperature, growing degree days (8°C to 29°C), extreme degree days (>29C), and total precipitation (mm) were pulled. Interval growing season average temperature, growing degree days, and extreme degree days were calculated along with annual total precipitation (cm). Weather variables are summarized in Table 3.1. The temperature and precipitation averages vary by hay and grazing in the summary table, even though the variables do not vary by crop. This is because not every county has both grazing and hay acres, so the counties summarized by crop are different.

3.4 Methods

We employ a two-way fixed effects approach with lagged variables to estimate the relationship between recent events and current PRF decision making. Weather influences current decision making directly, by increasing the mental availability of weather events, and indirectly through PRF and LFP program payments. Since rainfall is the determinant of PRF and LFP payments, disentangling the relationship between current PRF decision making and weather events, LFP payments, and PRF indemnities concurrently is infeasible. Instead, we estimate the indirect and direct channels separately. The county level fixed effects account for time-invariant factors like program availability, local production conditions, and crop insurance agent, which is closely tied to crop insurance awareness (Goodrich and Davidson, 2024). The year fixed effects account for location-invariant factors, like changes in national agricultural policy, prevailing market conditions, and the cattle cycle.

We are interested in the recency effects on two dimensions of the PRF enrollment decision- the extensive margin and the risk level. The extensive margin captures the decision to enroll in PRF or not. We hypothesize that the recent negative weather events, such as hot temperatures or low rainfall, will be associated with increased PRF enrollment in the following year. We also expect past PRF gains to have a positive relationship with current PRF enrollment. As noted briefly in Section 3.1, we expect the relationship between LFP and PRF to be variable, dependent on LFP requirements.

The risk level decision is based on the intervals selected. The eleven possible intervals are each assigned as growing season and non-growing season. Prior literature indicates that the growing and non-growing seasons have different risk profiles, where non-growing season intervals are risk-increasing (Davidson and Goodrich, 2023; Goodrich, Yu, and Vandeveer, 2020). We hypothesize that recent LFP gains are associated with risk-increasing behavior, captured by increased enrollment in the non-growing season. Having recently received an LFP payment may cause producers to overstate the probability of an LFP payment in the following year. If a producer thinks an LFP payment will cover losses in the important forage growth months, then the PRF may be used in a risk-increasing way. Recent PRF success in the growing season may be associated with continued enrollment in the growing season.

For the extensive margin, the dependent variable is the logit transformation of the share of 2002 pastureland acres enrolled in PRF for a given year. So that $ShareEnrolled^{c}_{f,y} = \frac{Acres^{c}_{f,y}}{TotalPastureland_{f,2002}}$ for each crop (c), fips (f), and year (y). The fractional nature of our dependent variables poses a methodological challenge, as the values from an OLS regression may feasibly fall outside of the [0,1] range. To overcome this challenge, we employ the logit transformation, which was originally proposed by Johnson (1949) and tested by Papke and Wooldridge (1996). Recently, this strategy was used in the recency literature space by Che, Feng, and Hennessy (2020) and in a crop insurance application by Connor and Katchova (2020).

By definition, a share of a total should fall within the [0,1] range. However, the total pastureland of a county may have increased over time due to PRF availability, so a share greater than one is possible. Observations outside of the allowable (0,1) range were top coded as 0.99999, as the logit transformation precludes observations outside of the zero one interval. Notably, we do not have any zero observations. Since we are using PRF enrollment data, all counties have some positive enrollment. As a robustness check, we also use a log transformation presented in Section 3.5.3. The extensive margin dependent variable is defined in Equation 3.1.

$$SE^c_{f,y} = log\left(\frac{ShareEnrolled^c_{f,y}}{1 - ShareEnrolled^c_{f,y}}\right) \tag{3.1}$$

Recent weather events and recent program events are estimated separately. This gives two time and county fixed effects equations as follows:

$$SE^c_{f,y} = \beta^c_0 + \beta^c_1 GS\_Temp_{f,y-1} + \beta^c_2 Precip_{f,y-1} + \theta^c_f + \Phi^c_y + \epsilon^c_{i,t} \tag{3.2}$$

$$SE^c_{f,y} = \beta^c_3 + \beta^c_4 IR^c_{f,y-1} + \beta^c_5 LFP^d_{f,y-1} + \theta^c_f + \Phi^c_y + \epsilon^c_{i,t} \tag{3.3}$$

In Equation 3.2, the recent effect variables are the average growing season temperature variable ($Temp_{f,y-1}$), and the total precipitation ($Precip_{f,y-1}$) from the previous year for county f. In Equation 3.3, the recent event variables are the lagged indemninty ratio ($IR^c_{f,y-1}$) and the LFP payment ($LFP^d_{f,y-1}$), by two definitions. In both equations, θ^c_f are the county fixed effects and Φ^c_y are the year fixed effects.

The indemnity ratio for practice c is defined as $IR^c = \frac{Indemnities}{Premiums}$. We opted not to add the subsidy to the denominator, which would capture the loss ratio, as the premium is the perceived cost of the producer. An indemnity ratio greater than one would indicate for that practice (hay or grazing) in county f in year y, the total indemnities paid exceed the total payments made. The productivity factor selection influences the premium paid, so it is not included as a separate decision factor.

As noted in the Section 3.3, the LFP variable has a significant number of large positive payments in addition to many zero payments. Including the unscaled LFP payment presents an interpretation challenge. As such, the LFP payment was scaled using the inverse hyperbolic sine (IHS) transformation. The IHS transformation allows for zero observations in the transformation, unlike a log transformation (Burbidge, Magee, and Robb, 1988). The interpretation of the IHS transformed variable coefficient mirrors that of a log transformation. However, recent literature has highlighted that the IHS transformation is sensitive to the arbitrary scaling of the input variable (Aihounton and Henningsen, 2021). We confirm this finding, showing that the distribution of the IHS transformed LFP variable is highly sensitive to the scale of the input LFP variable (dollars compared to thousand dollars), highlighted in Appendix Figure C.1. The impact of the unit on the IHS transformation muddles the interpretation of the transformed variable coefficient, so that it can no longer be interpreted like a log transformation. With the ambiguity surrounding the optimal transformation, we also include the LFP payment as an indicator variable, as an alternative specification. Therefore, the definition d of variable $LFP^d_{f,y-1}$ is either the IHS transformed variable in dollars or the binary LFP payment variable. We refrain from selecting a preferred specification and present both specifications in Section 3.5.

The average temperature was selected in our base model as the most similar variable to a weather heuristic. While a more granular temperature bin approach may be more closely related to forage outcomes, a variable that captures the typical weather in the previous year is cognitively simpler. A measure like average temperature is more in line with the type of simplifying heuristics we are interested in measuring. Other weather specifications should be further explored as a robustness check of this work. Similarly, total precipitation was used as that is a common unit of measurement, matching heuristics. We opted to do a log transformation to improve the interpretability of the coefficient, as there was a wide range of precipitation.

The risk level depedent varaible is defined as $ShareGrowingSeason^c_{f,y} = \frac{AcresGrowingSeason^c_{f,y}}{TotalEnrolledAcres_{f,y}}$ for each crop (c), fips (f), and year (y). The growing season is May through July, which are the most critical rainfall months for forage in large pastureland areas (Smart et al., 2021). May to July would include intervals four through seven. The fourth interval is April-May, and the seventh interval is July-August, which captures the widest definition of the May to July growing season. As a robustness check, we define the growing season more broadly, including the third and eighth intervals. The same logit transformation used for the extensive variable was implemented for the growing season interval selection variable. Here, since the total acreage base is from the same year and the growing season definition was broad enough, there are no observations outside of the allowable range. This gives the risk level dependent variable defined in Equation 3.4.

$$GS^c_{f,y} = log\left(\frac{ShareGrowingSeason^c_{f,y}}{1 - ShareGrowingSeason^c_{f,y}}\right) \tag{3.4}$$

$$GS^c_{f,y} = \beta^l_0 + \beta^c_1 GS\_Temp_{f,y-1} + \beta^c_2 GS\_Precip_{f,y-1} + \theta^c_f + \Phi^c_y + \epsilon^c_{i,t} \tag{3.5}$$

$$GS^c_{f,y} = \beta^c_3 + \beta^c_4 GS\_IR^c_{f,y-1} + \beta^c_5 LFP^d_{f,y-1} + \theta^c_f + \Phi^c_y + \epsilon^c_{i,t} \tag{3.6}$$

Equation 3.5 is distinct from the enrollment equation 3.2 in the definition of the precipitation variable, which now only measures precipitation in the growing season. Similarly, 3.6 measures the indemnity ratio for the previous year's growing season only. The LFP payments cannot be disaggregated by season, and so both definitions of the LFP payments are unchanged.

3.5 Results

3.5.1 Share Enrolled in PRF

The recency effects for lagged weather variables (where the =L indicates a lagged variable in all results tables) are similar for hay and grazing practices (Table 3.2). More rain in the previous year is associated with less PRF enrollment, while a warmer growing season in the prior year is associated with more PRF enrollment. The signs of the weather coefficients are in line with our expectations, as poor weather is positively related to future insurance enrollment. The coefficients are statistically different from zero (1% significance level).

Unlike the weather specifications, the size of the lagged PRF and LFP coefficients diverges between hay and grazing practices. When looking at hay, we can see that the lagged indemnity ratio coefficient is stable across LFP payment variable specifications. In columns two and three, the indemnity ratio coefficient is statistically significant (1% significance level) and positive. The indemnity ratio coefficient is also positive and significant at the 1% significance level for grazing, however, it is larger in magnitude. The positive coefficient for hay and grazing indicates that a higher indemnity ratio in the prior year, meaning higher indemnities relative to premiums, is associated with more PRF enrollment in the present year. The difference in magnitude of the coefficient indicates that the prior year's indemnity ratio is more strongly related to PRF enrollment in grazing relative to hay acres.

The impact of LFP payments, by both definitions, is positive and statistically significant at the 1% significance level across both hay and grazing practices and the IHS and binary variable. Recent LFP payments are positively related to enrollment in PRF. One potential explanation for the significance of an LFP payment in the PRF enrollment decision is the LFP insurance requirement. Until 2014, to be eligible for an LFP payment, producers must have enrolled in an insurance product, like PRF. This requirement may contribute to the significance of the LFP payments in PRF decision making. We revisit this hypothesis when discussing Table 3.3 below. Furthermore, receiving an LFP payment this year may lead to overestimating the probability of extreme drought in the following year, since the recent drought and LFP payment come to mind easily. More producers may enroll in PRF because of the increased perceived likelihood of low rainfall, due to the availability heuristic.

Next, we constrain the timeline to the years that PRF is universally available, from 2016 to 2023. In this subset, the coefficients change only slightly from the full 2010 to 2023 time panel. The results of the universally available subset are presented in Table 3.3. One notable difference between the full sample and the subsample is the magnitude of the indemnity ratio for hay. In Table 3.3, we see the magnitude is much closer to the grazing coefficient and similarly significant at the 1% significance level. Figure 3.2 shows a large increase in enrollment, specifically among hay acreage from 2016 to 2023. The large increase in enrollment in hay reduces the gap between hay and grazing practices, and so hay begins to mirror grazing decision making.

LFP payments are still positively associated with subsequent PRF enrollment within the short timeframe. This is noteworthy, as the LFP requirements changed within the longer time frame (2014) so that PRF (or a similar product) enrollment was no longer a requirement to receive an LFP payment. Since there is no similar mandate in the shorter timeframe, we might expect to find smaller coefficients, or even a negative sign. However, the LFP coefficients remain remarkably steady, even slightly increasing. This stable relationship between LFP and PRF enrollment indicates that the two programs do not appear to be viewed as substitutes, as we hypothesized. Instead, the consistent positive relationship suggests the two are complements. It could be that the recent LFP payment makes low rainfall years even more salient. Using the availability heuristic, poor weather comes to mind more easily, so producers further protect against dry conditions by enrolling in PRF.

3.5.2 Share Enrolled in the Growing Season

The weather specification results in Table 3.4 suggest that both rainfall and higher temperatures in the previous growing season are associated with less growing season PRF enrollment.

The negative rainfall coefficient is in line with our expectations, as increased rainfall makes a PRF payment less likely. However, the negative coefficient on the average growing season temperature is unexpected. Increasing temperature during the growing season could influence the risk preferences of decision-makers. If the growing season becomes riskier (due in part to high temperatures), then the relative difference in risk between the growing and non-growing season decreases, which could explain increased enrollment in the non-growing season. Future research is needed to better assess this hypothesis.

The results in Table 3.4 indicate that the indemnity ratio is unrelated to the growing season interval selection for both the hay and grazing practices. The insignificance of the PRF coefficients indicates that the past indemnities in the growing season are unrelated to the current growing season interval selection. This provides some evidence that the availability heuristic does not transfer to seasonal allocation decisions after the enrollment decision has been made.

Recent LFP payments have a positive and statistically significant coefficient for both hay and grazing by both definitions. Prior LFP payments are associated with increased enrollment in the growing season, though the magnitude of the impact is smaller than observed in the enrollment decision analysis. This finding further supports earlier findings in Table 3.2 that suggest LFP and PRF are viewed as complements. Our findings suggest that LFP payments continue to be used as a heuristic for PRF decisions.

Moving to the 2016 to 2023 subsample in Table 3.5, neither the weather nor the PRF and LFP payment variables are statistically different from zero at conventional significance levels. This suggests that in recent years, when the program has become more widely available and used, the availability heuristic does not appear to drive interval selection decision-making. Even the weather variables, which may have some predictive power for future weather events, do not predict future interval selection. Other factors, related to risk preferences, may better shed light on interval selection decisions.

3.5.3 Robustness Checks

Alternative Base Year

As noted in Section 3.3, we use 2002 as the base year for our primary specification. We use 2007 as an alternative base year to check the robustness of our results to base year selection, as 2007 is unlikely to be significantly impacted by PRF. We replicate Table 3.2 using 2007 as the alternative base year in Table 3.9. Comparing the coefficients, we see the results in terms of sign, magnitude, and significance are robust to the selection of the base year. This gives further confidence to our base model specification.

Alternative Transformation

Our main results employed a logit transformation of two dependent variables. The logit transformation was selected given the [0,1] nature of a share variable. As a robustness check, we replicate our base analysis with a log transformation of our dependent variables. The comparison of the distributions by variable transformation is available in Appendix Section C.2.

Comparing the results from the logit transformation in Table 3.2 to the log transformation in Table 3.7, we see the results are stable across specifications. The magnitude and sign of the coefficients are consistent. When comparing the interval selection results in Table 3.4 to 3.8, the findings are similarly robust to the model specification. In the log transformation, the coefficients are slightly smaller in magnitude for grazing. Still, the subtle differences in coefficients do not qualitatively change our overall findings.

Alternative Growing Season Definition

We define the growing season to be May to July, in keeping with rangeland and prior PRF literature (Smart et al., 2021; Yu et al., 2019). In our main analysis, we define the growing season to be intervals four through seven. At its core, the growing season definition is meant to capture the intervals where rainfall is most important for forage growth. Widening the definition of growing season is more likely to dilute the impact of the key months by including non-key months within the growing season definition. Still, this is a narrow growing season most applicable for the northern United States. More closely aligning the growing season to a given location would be a more ideal solution, though data challenges recognized by Davidson and Goodrich (2023) make a more granular definition of the growing season a challenge.

To test our definition of the growing season, we redo our risk level analysis with a wider definition of the growing season, intervals three through eight. This more closely balances the intervals in the growing and non-growing seasons. The indemnity ratio coefficient is still not statistically significant at generally accepted significance levels for both hay and grazing. The LFP payments, however, are robust to the definition of the growing season, with minimal changes between Tables 3.4 and 3.9. Temperature is no longer significant for grazing after the definition of the growing season changed. Since the weather data is available monthly, the growing season can be defined more precisely as May through July, so the temperature and precipitation variable definition does not change.

3.5.4 Combined Hay & Grazing

In previous specifications, we have considered hay and grazing practices separately. In our final robustness check, we use the combined crop level summary of business from RMA (RMA, 2024a). The aggregated dataset includes new metrics of the number of policies sold and inmenified. So, this robustness check allows us to check both the aggregation of hay and grazing and also our PRF indemnity ratio coefficient. The first column of Table 3.10 shows that the sign, significance, and magnitude of the weather variables are stable when the practices are combined.

Columns two and three mirror the findings of columns two and five, and three and six in Table 3.2. The magnitudes of the PRf and LFP coefficients are larger after combining the hay and grazing practices, but the sign and significance levels are consistent. Our findings would not significantly change if we instead combined the practices.

Finally, we use the count of policies paying premiums and policies indemnified rather than the value of premiums and indemnities to calculate an alternative specification of the indemnity ratio. Here, the sign and statistical significance of the LFP coefficients are consistent. The count indemnity ratio coefficient is similarly positive and statistically different from zero. We refrain from comparing the magnitude of the value and count indemnity ratio coefficients, since the two ratios are not directly comparable. Still, the consistency of sign and significance level indicates that there is a positive relationship between past PRF indemnities and current PRF enrollment, irrespective of specification. Data limitations prevent implementing the count indemnity ratio robustness check on the interval selection.

3.6 Conclusion

Complex decisions are not made in a vacuum. The behavioral economics literature points to recent events being a common heuristic for difficult to optimize decisions. This study evaluated the relationship between recent events and current decision making for PRF index insurance. Specifically, we looked at the impact of recent PRF indemnities relative to premiums, LFP payments, and weather events. Our PRF outcomes of interest were classified as the extensive decision to enroll and the risk level growing versus non-growing season interval allocation.

Leveraging county-level PRF data, we implement a two-way fixed effect approach to account for county and year fixed effects. Our findings suggest that recent PRF gains are associated with increased PRF enrollment but are unrelated to interval selection decisions. LFP payments, on the other hand, consistently encourage PRF enrollment and the selection of growing season intervals. These findings suggest PRF and LFP are viewed as complements. The salient connection between LFP and PRF decisions is relevant for policymakers, as the two programs appear to be linked in the decision making process. Our analysis suggests the availability heuristic provides a possible mechanism linking the two programs, as LFP payments reinforce the salience of potential PRF benefits.

These findings shed light on producer PRF decision making. Our results show that recent events are relevant for current decision making. However, recent events are not always good predictors of future events. Recent events may not be a sufficient proxy for profit-maximizing or risk-minimizing decision-making, leading to non-optimal choices. The impact of recent events may be useful in guiding future information campaigns and policy. In addition, future work should quantify the impact of using non-optimal proxies on PRF outcomes.

Prior work has found that recent events have persistent impacts, reaching beyond a single year (Atreya, Ferreira, and Michel-Kerjan, 2015; Cho and Brorsen, 2020; Gallagher, 2014). In the future, this work could be extended to consider the lasting effects of PRF and LFP payments. Unlike a flood, with a definitive classification, a PRF indemnity does not have a clear classification of extreme. Similarly, it is unclear whether the binary of receiving an LFP payment or the magnitude of an LFP payment is likely to have a lasting impact. Future work identifying the persistent impact of payments will need to consider many classifications of the severity of events. However, if done well, the results would be insightful for the long-term implications for PRF decision-making.

In its current form, our work does not allow us to look at potential spill-over or peer effects of recent events. Gallagher (2014) find that individuals who were not impacted by a flood, but were in the same media exposure area and would have therefore experienced media coverage of the flood, also had slight increases in flood insurance uptake. Santeramo (2019) suggests that indirect experience with an insurance product, through learning from insured farmers, may reduce knowledge frictions and similarly increase enrollment. Goodrich and Davidson (2024) advocate for targeted information campaigns, highlighting that the source of information matters. Taken together, these findings suggest future extensions of this work should investigate the impact of recent PRF experiences among peers.

In total, our findings highlight the importance of behavioral components in PRF decision making. Recent experiences play a critical role in current PRF enrollment and interval selection - a finding relevant for policymakers and future literature. Understanding the dynamic elements of PRF decision making is useful for enhancing the effectiveness of PRF insurance.

Figure 3.1: *Total Acres Enrolled in PRF by County & Year*

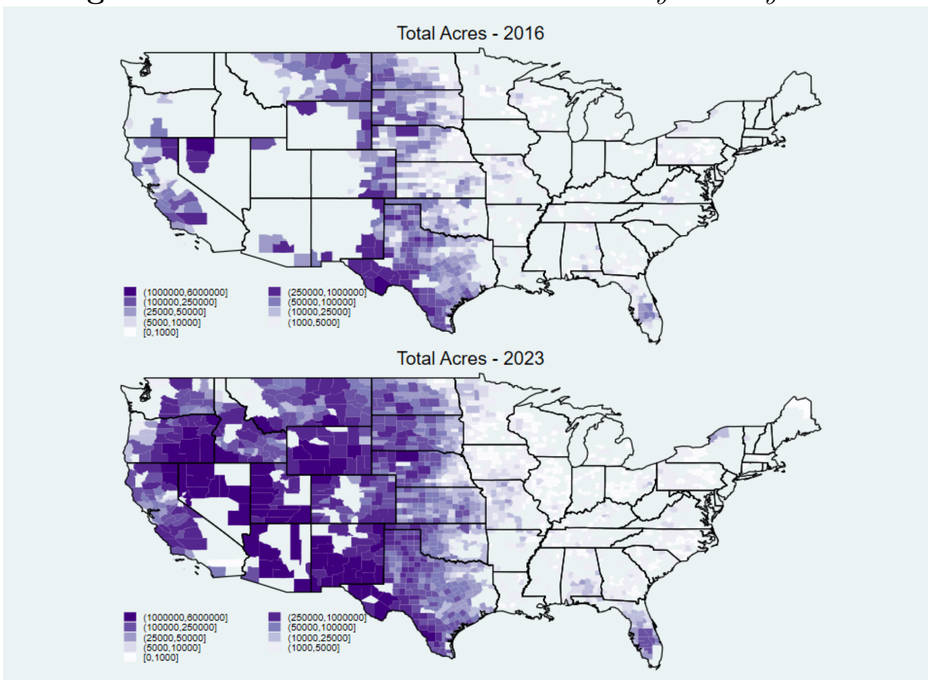


Figure 3.2: *Share of 2002 Pastureland Enrolled in PRF*

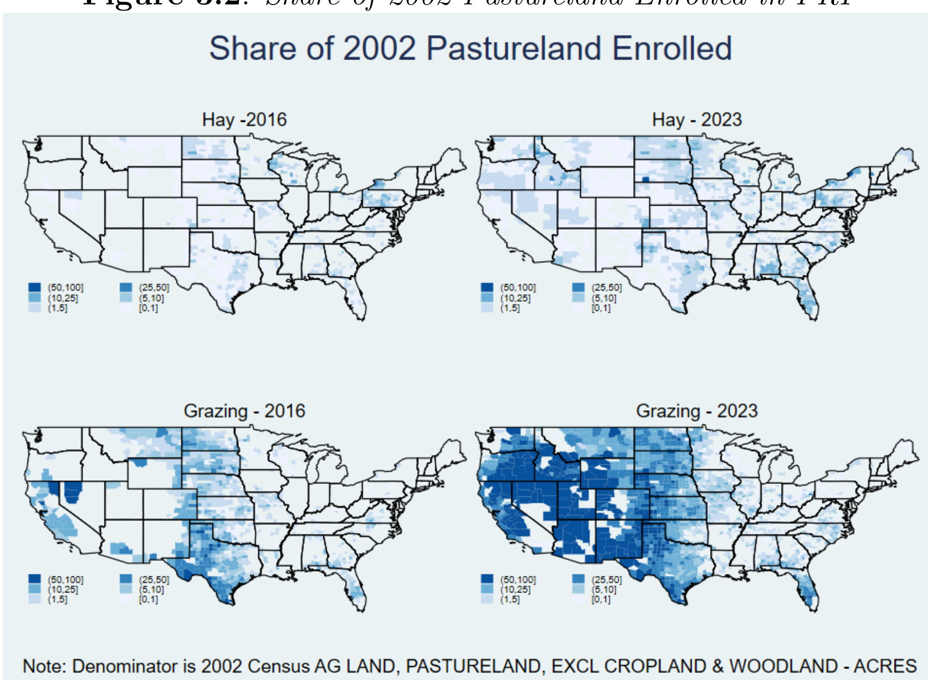


Table 3.1: *Summary Statistics*

	Hay	Grazing	Overall
Avg. Acres Enrolled	2,177.31	81,863.27	42,923.40
	(34.70)	(2538.21)	(1319.53)
Total Premium ($)	94,430.11	295,276.94	197,129.80
	(2163.23)	(8161.18)	(4346.27)
Total Indemnity ($)	87,159.20	294,367.94	193,111.94
	(2832.39)	(9610.27)	(5142.42)
LFP Payment ($)	581,635.63	604,663.13	593,410.38
	(15683.04)	(15318.28)	(10958.37)
LFP Payment Binary	0.46	0.48	0.47
	(0.00)	(0.00)	(0.00)
Precipication (cm)	93.37	89.70	91.49
	(0.36)	(0.36)	(0.25)
Avg. Growing Season Temp.	22.12	22.21	22.17
	(0.03)	(0.03)	(0.02)
Growing Degree Days (8-29 C)	2,923.82	2,951.18	2,937.81
	(8.84)	(8.52)	(6.13)
Extreme Degree Days (>29 C)	97.12	102.30	99.77
	(0.82)	(0.80)	(0.57)
Share of Enrolled Acres in Growing Season	0.45	0.41	0.43
	(0.00)	(0.00)	(0.00)
N	13733	14370	28103

Table 3.2: *Share of Acres Enrolled (2010-2023)*

Dependent Variable:	Logit Transformation of Share Acres Enrolled					
Variables / Pratice		Hay			Grazing	
Precipitation (cm)=L	-0.003***			-0.003***		
	(0.000)			(0.000)		
Avg. Growing Season Temp.=L	0.039***			0.046***		
	(0.010)			(0.010)		
Indemnity Ratio=L		0.099***	0.102***		0.145***	0.150***
		(0.015)	(0.015)		(0.015)	(0.015)
IHS of LFP Payment ($)=L		0.012***			0.018***	
		(0.002)			(0.001)	
LFP Payment Binary=L			0.138***			0.207***
			(0.022)			(0.019)
Constant	-5.969***	-5.581***	-5.574***	-3.919***	-3.413***	-3.405***
	(0.253)	(0.065)	(0.065)	(0.236)	(0.067)	(0.067)
Observations	11,498	11,498	11,498	11,913	11,913	11,913
R-squared	0.172	0.177	0.176	0.231	0.249	0.246
Number of fips	1,658	1,658	1,658	1,624	1,624	1,624
Fips FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Robust standard errors in parentheses. =L indicates the variable is lagged one year. Significance levels denoted as: *** p<0.01, ** p<0.05, * p<0.1

Figure 3.3: *Share of Enrolled Acres in Growing Season*

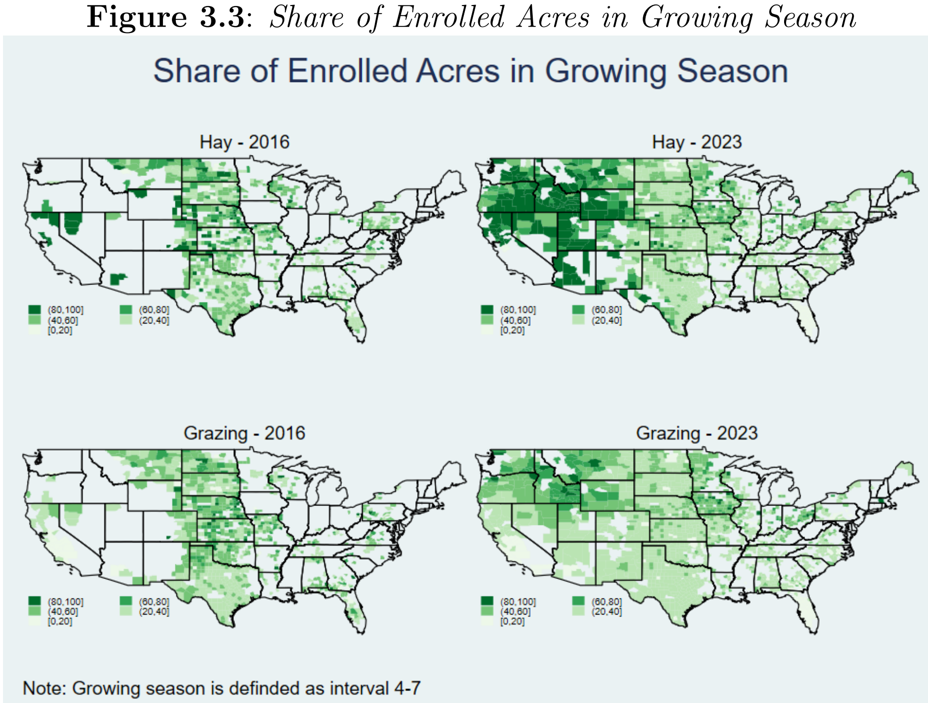


Table 3.3: *Share of Acres Enrolled (2016-2023)*

Dependent Variable:	Logit Transformation of Share Acres Enrolled					
Variables / Pratice	Hay			Grazing		
Precipitation (cm)=L	-0.005***			-0.004***		
	(0.001)			(0.000)		
Avg. Growing Season Temp.=L	0.046***			0.042***		
	(0.009)			(0.009)		
Indemnity Ratio=L		0.126***	0.134***		0.158***	0.169***
		(0.015)	(0.015)		(0.015)	(0.015)
IHS of LFP Payment ($)=L		0.017***			0.022***	
		(0.002)			(0.002)	
LFP Payment Binary=L			0.183***			0.246***
			(0.024)			(0.020)
Constant	-5.629***	-5.238***	-5.233***	-3.709***	-3.375***	-3.372***
	(0.234)	(0.030)	(0.031)	(0.229)	(0.027)	(0.027)
Observations	8,188	8,188	8,188	8,326	8,326	8,326
R-squared	0.171	0.187	0.182	0.259	0.297	0.290
Number of fips	1,622	1,622	1,622	1,588	1,588	1,588
Fips FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Robust standard errors in parentheses. =L indicates the variable is lagged one year. Significance levels denoted as: *** p<0.01, ** p<0.05, * p<0.1

Table 3.4: *Share of Enrolled in Acres in Growing Season (2010-2023)*

Dependent Variable:	Logit Transformation of Share of Enrolled Acres in Growing Season					
Variables / Pratice		Hay			Grazing	
Growing Season Precip. (cm)=L	-0.003***			-0.001**		
	(0.001)			(0.001)		
Avg. Growing Season Temp.=L	-0.036***			-0.010		
	(0.010)			(0.007)		
Growing Season Indemnity Ratio=L		0.012	0.012		0.006	0.006
		(0.008)	(0.008)		(0.006)	(0.006)
IHS of LFP Payment ($)=L		0.004***			0.005***	
		(0.001)			(0.001)	
LFP Payment Binary=L			0.062***			0.064***
			(0.019)			(0.014)
Constant	0.796***	-0.107***	-0.113***	0.044	-0.236***	-0.238***
	(0.240)	(0.041)	(0.041)	(0.177)	(0.031)	(0.031)
Observations	11,050	10,977	10,977	11,998	11,936	11,936
R-squared	0.061	0.061	0.062	0.133	0.138	0.138
Number of fips	1,615	1,610	1,610	1,648	1,644	1,644
Fips FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Robust standard errors in parentheses. =L indicates the variable is lagged one year. Growing season defined as intervals 4-7. Significance levels denoted as: *** p<0.01, ** p<0.05, * p<0.1

Table 3.5: *Share of Enrolled Acres in Growing Season (2016-2023)*

Dependent Variable:	Logit Transformation of Share of Enrolled Acres in Growing Season					
Variables / Pratice	Hay			Grazing		
Growing Season Precip. (cm)=L	-0.000			0.000		
	(0.001)			(0.001)		
Avg. Growing Season Temp.=L	0.004			0.001		
	(0.009)			(0.006)		
Growing Season Indemnity Ratio=L		0.002	0.001		0.003	0.002
		(0.008)	(0.008)		(0.006)	(0.006)
IHS of LFP Payment ($)=L		-0.002			-0.000	
		(0.001)			(0.001)	
LFP Payment Binary=L			-0.007			0.004
			(0.019)			(0.014)
Constant	-0.349	-0.248***	-0.252***	-0.432***	-0.406***	-0.409***
	(0.218)	(0.021)	(0.021)	(0.146)	(0.015)	(0.015)
Observations	7,866	7,816	7,816	8,473	8,433	8,433
R-squared	0.017	0.018	0.017	0.029	0.032	0.032
Number of fips	1,583	1,578	1,578	1,613	1,609	1,609
Fips FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Robust standard errors in parentheses. =L indicates the variable is lagged one year. Growing season defined as intervals 4-7. Significance levels denoted as: *** p<0.01, ** p<0.05, * p<0.1

Table 3.6: *Alternative Base Year - 2007*

Dependent Variable	Logit Transformation of Share Acres Enrolled					
Variables / Pratice		Hay			Grazing	
Precipitation (cm)=L	-0.003***			-0.003***		
	(0.000)			(0.000)		
Avg. Growing Season Temp.=L	0.038***			0.043***		
	(0.010)			(0.010)		
Indemnity Ratio=L		0.099***	0.102***		0.149***	0.154***
		(0.015)	(0.015)		(0.015)	(0.015)
IHS of LFP Payment ($)=L		0.012***			0.018***	
		(0.002)			(0.001)	
LFP Payment Binary=L			0.137***			0.207***
			(0.023)			(0.019)
Constant	-5.933***	-5.547***	-5.540***	-3.829***	-3.387***	-3.380***
	(0.254)	(0.065)	(0.065)	(0.243)	(0.065)	(0.065)
Observations	11,499	11,499	11,499	11,901	11,901	11,901
R-squared	0.172	0.178	0.176	0.228	0.247	0.244
Number of fips	1,658	1,658	1,658	1,623	1,623	1,623
Fips FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Robust standard errors in parentheses. =L indicates the variable is lagged one year. Significance levels denoted as: *** p<0.01, ** p<0.05, * p<0.1

Table 3.7: *Share of Acres Enrolled (2010-2023) - Log Transformation*

Dependent Variable:	Log Transformation of Share Acres Enrolled					
Variables / Pratice	Hay			Grazing		
Precipitation (cm)=L	-0.003***			-0.003***		
	(0.000)			(0.000)		
Avg. Growing Season Temp.=L	0.038***			0.027***		
	(0.010)			(0.008)		
Indemnity Ratio=L		0.097***	0.100***		0.125***	0.129***
		(0.014)	(0.014)		(0.013)	(0.013)
IHS of LFP Payment ($)=L		0.012***			0.014***	
		(0.002)			(0.001)	
LFP Payment Binary=L			0.138***			0.164***
			(0.022)			(0.017)
Constant	-5.981***	-5.594***	-5.586***	-3.579***	-3.444***	-3.438***
	(0.250)	(0.065)	(0.065)	(0.192)	(0.061)	(0.061)
Observations	11,505	11,505	11,505	12,218	12,218	12,218
R-squared	0.172	0.177	0.175	0.209	0.223	0.221
Number of fips	1,659	1,659	1,659	1,670	1,670	1,670
Fips FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Robust standard errors in parentheses. =L indicates the variable is lagged one year. Significance levels denoted as: *** p<0.01, ** p<0.05, * p<0.1

Table 3.8: *Share of Enrolled in Acres in Growing Season (2010-2023) - Log Transformation*

Dependent Variable:	Log Transformation of Share of Enrolled Acres in Growing Season					
Variables / Pratice	Hay			Grazing		
Growing Season Precip. (cm)=L	-0.002***			-0.001**		
	(0.000)			(0.000)		
Avg. Growing Season Temp.=L	-0.025***			-0.007		
	(0.006)			(0.005)		
Growing Season Indemnity Ratio=L		0.005	0.005		0.003	0.003
		(0.004)	(0.004)		(0.003)	(0.003)
IHS of LFP Payment ($)=L		0.003***			0.003***	
		(0.001)			(0.001)	
LFP Payment Binary=L			0.041***			0.045***
			(0.012)			(0.010)
Constant	-0.195	-0.834***	-0.837***	-0.697***	-0.890***	-0.892***
	(0.138)	(0.021)	(0.021)	(0.128)	(0.016)	(0.016)
Observations	11,372	11,298	11,298	12,057	11,995	11,995
R-squared	0.053	0.053	0.053	0.100	0.107	0.107
Number of fips	1,644	1,639	1,639	1,651	1,647	1,647
Fips FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Robust standard errors in parentheses. =L indicates the variable is lagged one year. Growing season defined as intervals 4-7. Significance levels denoted as: *** p<0.01, ** p<0.05, * p<0.1

Table 3.9: *Share of Enrolled in Acres in Growing Season (2010-2023)*

Dependent Variable:	Logit Transformation of Share of Enrolled Acres in Long Growing Season					
Variables / Pratice	Hay			Grazing		
Growing Season Precip. (cm)=L	-0.002**			-0.001***		
	(0.001)			(0.001)		
Avg. Growing Season Temp.=L	-0.020**			-0.002		
	(0.010)			(0.007)		
Indemnity Ratio=L		0.014	0.014		-0.011	-0.011
		(0.013)	(0.013)		(0.009)	(0.009)
IHS of LFP Payment ($)=L		0.004***			0.005***	
		(0.001)			(0.001)	
LFP Payment Binary=L			0.057***			0.065***
			(0.016)			(0.014)
Constant	1.129***	0.599***	0.592***	0.669***	0.553***	0.549***
	(0.227)	(0.049)	(0.049)	(0.167)	(0.039)	(0.039)
Observations	10,804	10,804	10,804	11,897	11,897	11,897
R-squared	0.052	0.053	0.053	0.131	0.132	0.132
Number of fips	1,611	1,611	1,611	1,642	1,642	1,642
Fips FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Robust standard errors in parentheses. =L indicates the variable is lagged one year. Growing season defined as intervals 3-8. Significance levels denoted as: *** p<0.01, ** p<0.05, * p<0.1

Table 3.10: *Combined Hay & Grazing Alternative Specification*

Dependent Variable:	Logit Transformation of Share Acres Enrolled				
Variables	(1)	(2)	(3)	(4)	(5)
Precipitation (cm)=L	-0.003***				
	(0.001)				
Avg. Growing Season Temp.=L	0.064***				
	(0.019)				
Indemnity Ratio (Value)=L		0.161***	0.168***		
		(0.020)	(0.020)		
Indemnity Ratio (Count)=L				0.226***	0.234***
				(0.040)	(0.039)
IHS LFP transformation ($)=L		0.020***		0.021***	
		(0.002)		(0.002)	
LFP Payment Binary=L			0.216***		0.230***
			(0.030)		(0.030)
Constant	-3.721***	-2.812***	-2.798***	-2.888***	-2.875***
	(0.439)	(0.064)	(0.064)	(0.068)	(0.069)
Observations	12,247	12,240	12,240	12,240	12,240
R-squared	0.157	0.168	0.166	0.163	0.161
Number of fips	1,650	1,650	1,650	1,650	1,650
Fips FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes

Notes: Robust standard errors in parentheses. =L indicates the variable is lagged one year. Significance levels denoted as: *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Bibliography

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

Aihounton, G.B.D., and A. Henningsen. 2021. "Units of measurement and the inverse hyperbolic sine transformation." *The Econometrics Journal* 24:334–351.

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.

Ashraf, N., B.K. Jack, and E. Kamenica. 2013. "Information and subsidies: Complements or substitutes?" *Journal of Economic Behavior & Organization* 88:133–139.

Atreya, A., S. Ferreira, and E. Michel-Kerjan. 2015. "What drives households to buy flood insurance? New evidence from Georgia." *Ecological Economics* 117:153–161.

Berry, J., G. Fischer, and R. Guiteras. 2020. "Eliciting and Utilizing Willingness to Pay: Evidence from Field Trials in Northern Ghana." *Journal of Political Economy* 128:1436–1473, Publisher: The University of Chicago Press.

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.

Breidert, C., M. Hahsler, and T. Reutterer. 2006. "A Review of Methods for Measuring Willingness-to-Pay." *Innovative Marketing* 2.

Burbidge, J.B., L. Magee, and A.L. Robb. 1988. "Alternative Transformations to Handle Extreme Values of the Dependent Variable.", pp. .

Carlsson, F., D. Daruvala, and H. Jaldell. 2010. "Do you do what you say or do you do what you say others do?" *Journal of Choice Modelling* 3:113–133.

Carlsson, F., M. Kataria, and E. Lampi. 2018. "Demand effects in stated preference surveys." *Journal of Environmental Economics and Management* 90:294–302.

Centre, S.E.a.P. 2023. "Cost of Cultivation of Agricultural Crops." Working paper No. Volume 84, Department of Agriculture, Peradeniya, Dec.

Champ, P.A., R. Moore, and R.C. Bishop. 2009. "A Comparison of Approaches to Mitigate Hypothetical Bias." *Agricultural and Resource Economics Review* 38:166–180.

Channa, H., A.Z. Chen, P. Pina, J. Ricker-Gilbert, and D. Stein. 2019. "What drives smallholder farmers' willingness to pay for a new farm technology? Evidence from an experimental auction in Kenya." *Food Policy* 85:64–71.

Che, Y., H. Feng, and D.A. Hennessy. 2020. "Recency effects and participation at the extensive and intensive margins in the U.S. Federal Crop Insurance Program." *The Geneva Papers on Risk and Insurance - Issues and Practice* 45:52–85.

Cho, W., and B.W. Brorsen. 2020. "Design of the Rainfall Index Crop Insurance Program for Pasture, Rangeland, and Forage.", pp. . Publisher: Unknown.

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.

Cohen, J., and P. Dupas. 2010. "Free Distribution or Cost-Sharing? Evidence from a Randomized Malaria Prevention Experiment*." *The Quarterly Journal of Economics* 125:1–45.

Cole, S., D. Stein, and J. Tobacman. 2014. "Dynamics of Demand for Index Insurance: Evidence from a Long-Run Field Experiment." *American Economic Review* 104:284–290.

Collier, B., J. Skees, and B. Barnett. 2009. "Weather Index Insurance and Climate Change: Opportunities and Challenges in Lower Income Countries." *The Geneva Papers on Risk and Insurance - Issues and Practice* 34:401–424.

Colombo, S., K. Glenk, and B. Rocamora-Montiel. 2016. "Analysis of choice inconsistencies in on-line choice experiments: impact on welfare measures." *European Review of Agricultural Economics* 43:271–302.

Connor, L., and A.L. Katchova. 2020. "Crop Insurance Participation Rates and Asymmetric Effects on U.S. Corn and Soybean Yield Risk.", pp. . Publisher: Unknown.

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

Cummings, R.G., and L.O. Taylor. 1999. "Unbiased Value Estimates for Environmental Goods: A Cheap Talk Design for the Contingent Valuation Method." *American Economic Review* 89:649–665.

Davidson, K.A., and B.K. Goodrich. 2023. "Nudge to insure: Can informational nudges change enrollment decisions in pasture, rangeland, and forage rainfall index insurance?" *Applied Economic Perspectives and Policy* 45:534–554.

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.

Demnitz, R., and S. Joslyn. 2020. "The Effects of Recency and Numerical Uncertainty Estimates on Overcautiousness.", Apr., pp. . Section: Weather, Climate, and Society.

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. .

Entem, A., P. Lloyd-Smith, W..V.L. Adamowicz, and P.C. Boxall. 2019. "Using inferred valuation to quantify survey and social desirability bias in stated preference research." *American Journal of Agricultural Economics* 104:1224–1242, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ajae.12268.

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.

Fezzi, C., L. Menapace, and R. Raffaelli. 2021. "Estimating risk preferences integrating insurance choices with subjective beliefs." *European Economic Review* 135:103717.

Fischer, K., J. van den Berg, and C. Mutengwa. 2015. "Is Bt maize effective in improving South African smallholder agriculture?" *South African Journal of Science* 111:1–2, Num Pages: 2 Place: Pretoria, South Africa Publisher: Academy of Science of South Africa Section: Commentary.

FSA. 2020. "Payment Files Information." Journal Abbreviation: Payment Files Information Last Modified: 2023-08-28T09:56:42.

Fuller, A.J., and J. Ricker-Gilbert. 2021. "Estimating Demand for Third-party Quality Testing in Rural Grain Markets: Evidence from an Experimental Auction for Measuring Moisture Content in Kenya." *Journal of African Economies* 30:389–417.

Gallagher, J. 2014. "Learning about an Infrequent Event: Evidence from Flood Insurance Take-Up in the United States." *American Economic Journal: Applied Economics* 6:206–233, Publisher: American Economic Association.

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.

Goodrich, B., J. Yu, and M. Vandeveer. 2020. "Participation patterns of the rainfall index insurance for pasture, rangeland and forage programme." *The Geneva Papers on Risk and Insurance - Issues and Practice* 45:29–51.

Goodrich, B.K., and K.A. Davidson. 2024. "Enrollment in Pasture, Rangeland, and Forage Rainfall Index Insurance: Awareness Matters." *Journal of Agricultural and Resource Economics* Preprint, Publisher: Unknown.

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. .

Hrozencik, R.A. 2024. "The Stocking Impact and Financial-Climate Risk of the Livestock Forage Disaster Program.", pp. .

Ifft, J., S. Wu, and T. Kuethe. 2014. "The Impact of Pasture Insurance on Farmland Values." *Agricultural and Resource Economics Review* 43:390–405, Num Pages: 16 Place: Durham, United Kingdom Publisher: Cambridge University Press.

Iizuka, T., and H. Shigeoka. 2022. "Is Zero a Special Price? Evidence from Child Health Care." *American Economic Journal: Applied Economics* 14:381–410.

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.

Johnson, N.L. 1949. "Systems of Frequency Curves Generated by Methods of Translation." *Biometrika* 36:149–176, Publisher: [Oxford University Press, Biometrika Trust].

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.

Kahneman, D., and S. Frederick. 2002. "Representativeness revisited: Attribute substitution in intuitive judgment." *Heuristics and biases: The psychology of intuitive judgment*, pp. .

Karlan, D., R. Osei, I. Osei-Akoto, and C. Udry. 2014. "Agricultural Decisions after Relaxing Credit and Risk Constraints *." *The Quarterly Journal of Economics* 129:597–652.

Keller, J.B., and T.L. Saitone. 2022. "Basis risk in the pasture, rangeland, and forage insurance program: Evidence from California." *American Journal of Agricultural Economics* 104:1203–1223, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ajae.12282.

Lee, S. 2024. "Aaron Smith - Download US Weather Data."

Leggett, C.G., N.S. Kleckner, K.J. Boyle, J.W. Duffield, and R.C. Mitchell. 2003. "Social Desirability Bias in Contingent Valuation Surveys Administered through In-Person Interviews." *Land Economics* 79:561–575, Publisher: [Board of Regents of the University of Wisconsin System, University of Wisconsin Press].

Levitt, S.D., and J.A. List. 2007. "What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World?" *Journal of Economic Perspectives* 21(2):153–174.

List, J.A., R.P. Berrens, A.K. Bohara, and J. Kerkvliet. 2004. "Examining the Role of Social Isolation on Stated Preferences." *American Economic Review* 94:741–752.

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.

Lusk, J.L., and F.B. Norwood. 2009a. "Bridging the gap between laboratory experiments and naturally occurring markets: An inferred valuation method." *Journal of Environmental Economics and Management* 58(2):236–250.

—. 2010. "Direct Versus Indirect Questioning: An Application to the Well-Being of Farm Animals." *Social Indicators Research* 96:551–565.

—. 2009b. "An Inferred Valuation Method." *Land Economics* 85:500–514, Publisher: [Board of Regents of the University of Wisconsin System, University of Wisconsin Press].

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.

Menapace, L., and R. Raffaelli. 2020. "Unraveling hypothetical bias in discrete choice experiments." *Journal of Economic Behavior & Organization* 176:416–430.

Miller, K.M., R. Hofstetter, H. Krohmer, and Z.J. Zhang. 2011. "How Should Consumers' Willingness to Pay be Measured? An Empirical Comparison of State-of-the-Art Approaches." *Journal of Marketing Research* 48:172–184, Publisher: SAGE Publications Inc.

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

Miranda, M.J., and K. Farrin. 2012. "Index Insurance for Developing Countries." *Applied Economic Perspectives and Policy* 34:391–427.

Moore, F.C., N. Obradovich, F. Lehner, and P. Baylis. 2019. "Rapidly declining remarkability of temperature anomalies may obscure public perception of climate change." *Proceedings of the National Academy of Sciences* 116:4905–4910, Publisher: Proceedings of the National Academy of Sciences.

Murphy, J.J., P.G. Allen, T.H. Stevens, and D. Weatherhead. 2005. "A Meta-Analysis of Hypothetical Bias in Stated Preference Valuation.", pp. .

Norwood, F.B., and J.L. Lusk. 2011. "Social Desirability Bias in Real, Hypothetical, and Inferred Valuation Experiments." *American Journal of Agricultural Economics* 93:528–534, Publisher: [Agricultural & Applied Economics Association, Oxford University Press].

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.

Papke, L.E., and J.M. Wooldridge. 1996. "Econometric methods for fractional response variables with an application to 401(k) plan participation rates." *Journal of Applied Econometrics* 11:619–632, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/%28SICI%291099-1255%28199611%2911%3A6%3C619%3A%3AAID-JAE418%3E3.0.CO%3B2-1.

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.

PRISM. 2024. "PRISM Climate Group at Oregon State University."

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.

RMA. 2024a. "Pasture, Rangeland, Forage Pilot Insurance Program.", Oct., pp. .

—. 2024b. "PRF Rainfall."

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.

Santeramo, F.G. 2019. "I Learn, You Learn, We Gain Experience in Crop Insurance Markets." *Applied Economic Perspectives and Policy* 41:284–304, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1093/aepp/ppy012.

Schmidt, J., and T.H.A. Bijmolt. 2020. "Accurately measuring willingness to pay for consumer goods: a meta-analysis of the hypothetical bias." *Journal of the Academy of Marketing Science* 48:499–518.

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.

Schwab, B., and J. Yu. 2024. "Guaranteed storage? Risk and credit constraints in the demand for postharvest technology." *Oxford Economic Papers*, Jul., pp. gpae027.

Shampanier, K., N. Mazar, and D. Ariely. 2007. "Zero as a Special Price: The True Value of Free Products." *Marketing Science* 26:742–757, Publisher: INFORMS.

Shrum, T.R., and W.R. Travis. 2022. "Experiments in ranching: Rain-index insurance and investment in production and drought risk management." *Applied Economic Perspectives and Policy* 44:1513–1533, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/aepp.13304.

Shukla, P., H.K. Pullabhotla, and K. Baylis. 2022. "Trouble with zero: The limits of subsidizing technology adoption." *Journal of Development Economics* 158:102920.

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.

Smart, A.J., K. Harmoney, J.D. Scasta, M.B. Stephenson, J.D. Volesky, L.T. Vermeire, J.C. Mosley, K. Sedivec, M. Meehan, T. Haigh, J.D. Derner, and M.P. McClaran. 2021. "Forum: Critical Decision Dates for Drought Management in Central and Northern Great Plains Rangelands." *Rangeland Ecology & Management* 78:191–200.

Stein, D. 2016. "Dynamics of Demand for Rainfall Index Insurance: Evidence from a Commercial Product in India." *The World Bank Economic Review*, Oct., pp. lhw045.

Sælensminde, K. 2002. "The Impact of Choice Inconsistencies in Stated Choice Studies." *Environmental and Resource Economics* 23:403–420.

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.

Turner, G., F. Said, and U. Afzal. 2014. "Microinsurance Demand After a Rare Flood Event: Evidence From a Field Experiment in Pakistan." *The Geneva Papers on Risk and Insurance. Issues and Practice* 39:201–223, Publisher: Palgrave Macmillan Journals.

Tversky, A., and D. Kahneman. 1973. "Availability: A heuristic for judging frequency and probability." *Cognitive Psychology* 5:207–232.

—. 1974. "Judgment under Uncertainty: Heuristics and Biases." *Science* 185:1124–1131, Publisher: American Association for the Advancement of Science.

USDA. 2002. "2002 – AgCensus."

Westerhold, A., C. Walters, K. Brooks, M. Vandeveer, J. Volesky, and W. Schacht. 2018. "Risk implications from the selection of rainfall index insurance intervals." *Agricultural Finance Review* 78:514–531.

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.

Yadav, L., T.M. Van Rensburg, and H. Kelley. 2013. "A Comparison Between the Conventional Stated Preference Technique and an Inferred Valuation Approach." *Journal of Agricultural Economics* 64:405–422.

Yu, J., B. Goodrich, and A. Graven. 2022. "Competing farm programs: Does the introduction of a risk management program reduce the enrollment in the Conservation Reserve Program?" *Journal of the Agricultural and Applied Economics Association* 1:320–333, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jaa2.29.

Yu, J., M. Vandeveer, J.D. Volesky, K. Harmoney, J. Yu, M. Vandeveer, J.D. Volesky, and K. Harmoney. 2019. "Estimating the Basis Risk of Rainfall Index Insurance for Pasture, Rangeland, and Forage." *Journal of Agricultural and Resource Economics* 44, Publisher: Unknown.

Zapata, S.D., and J.M. García. 2022. "Risk-Efficient Coverage Selection Strategies for the Pasture, Rangeland, Forage (PRF) Insurance Program." *Journal of Agricultural and Applied Economics* 54:286–305.

Zizzo, D.J. 2010. "Experimenter demand effects in economic experiments." *Experimental Economics* 13:75–98.

Appendix A

Appendix: Chapter 1

A.1 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 A.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 A.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

Appendix B

Appendix: Chapter 2

B.1 Internal Validity

Table B.1: *Internal Validity Check*

	Consistent	Inconsistent	p-value
Household size	4.072	4.010	0.528
Share with Male Head	0.505	0.494	0.403
Maximum Education	13.704	13.276	0.007
Number of Plots	3.629	2.877	0.000
Total Plot Area	5.028	7.266	0.003
Hired Male labor	5.770	10.281	0.000
Hired Female labor	0.460	0.158	0.006
HH male labor	1.165	0.557	0.013
HH female labor	0.393	0.063	0.002
Harvest (kg)	1655.168	4601.477	0.000
Largest plot	3.035	5.017	0.001
N	723	312	

B.2 Power Calculations

Prior to survey implementation, we conducted power calculations to determine the appropriate number of questions and prices to include. The RCT treatment group had a total of 1,100 participants, making that our maximum potential sample size. Given the existing survey length and the complexity of the questions, the maximum number of list questions to be included was set at three per person. We then used the number of questions, the maximum potential sample, and the number of prices to calculate the sample size per price level for each scenario ($Sample_p = \frac{1100*Questions}{Prices}$). Table B.2 summarizes the scenarios considered.

Next, we needed to identify how often we could detect (1) the share WTP different from zero at standard significance levels and (2) a statistically significant difference between the direct and indirect questioning. The first objective assumes a 20% true use rate of SEED Hub, which is drawn from the midline survey. Prevalence of the non-sensitive items in the list is assumed based on the mid-line survey and publicly available data sets. The level of bias in WTP is allowed to vary from 10% to 30%, based on social science literature levels of social desirability bias.

Based on the assumptions outlined, a data set was generated, objectives 1 and 2 were tested, and the p-value of the test was saved. The process was iterated 1,000 times. Finally, the share of the sample for each objective with $p < 0.05$ is reported in Table B.3. Detect refers to the ability to detect social desirability bias and Difference refers to the ability to distinguish the indirect from the direct measure statistically. The final selection of three questions with four prices was selected based on the available results. Future applications of this process should include an additional variable to account for the possibility of inconsistent responses, which may further decrease sample size though the attrition need not be related to the price level.

Table B.2: *Power Calculation Key*

Sample Size per price level	List Experiments	Prices
733	2	3
550	2	4
440	2	5
1100	3	3
825	3	4
660	3	5
550	3	6
471	3	7
413	3	8

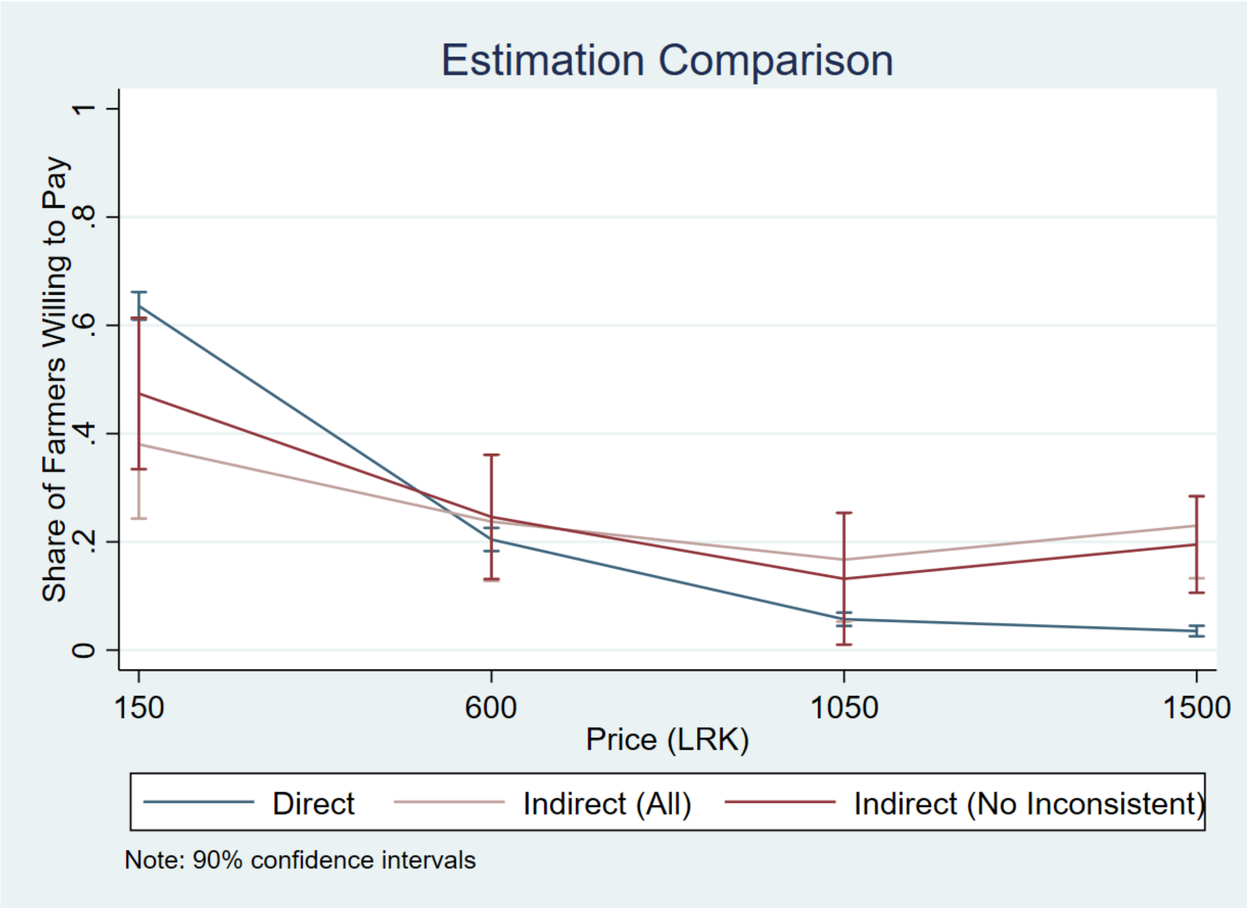
Table B.3: *Power Calculation Results*

Bias	Sample Size	Detect	Difference
0.1	367	70.80	17.50
0.15	367	71.40	32.10
0.2	367	72.20	52.00
0.25	367	74.70	71.60
0.3	367	71.70	86.90
0.1	412	75.20	19.30
0.15	412	76.10	36.40
0.2	412	77.50	54.80
0.25	412	74.60	76.10
0.3	412	77.20	86.90
0.1	440	81.30	17.40
0.15	440	78.80	37.70
0.2	440	79.70	60.40
0.25	440	79.40	78.50
0.3	440	80.30	90.70
0.1	471	81.70	20.60
0.15	471	81.60	41.40
0.2	471	82.10	62.70
0.25	471	79.60	83.20
0.3	471	82.70	93.40
0.1	550	87.30	23.40
0.15	550	87.90	43.00
0.2	550	86.10	69.20
0.25	550	86.50	86.80
0.3	550	85.40	97.10
0.1	550	87.10	23.50
0.15	550	88.40	44.10
0.2	550	86.70	69.20
0.25	550	87.30	87.00
0.3	550	86.30	95.30
0.1	660	92.70	27.60
0.15	660	91.50	52.20
0.2	660	92.60	76.20
0.25	660	92.70	90.30
0.3	660	93.20	98.20
0.1	733	95.10	25.20
0.15	733	96.30	57.40
0.2	733	94.40	80.30
0.25	733	94.00	92.80
0.3	733	94.40	99.30
0.1	825	96.60	32.70
0.15	825	96.60	61.40
0.2	825	96.20	86.30
0.25	825	96.00	96.50
0.3	825	97.90	99.90
0.1	1100	99.40	43.70
0.15	1100	99.40	75.90
0.2	1100	99.20	94.50
0.25	1100	99.80	98.90
0.3	1100	99.60	100.00

Detect is how often we could detect a statistically significant effect, given a 20% true prevalence of SEED hub usage. Difference is how often we could detect a statistically significant difference between list and direct questioning.

B.3 Inconsistent Response

Figure B.1: *With and Without Inconsistent Responses*



Appendix C

Appendix: Chapter 3

C.1 Transformation Comparison

Figure C.1: *IHS Transformation Unit Comparison*

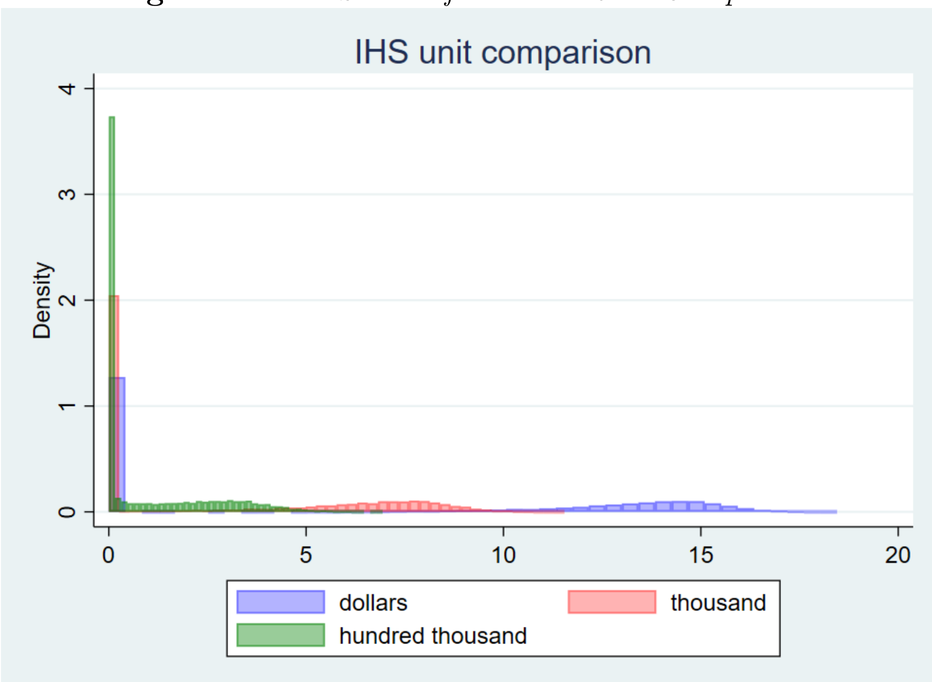


Figure C.2: *Dependent Variable Transformation Comparison*