IFPRI Discussion Paper 02393

December 2025

Targeting of Food Aid Programs

Evidence from Egypt

Mai Mahmoud

Sikandra Kurdi

Development Strategies and Governance Unit

INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE

The International Food Policy Research Institute (IFPRI), a CGIAR Research Center established in 1975, provides research-based policy solutions to sustainably reduce poverty and end hunger and malnutrition. IFPRI's strategic research aims to foster a climate-resilient and sustainable food supply; promote healthy diets and nutrition for all; build inclusive and efficient markets, trade systems, and food industries; transform agricultural and rural economies; and strengthen institutions and governance. Gender is integrated in all the Institute's work. Partnerships, communications, capacity strengthening, and data and knowledge management are essential components to translate IFPRI's research from action to impact. The Institute's regional and country programs play a critical role in responding to demand for food policy research and in delivering holistic support for country-led development. IFPRI collaborates with partners around the world.

AUTHORS

Mai Mahmoud (mmahmoud@povertyactionlab.org) is a Postdoctoral Research Fellow at the Abdul Lateef Jamil Poverty Action Lab (J-PAL) Middle East and North Africa, Cairo, Egypt.

Sikandra Kurdi (S.Kurdi@cgiar.org) is a Research Fellow in the Development Strategies and Governance (DSG) Unit and Egypt Country Program Leader at the International Food Policy Research Institute (IFPRI), Cairo, Egypt.

Abstract

In-kind food aid programs remain prominent world-wide. Targeting in these programs is complex due to potential distortions in consumption. This paper advances the literature by moving beyond poverty-based targeting to address nutritional objectives. Using data from a randomized controlled trial (RCT), we apply machine learning (ML) techniques to analyze heterogeneity in impacts across nutritional outcomes, aiming to inform targeting based on observable characteristics. We find that such characteristics significantly predict heterogeneity in treatment effects, though relevant predictors differ by outcome and treatment type. Building on recent literature advocating for balancing of deprivation and expected impact, we show that, in our context, the trade-off between targeting the most impacted versus the most deprived households is limited. Instead, the main challenge is prioritizing among competing nutritional objectives. Our findings indicate that ML methods can inform outcome-specific targeting criteria, though these criteria vary across outcomes and are imperfectly correlated.

Keywords: Social assistance, targeting, nutrition, machine learning (ML), proxy means testing (PMT), Middle East and North Africa, Egypt

Acknowledgments

We gratefully acknowledge the financial support of Sawaris Foundation for Social Development and strong partnership from the Egyptian Food Bank for the experimental study described in this paper. The dietary analysis and conversion of diets into macro and micro-nutrients was conducted by a team of researchers from Egypt's National Research Center (NRC). This work was undertaken as part of the CGIAR Science Program on Policy Innovations. The Policy Innovations Science Program delivers evidence-based recommendations to strengthen policies, markets and institutions, through leveraging expertise in analyzing policy trade-offs and co-creating solutions in partnership with policymakers, think tanks, and other stakeholders in the Global South. We would like to thank all funders who supported this research through their contributions to the CGIAR Trust Fund: https://www.cgiar.org/funders/.

1 Introduction

In-kind food aid programs represent an integral part of social safety nets across both low- and high-income countries, with over 90% of low-income countries incorporating in-kind transfers into their social protection systems (Gentilini, Honorati, and Yemtsov, 2014). Yet despite their ubiquity, the effectiveness of food aid programs is often questioned, especially in settings where in-kind transfers provide goods that are extramarginal to typical household consumption. In such contexts, food aid has the potential to shift consumption patterns (Cunha, 2014). For instance, food transfers offering diverse food items can improve dietary quality for households that might otherwise under-invest in nutritious diets (Hidrobo et al., 2014). However, in-kind transfers—particularly those that include higher-value food items—are costly, prompting policymakers to make strategic targeting decisions.

Targeting is already contentious in non-food social protection programs that rely on poverty indicators (Jayne et al., 2002; Alatas et al., 2012; Stoeffler, Mills, and Del Ninno, 2016; Brown, Ravallion, and Van de Walle, 2018; Bah et al., 2019). For food transfers, these challenges are even more nuanced. Nutritional gains from in-kind aid depend not only on baseline diet adequacy but also on the specific composition of the transfer, as well as on household substitution patterns and behavioral responses. As a result, targeting to improve nutritional outcomes may differ from targeting based on deprivation. Understanding how program impacts vary across households and across different types of food transfers is therefore essential for informing how food aid programs can best meet their nutritional objectives.

In this paper, we examine how data-driven approaches can inform the targeting of food aid by leveraging machine learning (ML) methods to predict household-level heterogeneity in treatment effects across nutritional outcomes. We analyze a nationwide in-kind food aid program in Egypt, a country characterized by the double burden of malnutrition—high rates of both obesity and childhood stunting (Ecker et al., 2016; UNICEF, WHO, and the World Bank Group, 2021). Using data from a randomized controlled trial (RCT) involving 150 communities, we study two types of food aid interventions: (1) a staple-heavy food box designed to ensure basic caloric sufficiency, and (2) a more expensive nutrition-sensitive food box designed to both ensure caloric sufficiency and improve dietary diversity. Each community was randomly assigned to receive one of the two food box types or serve as a control. Within communities, all beneficiaries were pre-selected through community-based targeting, reflecting a prevailing approach to targeting in low- and middle-income country settings. We measure nutritional outcomes using three standardized indicators: (1) calorie intake, (2) food insecurity experience scale (FIES), and (3) household dietary diversity score (HDDS).

Our analysis is guided by a simple conceptual framework that highlights two key tradeoffs in the targeting of food aid programs. The first tradeoff concerns whether to prioritize households based on their baseline need or on the predicted nutritional impact of food aid. The second tradeoff involves balancing competing nutritional outcomes—such as calorie intake versus perceived food security—when designing targeting criteria. Recognizing these tradeoffs clarifies how alternative targeting rules may shape the effectiveness of food aid interventions.

A critical distinction in the provision of food aid is whether transfers are inframarginal or extramarginal to household consumption patterns. Inframarginal transfers do not change overall consumption relative to equivalent cash, while extramarginal transfers may result in increased consumption of the provided food items. For staple-heavy food aid, effective targeting should aim to identify households for whom the transfer addresses calorie insufficiency without resulting in overconsumption of staple calories. This scenario is more likely when the staple calories provided are inframarginal for households starting from low baseline calorie status. In contrast, for nutrition-sensitive food aid, targeting should prioritize households for whom the food items provided are extramarginal or who demonstrate high income elasticity of demand for diverse diets. However, low baseline dietary diversity alone may not reliably predict large impacts in the absence of information about a household's likely response to the in-kind transfer. Hence, a successful targeting strategy should ideally match food aid to the consumption needs and responsiveness of recipient households.

Perceived food security can serve as an alternative targeting criterion, yet it is a subjective measure that may be orthogonal to both calorie consumption and dietary diversity. While households may report improvements in perceived food security—as reflected in a lower FIES score— these perceptions may not correspond to genuine improvements in diet quality or may even mask calorie overconsumption. Moreover, if improving perceived food security is the primary policy objective, it may be addressed with either staple-heavy or nutrition-sensitive transfers. However, since nutrition-sensitive options are typically more costly, this choice involves important budgetary and strategic considerations for policymakers. Thus, the choice of targeting criterion not only shapes who benefits, but can also have implications on the design of food aid programs.

From a feasibility perspective, it is also important to assess whether meaningful observable proxies exist for targeting households with high baseline need or high nutritional impacts. Directly measuring nutritional outcomes across a large population for targeting purposes is impractical and prone to manipulation, especially in the absence of objective verification. Because average calorie intake, food insecurity experience, and dietary diversity are not externally verifiable, they are not suitable as direct targeting criteria. Similarly, estimating households' underlying demand for calories or diverse diets to predict their response to in-kind transfers is not feasible. The survey data analyzed in this paper allows us to explore whether observable characteristics, such as household demographics, education, employment status, and asset ownership, can serve as reliable proxies for both baseline deprivation and predicted nutritional impact.

Empirically, our results reinforce the findings of Abay et al. (2023) by demonstrating that both types of food transfers improve calorie intake, food security experience, and dietary diversity on average. Nevertheless, the magnitude and distribution of impacts vary markedly across recipient households. This variation highlights the necessity of examining not only the average effects but also the underlying heterogeneity in program outcomes. In this sense, our paper complements Abay et al. (2023), which focuses on average treatment effects, by documenting systematic heterogeneity in household-level responses and exploring its implications for targeting food aid programs.[1]

Using ML techniques, we demonstrate that heterogeneity in treatment effects can be effectively predicted using observable household characteristics, akin to those used in proxy means tests. Notably, subgroups with the highest predicted impacts are expected to experience nutritional gains more than double the average treatment effect—a pattern consistent across all nutritional outcomes and both food aid types. In particular, the nutritious food box consistently yields broader and more substantial improvements in dietary diversity and food security experience (as measured by FIES), whereas the staple-heavy box's benefits tend to be concentrated among a subgroup of households. These patterns highlight the value of using predictive analytics to improve targeting across diverse beneficiaries.

When evaluating potential synergies and trade-offs in targeting strategies, our analysis shows that for calorie intake and dietary diversity, predicted impacts are strongly aligned with baseline deprivation for both food aid types. This finding suggests that targeting the most deprived groups also captures those with the greatest potential for impact along these dimensions. In contrast, for food insecurity experience (FIES)—particularly in the case of the staple box—this alignment breaks down and trade-offs emerge. In such cases, households with the highest predicted impacts are not necessarily the most deprived, emphasizing the importance of tailoring targeting strategies to specific program objectives.

[1]The analysis approach, including the use of machine learning methods to explore heterogeneity and implications for targeting, was broadly outlined in the study's pre-analysis plan. However, the detailed implementation and focus presented here extend beyond the initial description.

Further, we explore trade-offs between the two food box types. We find that for calorie intake and dietary diversity, predicted treatment effects of the staple-heavy and nutrition-sensitive boxes are positively correlated. However, there is low correlation in treatment effects between the box types for food insecurity experience. This divergence suggests that different targeting approaches are required for staple-heavy and nutrition-sensitive interventions when program goals relate to reducing food insecurity experience.

Our analysis also shows that the characteristics predicting high treatment impact differ across food aid types and nutritional outcomes. For instance, unemployment-related vulnerability is a key predictor of impacts from the staple box on both calorie intake and food security experience. For the nutrition-sensitive box, variables associated with low socioeconomic status predict stronger impacts on calorie intake, while variables associated with higher socioeconomic status predict greater improvements in food security experience. In the case of dietary diversity, higher impacts are predicted by characteristics such as a more educated household head, smaller household size, and lack of a ration card. These findings indicate that policymakers may need to prioritize specific outcomes when setting targeting objectives, as the same households may not benefit equally across all dimensions.

Overall, our findings indicate that data-driven targeting strategies, informed by both empirical evidence and a conceptual understanding of household behavior, can enhance the effectiveness of food aid programs. By identifying which households are most likely to benefit from different types of food aid, policymakers can tailor program design and targeting criteria to specific nutritional goals. This is particularly important in resource-constrained environments, where improving the effectiveness of aid delivery can have substantial welfare implications. At the same time, we acknowledge that targeting based solely on observable characteristics has limitations. A subset of these characteristics may be subject to manipulation or distortion, especially when influenced by social pressures or strategic behavior in anticipation of program benefits, as recent evidence on signaling and social image suggests (Dupas, Fafchamps, and Hernandez-Nunez, 2024).

This paper contributes to several strands of literature. First, we add to research on the targeting of social protection programs in low- and middle-income countries.[2] Prior work has highlighted the strengths and limitations of proxy means tests and community-based targeting, but has largely focused on targeting for poverty or income.[3] We advance this literature by examining trade-offs between different targeting objectives aimed at improving nutritional outcomes. In doing so, we build on recent contributions by Haushofer et al. (2025), who advocate for balancing both deprivation and expected impact rather than targeting exclusively the poorest households. We extend this work by showing that, for the in-kind food aid program in our study, the trade-off between targeting the most impacted versus the most deprived households is relatively limited. Instead, a more salient trade-off emerges when prioritizing targeting based on competing nutritional outcomes or alternative intervention modalities.

Second, we contribute to the literature on nutrition-sensitive social protection (Ruel and Alderman, 2013; Alderman, 2014). Although many social safety net programs are motivated by nutritional objectives, few explicitly assess how targeting can be optimized to maximize nutritional gains. While Premand and Schnitzer (2021) demonstrates the viability of formula-based targeting for a cash transfer program using food insecurity scores, we show how machine learning can refine targeting for in-kind food aid programs. Our findings indicate that even among economically deprived populations, observable characteristics such as education, unemployment, and household composition predict meaningful heterogeneity in treatment effects across key nutritional outcomes.

Third, we contribute to an emerging literature applying ML methods to estimate heterogeneous

[2] For a review of targeting methods in developing countries see Hanna and Olken (2018).

[3] For example, Alatas et al. (2012); Stoeffler, Mills, and Del Ninno (2016); Karlan and Thuysbaert (2019); Banerjee et al. (2020); Gazeaud (2020); Beaman et al. (2021); Dupas, Fafchamps, and Houeix (2022); Trachtman, Permana, and Sahadewo (2022); Sumarto et al. (2025).

treatment effects and inform targeting strategies (e.g., targeting of workfare programs as in Bertrand et al. 2021; or allocation of business loans as in Bryan, Karlan, and Osman, 2024). The studies most closely related to ours are Lyons et al. (2025) and Altındağ et al. (2021), which apply machine learning methods to improve poverty targeting among forcibly displaced populations. While these papers address targeting, their primary focus is on the construction of poverty metrics. In contrast, our paper centers on treatment effect heterogeneity and outcome-specific targeting trade-offs. Specifically, we examine the synergies and trade-offs that arise when targeting for improvements in different nutritional outcomes. To our knowledge, this is the first study to apply ML methods to empirically assess food aid targeting based on predicted nutritional impact, rather than poverty alone.

The remainder of the paper is structured as follows. Section 2 presents the conceptual framework. Section 3 describes the research design, data, measurement of the primary nutritional outcomes, and average treatment effects. Section 4 outlines the methodological approach. Section 5 reports the empirical results including heterogeneity in treatment effects, synergies and trade-offs, and implications for targeting. Section 6 concludes.

2 Conceptual Framework

The nutritional impact of food aid depends not only on households' baseline status but also on how they adjust their consumption in response to different types of transfers. To illustrate these dynamics, we develop a simple conceptual framework to clarify the trade-offs involved in targeting two distinct types of food aid: (1) staple-heavy, and (2) nutrition-sensitive transfers.

Targeting Criteria and Household Typology

The ideal targeting criteria depends on the primary objectives of each food aid type. Staple-heavy food transfers are designed to prevent calorie shortfalls among food-insecure households. Accordingly, targeting in these programs should prioritize calorie-deficient households, while minimizing the risk of promoting over-consumption. Nutrition-sensitive transfers, by contrast, aim to increase dietary diversity by providing nutrient-rich foods that households may otherwise under-consume. Because households may under-invest in diverse diets for reasons beyond income—such as limited nutrition knowledge or failure to internalize long-term health returns—targeting solely on the basis of baseline dietary diversity may not maximize impacts. Instead, high predicted responsiveness to diverse foods, combined with low or moderate baseline diversity, is more informative for targeting. Behavioral responses such as waste or side-selling further complicate these decisions and may attenuate the benefits of transfers that are not well matched to household demand.

An alternative targeting objective relates to perceived food security, measured through subjective experiences rather than objective consumption. These measures do not always align with standard food consumption indicators (Lain, Tandon, and Vishwanath, 2024). In contexts characterized by price or supply volatility, in-kind transfers may meaningfully reduce households' anxiety or uncertainty even when food consumption indicators show modest change.

Table 1 formalizes these concepts by mapping households according to two key dimensions: their baseline status and their income elasticity of demand for additional calories (in the case of staple-heavy transfers) or for greater dietary diversity (in the case of nutrition-sensitive transfers). The table illustrates that extramarginal in-kind transfers, those that change consumption patterns relative to cash, can characterize transfers either to households with low baseline levels of consumption or households with low income elasticity of demand at baseline. Additionally, the framework distinguishes between impacts on the absolute level of calorie consumption and on achieving calorie sufficiency. This is important because the health gains from increased calorie consumption are not uniform and may turn negative when calorie sufficiency is exceeded. The framework also accounts for censoring in the dietary diversity measure, given the natural upper bound on the number of food groups that can be incorporated into a household's diet.

Staple-Heavy Transfers

For staple-heavy food transfers, targeting on the basis of deprivation would include all households with low baseline calorie intake (rows A1 and A2 in Table 1). In-kind provision is most justified when demand for calories is low and the transfer is extramarginal (A1). When demand is high (A2), a cash transfer with lower administrative costs could achieve similar outcomes. Since calorie sufficiency represents the most basic human need, households of type A1—those that would not prioritize staple food spending despite calorie deficits—are likely rare. Most targeting decisions concern households for whom calorie transfers are inframarginal and could improve both calorie intake and perceived food security (A2–A4). However, targeting based solely on perceived food security can promote over-consumption by including households with already sufficient calorie intake and high demand for additional calories (A4). Thus, for staple-heavy transfers, the central trade-off is less about deprivation versus predicted impact and more about whether to target improvements in calorie intake or reductions in food insecurity experience.

Nutrition-Sensitive Transfers

For nutrition-sensitive transfers, targeting decisions requires balancing deprivation, predicted impact, and competing nutritional outcomes. In-kind transfers are extramarginal for households with low baseline dietary diversity (B1), or moderate baseline diversity but low demand elasticity for diverse diets (B3). However, the greatest impacts are likely to occur among households with low to moderate baseline diversity and a high demand for diverse diets (B2 and B4), where transfers both complement preferences and increase diet quality. In such cases, in-kind transfers (rather than cash) may remain appropriate if they also enhance perceived food security. As with calorie intake, there is also an upper limit: once almost all food groups are already consumed at baseline (B5), additional transfers may not produce any measurable changes. Thus, for nutrition-sensitive food transfers, both the tradeoff between targeting impact versus deprivation and the tradeoff between competing nutritional outcomes are important considerations.

Behavioral and Institutional Considerations

These conceptual distinctions are further clarified in Appendix B, which presents stylized Engel curves illustrating how the nonlinear and composition-dependent effects of in-kind transfers produce heterogeneous impacts on calorie consumption and dietary diversity. Our conceptual framework also allows for behavioral and institutional frictions that shape household responses to food transfers. Preferences can evolve as beneficiaries receive informational signals about desirable diets through program delivery. Even inframarginal transfers may push consumption above the utility-maximizing level because of "stickiness" in purchasing patterns, which likely differs between staple and more nutritious items. For extramarginal transfers, households may consume most of the food delivered, though some waste may arise when items do not align with local foodways, and side-selling may occur even if it is inefficient. Together, these features underscore that both consumer demand and behavioral frictions are central to understanding heterogeneous impacts of in-kind food aid.

Implications and Motivation for Empirical Analysis

A limitation of our analysis is that we cannot distinguish empirically between households who would have realized similar nutritional outcomes with cash and those uniquely affected by in-kind transfers. Still, the empirical evidence indicates that many households prefer in-kind aid—often due to perceived certainty and insurance against inflation (Gadenne et al., 2021).[4] These practical considerations underscore the value of understanding heterogeneity in nutritional impacts of different types of in-kind aid, particularly where cash is not a viable alternative.

In sum, the conceptual framework motivates an approach that goes beyond targeting solely on baseline deprivation—whether calorie shortfalls or low dietary diversity. Although all of the household types outlined in Table 1 are theoretically possible, the extent to which trade-offs arise across baseline status, predicted impacts, and outcome dimensions is an empirical question. While we cannot directly observe household demand curves, we can examine whether observable characteristics systematically predict treatment effect heterogeneity in ways that can inform food aid targeting (as we do in Section 5). By leveraging these predictions, we assess the prevalence of households with low baseline need and low demand for additional calories or dietary diversity—such as types A1, B1, and B3 in Table 1—that may exhibit low predicted impacts. This analysis informs not only who should be targeted, but also how in-kind program design might maximize nutritional returns.

3 Research Design

We study targeting in the context of a large-scale evaluation of nation-wide food aid program in Egypt. This section provides a brief overview of the study setting and experimental design, with further details available in Abay et al. (2023). We then describe how key nutritional outcomes are measured and present average program impacts on calorie intake, food security, and dietary diversity.

3.1 Study Setting

This study was conducted in Egypt, a middle-income country with a complex nutritional landscape marked by the double burden of malnutrition: high rates of both obesity and childhood stunting (Ecker et al., 2016). For instance, the childhood stunting rate stands at 22.3% (UNICEF, WHO, and the World Bank Group, 2021), while the prevalence of iron-deficiency anemia is 47.2% among mothers and 39.6% among preschool children (Tawfik, Hanna, and Abdel-Maksoud, 2015). These nutritional challenges were further compounded at the time of data collection by sharp inflationary pressures, partly triggered by the Russian war in Ukraine. For instance, in September 2023, Egypt's inflation rate peaked at 38%, with food prices being a major contributor (CAPMAS, 2023; Reuters, 2023).

At the same time, Egypt's long-standing food subsidy system—the ration card program—has been criticized for poor targeting (Ahmed and Bouis, 2002; Gutner, 2002; Breisinger et al., 2024). Covering the majority of households and focusing heavily on calorie-dense staples, the program has contributed to the country's nutritional paradox: high levels of both undernutrition and obesity (Ecker et al., 2016). This context makes Egypt a particularly relevant setting for examining the trade-offs in targeting in-kind food transfers.

The food aid program evaluated in this study represents an effort to protect the poor and vulnerable against hunger and food insecurity. It is implemented by the Egyptian Food Bank (EFB), one of the largest nonprofit, non-governmental organizations in the country. The program distributes monthly

[4] Indeed, Abay et al. (2023) document persistent stated preferences for in-kind over cash among households in our sample.

food boxes containing non-perishable food items such as rice, pasta, cooking oil, and lentils. Priority is given to female-headed households with no stable income, especially those where the head is a primary caregiver for children or individuals with disabilities. Community-based organizations identify households that meet these criteria, and EFB caseworkers verify eligibility to ensure that assistance reaches those in need. This study examines whether the targeting of the EFB's food aid program can be improved beyond the baseline approach of community-based targeting.

3.2 Experimental Design

This paper uses data from a clustered randomized controlled trial (RCT), with the third administrative unit in Egypt serving as the level of clustering, henceforth referred to as a community.[5] The sample includes 150 communities spanning 22 governorates (the first administrative unit) in Egypt. Treatment randomization is stratified by governorate and urban/ rural status. Each community is randomly assigned to: (1) staple-heavy food box, (2) nutrition-sensitive food box, (3) a control group. Within each community, 16 households were randomly selected from those identified as eligible to benefit from the expansion of the food aid program. That is, all households in our sample are identified by community-based organizations as potential beneficiaries.

Table 2 provides a detailed list of the contents of the staple and nutritious food boxes. The staple food box contains 9 food items that are calorie-heavy but are relatively low in nutritional variety. On the other hand, the nutritious food box includes 17 food items that provide substantially higher nutritional variety and around 25% more calories compared to the staple food box. Therefore, the nutritious food box can be considered an upgrade to the staple-heavy food box that has a higher nutritional value but also a higher cost. The cost of the nutrition-sensitive food box was double the cost of the staple-heavy food box when this study was implemented in the last quarter of 2022.

3.3 Data and Measurement of Nutritional Outcomes

The data collected through the RCT not only enables the evaluation of the nutritional impacts of the two types of food boxes, as reported in (Abay et al., 2023), but also allows us to examine how these impacts vary with beneficiaries' characteristics. The baseline survey was conducted between August and October 2022, immediately before the food aid program was extended to treatment group beneficiaries. It includes detailed information on household characteristics, particularly those used in the proxy means test (PMT) targeting method. Appendix C provides a detailed list of the baseline covariates that we used in this heterogeneity analysis. The midline survey was carried out in March 2023, after beneficiaries in the treatment groups had received at least three rounds of monthly food box distributions. This timing ensures that the survey captures meaningful exposure to the intervention.

Both the baseline and midline surveys include data on our primary outcome variables: (1) household food energy intake (calorie intake), (2) the Food Insecurity Experience Scale (FIES), and (3) the Household Dietary Diversity Score (HDDS).

To compute data on food energy intake (kcal), we use a survey module that prompts the female household head to recall all food and beverages consumed over a 24-hour period, along with their respective ingredients and quantities. This detailed information on food consumption is then analyzed by researchers from Egypt's National Research Center, who break it down into nutrient components to estimate calorie intake as well as the intake of other macro and micro-nutrients.

To measure household-level food security, we use the Food Insecurity Experience Scale (FIES), a widely adopted self-reported metric developed by the FAO to assess challenges in accessing adequate

[5]The third administrative unit in Egypt is either a village or an urban census unit.

food due to financial or other resource constraints (Saint Ville et al., 2019). The FIES consists of an eight-question module capturing respondents' experiences with food insecurity over a seven-day recall period. The questions address the following dimensions: (1) worrying about not having enough food, (2) inability to eat healthy food, (3) eating a limited variety of foods, (4) skipping meals, (5) eating less than needed, (6) running out of food, (7) being hungry but not eating, and (8) going a whole day without eating—all due to a lack of money or resources. Responses to these questions are aggregated into an overall FIES score ranging from 0 to 8, where a score of 0 indicates no reported experience of food insecurity across all dimensions. Thus, lower FIES scores represent better food security outcomes.

The components of the Food Insecurity Experience Scale (FIES) can be further categorized into three levels of food insecurity: mild, moderate, and severe. These categories capture the progression of food insecurity experiences, ranging from worrying about food adequacy to more extreme conditions of hunger and deprivation. Mild food insecurity is typically reflected in concerns about food access and compromises in dietary quality (dimensions 1-3 above), while moderate food insecurity involves reducing food quantity or skipping meals (dimensions 4-5). Severe food insecurity represents the most extreme experiences, including going a whole day without eating due to a lack of resources (dimensions 6-8). By analyzing these categories, we can gain a deeper understanding of the varying intensities of food insecurity within the study sample and identify households most at risk.

To measure dietary diversity, we employ an adjusted version of the household dietary diversity score (HHDS) developed by the Food and Agriculture Organization (FAO) with a recall period of one week (Kennedy, Ballard, and Dop, 2013). Our HDDS measure is a simple count of the number of food groups consumed by all household members over the past week, out of a total of 12 food groups.[6]

3.4 Average Program Impacts on Calorie Intake, Food Security, and Dietary Diversity

As shown in Abay et al. (2023), both types of transfers have beneficial impacts on calorie intake, food insecurity experience (FIES), and dietary diversity.[7] Building on those results, we additionally examine calorie sufficiency—measured as the share of the recommended daily allowance (RDA) consumed or the shortfall relative to the RDA. This outcome is particularly relevant to concerns raised in the conceptual framework about the risk of over-consumption among some households (e.g., group A4).

According to Table 3, the control group maintains an average calorie intake (by household head) representing 78% of the recommended dietary allowance (RDA) (column 2), scores 6 out of 8 on the FAO's food insecurity experience scale (FIES) (column 4), and scores 9 out of 12 on the household dietary diversity score (HDDS) (column 5). These summary statistics suggest that households in our sample do struggle to consume sufficient quantities of food to meet the recommended amount of calorie intake or feel food secure. Table 3 shows that the staple food box improves calorie intake by 3% of the RDA, though this effect is statistically insignificant. It significantly reduces FIES by one-third of a point and increases HDDS by one-third of a point. By contrast, the nutritious food box significantly boosts calorie intake by 6% of the RDA, reduces FIES by 0.7 points, and increases HDDS by 0.9 points. The impact of the nutritious food box is significantly greater than that of the staple food box in improving dietary diversity and reducing food insecurity.

To better assess the cost-effectiveness of these interventions, we conduct additional analysis comparing the nutritional impacts relative to intervention cost. Although the nutritious box delivers

[6]These 12 food groups are: (1) foods made from grains, (2) white roots and tubers, (3) legumes nuts and seeds, (4) milk and milk products, (5) meat and organ meat, (6) fish and seafood , (7) eggs, (8) fruits, (9) vegetables, (10) oil and fats, (11) spices condiments and beverages, and (12) sweets and sweetened beverages.

[7]We follow a a standard ANCOVA regression framework to estimate ITT effects as presented in Appendix D

statistically larger improvements in dietary diversity and food security, these gains need to be interpreted in the context of the box's roughly double cost compared to the staple box. A direct test fails to conclusively reject the null that both boxes deliver equal impacts proportional to their costs (see p-value for testing double effect in Table 3). This motivates further investigation into heterogeneity-based targeting approaches that may yield improved outcomes at fixed budget levels.

An important finding from the calorie intake results is that the food box intervention helps narrow (though not fully close) the gap between actual calorie intake and recommended daily allowances. These impacts alleviate concerns that the intervention might push calorie consumption on average beyond recommended levels and contribute to overweight or obesity. However, they still leave open the possibility that the transfer result in over-consumption for some households (group A4).

Regarding food insecurity experience, Figure A.1 and Table A.1 in the appendix provide a detailed summary of the program's impact on individual components of the FIES. The nutritious food box significantly reduces food insecurity experience across all FIES component, consistent with the higher value and diversity of the box. The staple food box, while less diverse, produces effects of similar magnitude for FIES components related to both moderate (e.g., skipping a meal, eating less) and severe (e.g., enduring a whole day without food) food insecurity. This suggests that calorie-dense staple items are particularly effective in addressing more acute forms of food insecurity, as indicated by reductions in FIES scores. This distinction also highlights, as outlined in the conceptual framework, that impacts on food insecurity experience reflect not only objective indicators—such as reduced food quantity or variety—but also subjective dimensions, including worry and perceptions of dietary adequacy.

Moving on to dietary diversity, Figure A.2 and Table A.2, the increase in dietary diversity attributed to the nutritious food box is primarily driven by higher consumption of food items included in the nutritious food box. However, there are also notable increases in the consumption of food items not included in the nutritious food box, such as fish and vegetables. This likely reflects a relaxation of budget constraint, allowing households to reallocate funds toward other food items.[8] Thus, the findings align with the conceptual framework's emphasis on household behavior and demand for diverse food items as important drivers of heterogeneity in impact.

Taken together, the results from the intent-to-treat (ITT) estimates suggest that the in-kind food transfer improves food security and nutritional outcomes of the beneficiaries, on average. The nutritious food box outperforms the staple food box in enhancing dietary diversity and reducing food insecurity, as measured by FIES. However, the distinction in the effects of the two types of food transfers on calorie intake is less pronounced. This can be attributed to two factors. Firstly, although the nutritious food box represents an upgrade from the staple food box, the items in the latter are calorie-dense. Despite containing 90% more items, the nutritious food box only provides 25% more calories than the staple food box. Consequently, the enhancement in nutritional variety is more noteworthy than the increase in total calories. Secondly, unlike HDDS and FIES measures, the assessment of calorie intake using survey questionnaires inherently involves measurement error, which could limit the precision of our findings. Further analysis of average treatment effects on calorie intake is provided in Appendix E.

Finally, although the nutritious box delivers greater improvements in dietary diversity and perceived food security, its higher cost raises the concern that it may not be more cost-effective on average (see detailed analysis in (Abay et al., 2023)). This finding motivates further analysis of heterogeneity-based targeting strategies and how they may vary by treatment type.

[8] Table A.3 shows that both food boxes influence household food expenditure patterns. For example, households receiving either food box reduce their spending on certain items included in the boxes, such as legumes, while increasing spending on complementary items, such as fruits and snacks in the case of the nutritious food box. However, the overall impact on total food expenditure is statistically insignificant.

4 Methodology: Heterogeneity Analysis Using Machine Learning Methods

We follow the approach of Chernozhukov et al. (2023) in analyzing heterogeneity in treatment effects using machine learning (ML) techniques. A key feature of this approach is allowing for theoretically valid inferences while using ML methods.

4.1 Predicted Treatment Effects and Predicted Counterfactual

We begin by estimating the conditional average treatment effects (CATE) and baseline conditional average (BCA) for each household. To clarify the estimates of CATE and BCA, consider a basic model of the form:

$$Y = b_0(Z) + Ds_0(Z) + U, \qquad E[U|Z, D] = 0,$$

where Z is a vector of covariates, D is a treatment indicator. The coefficients $b_0(Z)$ and $s_0(Z)$ can be broken down into:

$$b_0(Z) = E[Y|D = 0, Z] \tag{1}$$

which is the baseline conditional average (BCA), predicted at the household level conditional on baseline covariates (Z), and

$$s_0(Z) = E[Y|D = 1, Z] - E[Y|D = 0, Z] \tag{2}$$

which is the conditional average treatment effect (CATE), also , predicted at the household level based on baseline covariates (Z). The CATE, also sometimes referred to as individual-treatment-effect (ITE) as in Bryan, Karlan, and Osman (2024), provides estimates for the treatment effect of each observation enabling flexible analysis of heterogeneity that goes beyond what is possible in the regression framework. The BCA similarly represents the predicted counterfactual outcome for each observation in the absence of the treatment.

We estimate both the BCA and CATE using a repeated cross-fitting machine learning procedure (details in Appendix F). Multiple supervised algorithms—random forest, elastic net, support vector machines, and gradient boosted trees—are trained to predict potential outcomes under treatment and control using only baseline covariates. In each of at least 100 iterations, the sample is randomly split into training and validation sets. The ML algorithms are trained on a subset of the data from the treatment group to predict $E[Y|D = 1, Z]$ and a subset of the data from the control group to predict $E[Y|D = 0, Z]$ in each iteration. CATEs are computed as the out-of-sample difference between $E[Y|D = 1, Z]$ and $E[Y|D = 0, Z]$ outcomes for each household. The final CATE predictions are obtained by taking the median across all iterations. This procedure is implemented separately for each treatment contrast (i.e., staple food box vs. control, and nutritious food box vs. control) to ensure that predictions are tailored to each program type while avoiding overfitting.

We use the resulting BCA and CATE estimates to examine potential trade-offs in targeting. In particular, we assess whether prioritizing households with the lowest counterfactual outcomes (i.e., the most deprived) is in tension with targeting those predicted to benefit most from the intervention. To do so, we compute the correlation between the predicted counterfactual outcome, $b_0(Z)$, and the predicted treatment effect, $s_0(Z)$.

In addition, we test for heterogeneity in treatment effects conditional on baseline covariates. Following Chernozhukov et al. (2023), we estimate the best linear predictor (BLP) of the CATE, defined as:

$$BLP := \beta_1 + \beta_2(S(Z) - ES(Z)) \tag{3}$$

where $S(Z)$ is an estimator for the CATE. β_2 captures the extent of treatment effect heterogeneity. Rejecting the null hypothesis $\beta_2 = 0$ implies that there is heterogeneity and affirms that $S(Z)$ is its relevant predictor. We estimate the BLP using a weighted linear projection, as in Chernozhukov et al. (2023).

4.2 Sorted Group Average Treatment Effects (GATES) and Characteristics of Each Group

To examine the magnitude of predicted treatment effects across the distribution, we sort the sample into quantiles of the predicted treatment effect using $S(Z)$. The sorted group average treatment effects (GATES) are then estimated using the following regression:

$$Y = \alpha_1 + \alpha_2 B(Z) + \sum_{k=1}^{5} \gamma_k [D - P(Z)] \cdot 1_{G_k} + \nu, \tag{4}$$

where Y is the outcome of interest, and $P(Z)$ represents the projection parameter defined as $\hat{E}(D|Z)$. The projection coefficient γ_k, which is the parameter of interest, is defined as $E[S_0(Z)|G_k]$. This coefficient represents the Group Average Treatment Effects (GATES). The groups are defined as $G_k = \{S(Z) \in I_k\}$, where I_k corresponds to the k-th quantile of the predicted treatment effect. For example, G_5 captures households in the top quantile of the distribution of the predicted treatment effect for a specific outcome. This regression is estimated using the weights: $w(Z) = \{P(Z)(1 - P(Z))\}^{-1}$.[9]

Next, we compare the characteristics of households in the top and bottom quantiles of the predicted treatment effect distribution, i.e., those in G_5 versus G_1. This exercise aims to identify observable characteristics that differentiate households with the highest predicted treatment effects from those with the lowest. While this method does not explicitly reveal which characteristics drive heterogeneous treatment effects, examining these differences provides valuable insights into potential targeting criteria. By prioritizing beneficiary households with characteristics more prevalent among the most impacted groups, we can improve targeting strategies and enhance the effectiveness of interventions.

Additionally, we inspect whether households with high predicted treatment effects on calorie intake also experience larger impacts on other outcomes, such as dietary diversity and perceived food security. Formally, we analyze how belonging to a specific heterogeneity group for calorie intake, Y, influences the treatment effect on another outcome, $\tilde{Y}$. To perform this analysis, we use equation 4, replacing Y as the dependent variable with the alternative outcome variable $\tilde{Y}$.[10]

Identifying synergies or trade-offs between improving calorie intake and increasing dietary diversity or perceived food security is pivotal. For instance, if households with high predicted impact on calorie intake also experience substantial gains in dietary diversity and perceived food security, it suggests that interventions targeting calorie intake (e.g., a staple food boxes) may yield broader benefits. In contrast, if the impacts on other outcomes are limited or inconsistent, it may highlight the need for alternative interventions to address these dimensions effectively.

Finally, we extend the household-level heterogeneity analysis by treating the staple box as the counterfactual for the nutritious food box. Specifically, we drop the control group and use machine learning to estimate CATEs for the nutritious food box relative to the staple box. As before, we assess treatment effect heterogeneity using the BLP and GATES estimates. This approach is motivated by the cost differential between the two interventions. In particular, it enables simulations of targeted

[9]As noted by Chernozhukov et al. (2023), the interaction $[D - P(Z)] \cdot 1_{G_k}$ is orthogonal to all other regressors that are functions of Z under the weights $w(Z)$.

[10]This analysis parallels that conducted in Bertrand et al. (2021) and Bryan, Karlan, and Osman (2024), with the latter referring to it as conditional group average treatment effects (CGATEs).

allocations in which the more expensive nutritious box is provided only to households predicted to benefit most. Given the roughly twofold higher cost of the nutritious box during implementation, we assess whether prioritizing 50% of the sample that would benefit the most from an upgrade to the nutritious food box can improve aggregate outcomes without raising overall program costs.

5 Empirical Results: Heterogeneity, Trade-offs, and Targeting Implications

This section empirically tests the extent to which heterogeneous groups—defined by baseline need and predicted impacts across outcomes, as described in the conceptual framework—exist within the targeted population. We use machine learning methods to predict treatment effect heterogeneity based on observable characteristics. The analysis demonstrates significant heterogeneities in program impacts across households, with observable traits such as education, unemployment, and asset ownership effectively predicting which households benefit most. Subgroups with the highest predicted impacts are expected to experience nutritional gains exceeding 200% of the average treatment effects—a pattern observed consistently across all nutritional outcomes and both food aid types. These results suggest that data-driven targeting could significantly improve the effectiveness of food aid programs by identifying households with the greatest potential to benefit.

Our findings also shed light on the potential trade-offs and synergies involved in food aid targeting. For calorie intake and dietary diversity, targeting by predicted impact largely overlaps with targeting by deprivation, supporting the hypothesis that well-designed interventions can simultaneously address need and effectiveness. However, for food insecurity experience—particularly under the staple box—some trade-offs emerge. The most deprived households do not always experience the greatest improvements, highlighting the conceptual framework's caution against relying solely on baseline status. The analysis further reveals that unemployment-related vulnerability is a key predictor of high impact from the staple box on both calorie intake and food security, while education level and household structure emerge as key predictors of the nutritious food box's impact on dietary diversity. These nuanced patterns emphasize the importance of clearly defining program objectives and tailoring targeting criteria to specific outcomes when designing food aid interventions.

5.1 Significant Heterogeneity in Impacts Exists and Can Be Predicted Using Observable Characteristics

We demonstrate that observable characteristics –which can feasibly be used for targeting– are successful in predicting treatment effects as well as heterogeneity in treatment effects. We focus on baseline covariates that can either be obtained from administrative data or easily verified by surveyors during household visits.[11] While we did collect data on the baseline values of our primary nutrition outcomes, we excluded these variables from the heterogeneity analysis to ensure our findings offer practical, policy-relevant insights for targeting.

Using machine learning (ML) methods, we uncover three key findings. First, the distribution of predicted treatment effects accents the difference between the impacts of the nutritious and staple food boxes on dietary diversity and food security. Second, estimates based on group average treatment effects (GATES) reveal substantial heterogeneity, suggesting considerable scope to increase aggregate program benefits by prioritizing households with the largest predicted impacts. Third, accounting for the cost differential between the two food boxes, we show that using ML methods to target the more

[11] Appendix C provides a detailed list of the baseline covariates used in this heterogeneity analysis.

expensive nutritious box to high-benefit households delivers greater nutritional gains than the status quo of community-based targeting of the staple box under a comparable budget. These calculations, however, abstract from any additional operational costs entailed in the collection of administrative data and the implementation of targeted methods.

Graphical evidence in Figure 1 illustrates the contrast in effect distributions between the two food aid types. The figure compares between the distribution of the conditional average treatment effects (CATEs) for the nutritious food box and the staple food box treatments across our three main outcomes. For calorie intake, the two distributions exhibit some overlap. However, the CATE distribution for the nutritious food box is flatter, indicating more variation in household-level impacts, whereas the staple box shows a more concentrated distribution of effects. In contrast, for food insecurity experience and dietary diversity, the CATE distribution for the nutritious food box is consistently shifted outward, reflecting higher predicted treatment effects across the population. Notably, the nutritious food box shows consistently stronger predicted effects on dietary diversity across the entire distribution of household-level estimates.

Formal heterogeneity analysis using ML methods (as detailed in Section 4) demonstrates statistically significant heterogeneity in program effects for both boxes on most outcomes.[12] This is reflected in a significant heterogeneity coefficient (β_2) from the best linear predictor (BLP) estimates (panel A of Table 4) and a significant difference between predicted impacts for the top and bottom quantiles in the GATES estimates (panel B of Table 4). These findings indicate that observable characteristics, similar to those used in the Proxy Means Testing (PMT) approach, effectively predict heterogeneity in treatment effects across the three primary outcomes.

The magnitude of predicted treatment effect heterogeneity is remarkable. Comparing the GATES estimates for the quantile of households with the largest predicted improvements in nutritional outcomes, as shown in Panel B of Table 4, to the ATE results in Table 3 reveals striking patterns. For the staple food box, the groups with the greatest predicted benefits (G5 for calorie intake and dietary diversity, and G1 for FIES) are expected to experience improvements approximately 470%, 260%, and 250% of the average treatment effects for calorie intake, FIES, and HDDS, respectively. Similarly, for the nutritious food box, the group with the largest predicted gains also achieves improvements around 320%, 275%, and 200% of the average effects for these same outcomes. These findings highlight the potential for observable baseline characteristics to identify subgroups with substantially greater response to each food aid type.

To further understand the extent of heterogeneity, we examine the range of predicted treatment effects across the distribution. Panel B of Table 4 indicates that the difference in predicted impacts on calorie intake between the most and least affected groups exceeds 300 kcal/day for both the staple and nutritious food boxes. Given an average calorie intake of 1,581 kcal/day in the control group (as shown in Table 3), this difference represents 20% and 26% of the control group's calorie intake for the staple and nutritious food box treatments, respectively. Similarly, for the food insecurity experience (FIES), the difference in predicted impacts between the most and least affected groups corresponds to 15% and 30% of the control group's FIES score for the staple and nutritious food box treatments, respectively. Finally, for the household dietary diversity score (HDDS), the difference in predicted impacts between the most and least affected groups represents 9% and 18% of the control group's HDDS for the staple and nutritious food box treatments, respectively. These results further emphasize the substantial variation in impact across subgroups, underscoring the potential to improve targeting by using observable household characteristics that predict treatment effect heterogeneity.

Figure 2 provides further insight into these patterns. For calorie intake, significant impacts arise

[12] In Table 4, we present the results from the "best" ML algorithm, defined as the one that minimizes prediction errors. Results from the other three ML algorithms are available upon request.

only in the top quantile for both food boxes (Panel A), indicating that benefits are concentrated among a relatively small subset of households. A similar pattern appears for food insecurity experience with the staple food box (Panel B.1), where only the highest quantile shows significant reductions. In contrast, the nutritious food box demonstrates broader impacts, with significant reductions in food insecurity experience across multiple quantiles (Panel B.2). These findings suggest that the staple food box yields high impacts on calorie intake or perceived food security only for a limited subset of households.

Disaggregation of FIES components (Figure A.3 and Figure A.4) reveals that heterogeneity is driven by different underlying mechanisms for each box. For the staple food box, heterogeneity is concentrated in moderate and severe food insecurity indicators. By contrast, the nutritious food box exhibits significant heterogeneity across a broader set of food insecurity domains.

The dietary diversity outcome presents the clearest contrast. Panel C of Figure 2 shows that the staple food box yields no statistically significant improvements at any quantile. On the other hand, the nutritious box significantly improves dietary diversity across all but the lowest quantile. This broad distribution of significant impacts of the nutritious food box suggests that households are responsive to the design of the box, which prioritizes nutritional diversity.

Finally, to assess cost-effectiveness more explicitly, we re-estimate treatment effect heterogeneity using the staple box as the counterfactual. This reframing is motivated by the cost difference between the two interventions. Results in Table 5 show substantial heterogeneity: households in the top 50% of predicted impacts experience statistically significant and sizable improvements in calorie intake, food insecurity experience, and dietary diversity. Specifically, the GATES estimates for this group exceed the staple box ATE by 46% for calorie intake, 194% for food insecurity, and 170% for dietary diversity. Importantly, allocating the more costly nutritious box only to these higher-impact households—at roughly the same aggregate cost as the status quo provision of the staple box—would generate substantially larger nutritional gains. These results suggest a promising role for predictive targeting in enhancing the cost-efficiency of nutrition-sensitive aid.

In summary, the ML analysis confirms the extent to which the nutritious food box outperforms the staple food box in improving dietary diversity and food security experience. Importantly, the ML models, which do *not* include baseline values of the outcome variables as predictors, successfully predict heterogeneity in treatment effects using observable characteristics. These findings highlight the potential for data-driven targeting approaches to enhance the effectiveness of food aid programs. In the next sections, we will build on these results in three ways. First, we will examine potential synergies and trade-offs between nutritional outcomes (Section 5.2). Second, we will examine whether there is a tradeoff between targeting by impact versus targeting by deprivation across different types of food transfers (Section 5.3). Third, we will investigate whether the households with the highest predicted treatment effects significantly differ from those with the lowest predicted treatment effects on a set of observable characteristics (Section 5.4).

5.2 Synergies and Trade-offs Across Nutritional Outcomes: Linking Calorie Intake with Dietary Diversity and Food Security

We examine how the same food box intervention may generate synergies or trade-offs across different nutritional outcomes. To assess these dynamics, we stratify households by their predicted treatment effects on calorie intake and analyze corresponding impacts on dietary diversity and food insecurity.[13]

[13] In Appendix E, we examine how treatment effects on calorie intake vary across households with different baseline consumption levels. We also assess impact heterogeneity across the full distribution of midline calorie intake using quantile regressions.

Table 6 presents the treatment effect estimates for the staple and nutritious food boxes separately. At the bottom of the table, we provide tests for the equality of treatment effects across all quantiles and between the bottom and top quantiles.

For the staple food box, column (1) of Table 6 shows that households in higher quantiles of predicted impact on calorie intake experience greater reductions in food insecurity, though this relationship is non-linear. While we reject the null hypothesis of equal impacts across all quantiles (p-value of 0.03), the impact on FIES for the top and bottom quantiles does not differ significantly (p-value of 0.30). Expectedly, column (2) indicates that the staple food box has a weak impact on dietary diversity across most quantiles of predicted impact on calorie intake.

In contrast, columns (3) and (4) of Table 6 demonstrate that the nutritious food box has significant effects on both reducing food insecurity and improving dietary diversity across most quantiles. Households in the top quantile of predicted calorie impact experience stronger improvements in both food security and dietary diversity relative to those in the bottom quantile. The difference in impacts is statistically significant for food security (p-value = 0.05) and marginally significant for dietary diversity (p-value = 0.13).

All in all, these results highlight the potential of the nutritious food box to address multiple dimensions of nutritional well-being. However, the staple food box in our setting primarily mitigates calorie deficits while simultaneously reducing food insecurity. This finding indicates that instances where gains in calorie consumption do not improve perceived food security, as outlined in our theoretical framework, are not prevalent in our context.

5.3 Limited Trade-offs in Food Aid Targeting: Evidence from Predicted Impact and Predicted Deprivation

A critical question for improving targeting of social safety net programs is whether there is a trade-off between improving targeting based on one specific objective function versus another (Haushofer et al., 2025). In this study, we focus on targeting of two types of food transfers with distinct objectives: a staple-heavy food box that prioritizes the provision of calorie-dense foods, and a nutrition-sensitive food box that emphasizes nutritionally diverse food items. Thus, we aim to determine whether beneficiaries who benefit significantly from one type of food transfer are also expected to experience high impacts from the other, or if their impacts are specific to one type of intervention. In addition, we examine whether there is a trade-off between targeting households with the highest predicted treatment effects (the most impacted) versus those with the lowest predicted counterfactual outcomes (the most deprived). Our results indicate that targeting trade-offs are limited and emerge primarily in outcomes related to food insecurity experience.

Figure 3 demonstrates a positive relationship between the predicted treatment effects of the staple food box and that of the nutritious food box for the outcomes on calories intake and dietary diversity. Table A.4 in the Appendix confirms that these positive correlations are statistically significant. For the outcome on food insecurity experience, however, there is hardly any relationship between the predicted treatment effects of the staple and the nutritious food boxes. This could be due to the subjectivity of the FIES measure. Hence, we can assert that there is not a substantial trade-off between improving the targeting for the staple versus the nutritious food box interventions based on heterogeneity in predicted treatment effects.

Next, we turn to the relationship between predicted treatment effects and predicted counterfactual, testing the theoretical trade-offs outlined in Section 2. Our framework posits that potential trade-offs between targeting for impact and targeting for deprivation in the case of food transfers are likely to be limited to a few scenarios. Empirically, we find that such trade-offs are only evident in the staple food box's impact on food insecurity experience.

Using a machine learning approach, we estimate the conditional average treatment effect (CATE) and the baseline conditional average (BCA) based on household characteristics, as explained in Section 4.1. The correlation between the two captures whether households with lower baseline outcomes (greater deprivation) also experience larger predicted treatment effects. Figure 4 and Table 7 reveal a significantly negative correlation between CATE and BCA for the outcomes on calorie intake and dietary diversity for both types of food boxes. This pattern suggests that households with higher baseline deprivation (lower BCA) tend to exhibit stronger treatment effects (higher CATE), aligning with the hypothesis that targeting the most deprived can maximize impact for nutritional outcomes.

However, the staple food box's impact on the FIES deviates from this pattern. Here, CATE and BCA exhibit a weak positive correlation ($\rho = 0.34$), suggesting that households with moderate baseline food insecurity—not the most severely deprived—experience the greatest benefits. This relationship aligns with the hypothesis that, for subjective or threshold-based outcomes, households at intermediate levels of deprivation are most likely to experience discrete improvements in response to assistance.

Taken together, our findings suggest that targeting the most impacted households generally also reaches the most deprived, especially for calorie intake and dietary diversity. However, for food insecurity outcome (FIES), particularly under staple-heavy intervention, some trade-offs may emerge. If improving FIES is a key policy objective (e.g., as a proxy for perceived adequacy or satisfaction), further research is needed to refine the design of the food transfer or the targeting criteria to better serve the most vulnerable.

5.4 Identifying the Most Impacted: Unpacking Household Characteristics and Targeting Implications

Having established that observable characteristics can predict significant heterogeneity in treatment effects among beneficiaries, we next compare the attributes of households with the highest and lowest predicted impacts. This analysis seeks to identify consistent patterns in the characteristics of households with the highest predicted treatment effects and to assess whether these patterns persist across all primary outcomes.

We find that, for the staple food box, characteristics indicative of a temporary shockare associated with higher predicted impacts on increasing calorie consumption and, to a lesser extent, on reducing FIES. For the nutrition-sensitive food box, the primary predictors of higher impacts on HDDS include the education level of the household head, household size, and lack of a ration card. These distinctions in the characteristics of the most impacted beneficiaries across programs and objectives align with plausible proxies for demand for staple calories and dietary quality, suggesting potential benefits from a differentiated targeting approach.

Table 8 shows that, for the staple food box, households with the highest predicted treatment effects on calorie intake exhibit slightly higher socioeconomic status but face greater unemployment-related vulnerabilities compared to those with the lowest predicted treatment effect. This distinction is evident across three key sets of characteristics:

1. Household Demographics: The most impacted households tend to have a more educated head but a higher number of unemployed members.
2. Dwelling Characteristics: They are less likely to reside in a rural house or rented accommodations and less likely to have wooden roofs (as opposed to cement roofs). However, they are more likely to reside in dwellings without private bathrooms—an outlier characteristic given that less than

20% of the sample lacks private bathrooms.

3. Asset Ownership: They are more likely to own higher-category washing machines and have internet access.

Ownership of higher-category appliances alongside unemployment reflects an "asset-rich, cash-poor" scenario often observed in middle-income countries. This pattern is suggestive of a temporary shock that heightens a household's need for food aid, particularly during periods of high inflation and in contexts where complementary social assistance interventions may be slow to materialize.

On the other hand, the heterogeneity in the treatment effect of the nutritious food box on calorie intake reveals a distinct pattern. Table 8 shows that households with the highest predicted treatment effect exhibit lower socioeconomic status compared to those with the lowest predicted treatment effects.

1. Household Demographics: The most impacted households tend to have lower education levels and limited access to pension benefits.
2. Dwelling Characteristics: They are more likely to reside in rural areas or in dwellings lacking private bathrooms or modern toilets and tend to have lower electricity bills.
3. Asset Ownership: They are significantly more likely to own basic (non-smart) mobile phones and regular (low-category) washing machines.

We posit that the nutritious food box may generate higher impacts on calorie consumption among low-socioeconomic households not only due to greater need but also because structural constraints may incentivize direct consumption. Nutrient-dense items such as red meat and dried fruits may be extramarginal for these households due to cost prohibitions, aligning with theories of constrained consumption (Jensen and Miller, 2008). Lacking the means to purchase such items independently, these households are more likely to consume the provided food, increasing calorie intake. Furthermore, rural residence and low electricity bills may serve as proxies for limited market integration, reducing resale opportunities and further reinforcing direct consumption.

Turning to the staple food box's effects on reducing food insecurity, Table 9 reveals a less distinct pattern in the characteristics distinguishing households with the highest predicted treatment effects.

1. Household Demographics: The most impacted households tend to have heads with higher levels of education but are also less likely to have income from work. Additionally, they exhibit a lower dependency ratio but slightly more unemployed members, which is probably a consequence of having an unemployed head.
2. Dwelling Characteristics: They are less likely to have cement roofs.
3. Asset Ownership: They are more likely to own lower-category washing machines.

Unemployment-related vulnerability emerges as a common characteristic among households with the highest predicted impacts from the staple food box on both increasing calorie consumption and reducing food insecurity. Households facing unemployment are particularly responsive to staple food assistance, likely because the absence of labor income directly constrains their ability to meet basic food needs.

Table 9 shows that, for the nutritious food box treatment, the most impacted households in terms of decreasing reported food insecurity experience differ notably from those least affected households across three key sets of characteristics.

1. Household Demographics: Older, more educated heads who are more likely to earn income from work, fewer members enrolled in education, and slightly more members without pension benefits.

2. Dwelling Characteristics: Greater likelihood of residing in dwellings with cement roofs, private bathrooms, and modern toilets, as well as higher electricity bills.
3. Asset Ownership: More likely to own higher-category washing machines, such as half- or full-automatic models.

Accordingly, in contrast to the findings from the heterogeneity analysis of the treatment effects on calorie intake, households with the highest treatment effect on food security experience tend to have higher socioeconomic status. It is plausible that older and more educated household heads with higher socioeconomic status are more likely to appreciate the impact of the nutritious food box on their food security experience.

Table 10 shows that the characteristics distinguishing households with the highest predicted treatment effects on dietary diversity are consistently similar for both the staple and nutritious food box treatments.

1. Household Demographics: More educated household heads, smaller household size, lower dependency ratios, and fewer members enrolled in education.
2. Dwelling Characteristics: Greater likelihood of residing in rented units or in dwellings with cement (rather than wood) roofs. However, they are also more likely to lack private bathrooms and tend to incur lower electricity bills.
3. Asset Ownership: A higher likelihood of owning regular (low-category) washing machines but also internet access.

The most consistent trend observed among the characteristics distinguishing households with the highest treatment effects on dietary diversity pertains to household demographics. It is reasonable to expect that households with a more educated head would exhibit greater impacts on dietary diversity, if education level is positively correlated with awareness of the importance of a varied diet. This is particularly relevant for a sample comprising female-headed households in a society where prevailing social norms dictate that women are primarily responsible for food preparation. Moreover, the finding that smaller households with fewer dependents demonstrate higher predicted treatment effects on dietary diversity aligns with the fact that all beneficiary households receive the same food box, regardless of household size.

An important characteristic for heterogeneity in treatment effects of a food aid intervention is whether households have access to a ration card.[14] Table 8 and Table 10 show that heterogeneity in treatment effects with respect to access to ration card differs for the treatment effects on calorie intake and dietary diversity. For calorie intake, households with higher treatment effects are more likely to have ration cards, indicating a complementary relationship between ration cards and food boxes. Conversely, for dietary diversity, households with higher treatment effects are less likely to have ration cards, suggesting a substitution relationship between the two. This implies that while ration cards and food boxes may complement each other for calorie intake, they serve as substitutes for dietary diversity. This aligns with the notion that households who have greater access to subsidized food consume more and increase their calories intake by more. In contrast, for dietary diversity, households may target a certain level of dietary diversity. They can achieve that level through either ration cards or food boxes.

Overall, these findings suggest that the characteristics predicting high treatment impacts vary by both food box type and outcome measure. For example, unemployment-related vulnerability is a consistent predictor of greater impact from staple food box on both calorie intake and food insecurity,

[14] In Egypt, the ration card program gives households access to subsidized food items and traditional (*baladi*) bread allowance (Breisinger et al., 2024).

while lower socioeconomic status are more relevant for high impacts from nutritious food boxes on calorie intake. In contrast, higher socioeconomic status predicts greater improvements in food security experience from the nutritious food box. For dietary diversity, education level and household composition emerge as the most salient factors across both interventions. The observed differences in predictors across outcomes underscore the importance of clearly defining program objectives when designing targeting criteria. Given that calorie intake and FIES are conceptually distinct measures of food security, policymakers may need to prioritize one outcome over another in setting targeting objectives for this food aid program.

5.5 Discussion: Why Targeting Strategy Depends on What We Measure?

Understanding the diverse results presented above requires recognizing the multidimensional nature of both baseline need and treatment impact within nutritional outcomes. For more objective nutritional measures, such as calorie intake and dietary diversity, households with the greatest baseline deprivation tend to benefit the most. This pattern explains the synergies identified in Section 5.3 between need and impact, and, to a lesser extent, the distinct household-level predictors discussed in Section 5.4.

In contrast, the relationship between need and impact is more nuanced for subjective outcomes like the Food Insecurity Experience Scale (FIES). Here, households not necessarily among the most severely deprived sometimes report the largest improvements. This may reflect that households on the margins of food insecurity are more likely to cross salient thresholds in response to assistance, or that they possess greater awareness or ability to articulate changes in their food security status.

This distinction implies that targeting will be more effective when tailored to specific program objectives. For example, maximizing improvements in dietary diversity or calorie intake may be achieved by targeting based on indicators of deprivation. However, enhancing perceived food security requires a more nuanced strategy that considers behavioral responses, awareness, and the potential for threshold effects.

Taken together, our findings challenge the efficacy of a "one size fits all" targeting approach. While we find limited trade-off between targeting for impact and for deprivation, the households predicted to benefit most for one outcome are not necessarily those predicted to benefit most for others. By jointly examining the synergies and trade-offs across nutritional outcomes alongside the household characteristics that predict treatment effect heterogeneity, our analysis offers guidance for aligning food aid program design and targeting with diverse household needs and policy goals.

6 Conclusion

This study contributes new empirical evidence on how targeting of in-kind food transfers can be improved, drawing on two distinct food aid interventions: a staple-heavy food box and a nutrition-sensitive food box. Using data from a randomized evaluation of a nationwide food aid program in Egypt, we apply machine learning methods to uncover substantial heterogeneity in treatment effects across three key nutritional outcomes: calorie intake, food insecurity experience (FIES), and household dietary diversity (HDDS). We show that observable household characteristics, such as household demographics, dwelling characteristics, and asset ownership, effectively predict which households experience the largest benefits, underscoring the value of data-driven approaches to targeting.

We present a simple conceptual framework to motivate our empirical analysis. A central proposition of the conceptual framework is that targeting based solely on baseline need may not capture the full potential for impact, particularly when food aid distorts consumption patterns. In such cases, household demand for dietary quality may interact with baseline nutritional status to shape the response to in-kind transfers. This builds on Haushofer et al. (2025)'s equity-efficiency framework, but shifts the focus from poverty metrics to nutritional optimization, a dimension less explored in earlier targeting analyses.

Our results indicate that targeting for need and targeting for impact are highly, though not perfectly, correlated. The overlap is strong for outcomes such as calorie intake and dietary diversity, suggesting that well-designed interventions can simultaneously address deprivation and effectiveness. However, notable exceptions arise. For food insecurity experience, the correlation between deprivation and predicted impact is weaker, especially under the staple food box intervention. These nuances echo Alatas et al. (2012)'s finding that no single targeting method dominates across all welfare dimensions.

Moreover, the results highlight how program design influences the distribution of impacts. The nutritious food box delivers broader and more consistent improvements in dietary diversity and food security. In contrast, the staple food box delivers more concentrated benefits, particularly in calorie intake and food security, among a narrower group of households characterized by acute, often temporary, vulnerabilities such as unemployment. From a policy perspective, the lower cost of the staple food box may create a temptation to distribute it more widely. However, our analysis suggests the staple food box should be more narrowly targeted, given that a large share of the recipients are not predicted to experience significant nutritional benefits.

Synthesizing the evidence, we identify both synergies and trade-offs across program objectives. For calorie intake and dietary diversity, there are strong synergies: households with greater predicted deprivation also tend to be those who derive the most benefit, and predicted treatment effects are positively correlated across box types. In contrast, for food insecurity experience (FIES), a trade-off emerges: households predicted to gain the most from the staple-heavy food transfer are not always those with the highest baseline deprivation. These patterns suggest that targeting strategies may need to be adjusted depending on whether the goal is to reduce perceived food insecurity or address objective consumption shortfalls.

Ultimately, targeting decisions entail normative judgments about which objective function to optimize and which outcomes to prioritize. This study demonstrates that even when objectives are positively correlated on average, substantial heterogeneity in treatment effects remains. By leveraging observable data to predict impact, policymakers can better align food aid interventions with household needs, improving both the efficiency and equity of food aid programs. Continued research is warranted to refine targeting algorithms and to assess the long-term welfare impacts of differentiated food aid strategies.

References

Abay, Kibrom A, Lina Abdelfattah, Mohamed Elkaramany, Dalia Elsabbagh, and Sikandra Kurdi. 2023. "Nutrition-sensitive food distribution amidst inflationary shock: Evidence from a randomized intervention in Egypt." *IFPRI Discussion Paper* .

Ahmed, Akhter U and Howarth E Bouis. 2002. "Weighing what's practical: proxy means tests for targeting food subsidies in Egypt." *Food Policy* 27 (5-6):519–540.

Alatas, Vivi, Abhijit Banerjee, Rema Hanna, Benjamin A Olken, and Julia Tobias. 2012. "Targeting the poor: evidence from a field experiment in Indonesia." *American Economic Review* 102 (4):1206–1240.

Alderman, Harold. 2014. "Can transfer programs be made more nutrition sensitive?" .

Altındağ, Onur, Stephen D O'Connell, Aytuğ Şaşmaz, Zeynep Balcıoğlu, Paola Cadoni, Matilda Jerneck, and Aimee Kunze Foong. 2021. "Targeting humanitarian aid using administrative data: Model design and validation." *Journal of Development Economics* 148:102564.

Bah, Adama, Samuel Bazzi, Sudarno Sumarto, and Julia Tobias. 2019. "Finding the poor vs. measuring their poverty: Exploring the drivers of targeting effectiveness in Indonesia." *The World Bank Economic Review* 33 (3):573–597.

Banerjee, Abhijit, Rema Hanna, Benjamin A Olken, and Sudarno Sumarto. 2020. "The (lack of) distortionary effects of proxy-means tests: Results from a nationwide experiment in Indonesia." *Journal of Public Economics Plus* 1:100001.

Beaman, Lori, Niall Keleher, Jeremy Magruder, and Carly Trachtman. 2021. "Urban networks and targeting: Evidence from liberia." In *AEA Papers and Proceedings*, vol. 111. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203, 572–576.

Bertrand, Marianne, Bruno Crépon, Alicia Marguerie, and Patrick Premand. 2021. "Do workfare programs live up to their promises? Experimental evidence from Cote D'Ivoire." Tech. rep., National Bureau of Economic Research.

Breisinger, Clemens, Yumna Kassim, Sikandra Kurdi, Josee Randriamamonjy, and James Thurlow. 2024. "From food subsidies to cash transfers: Assessing economy-wide benefits and trade-offs in Egypt." *Journal of African Economies* 33 (2):109–129.

Brown, Caitlin, Martin Ravallion, and Dominique Van de Walle. 2018. "A poor means test? Econometric targeting in Africa." *Journal of Development Economics* 134:109–124.

Bryan, Gharad, Dean Karlan, and Adam Osman. 2024. "Big loans to small businesses: Predicting winners and losers in an entrepreneurial lending experiment." *American Economic Review* 114 (9):2825–2860.

CAPMAS. 2023. "Inflation Data." Data retrieved from the Central Authority for Public Mobilization and Statistics (CAPMAS) Database, https://www.capmas.gov.eg/.

Chernozhukov, Victor, Mert Demirer, Esther Duflo, and Ivan Fernandez-Val. 2023. "Generic Machine Learning Inference on Heterogenous Treatment Effects in Randomized Experiments, with Application to Immunization in India." .

Cunha, Jesse M. 2014. "Testing paternalism: Cash versus in-kind transfers." *American Economic Journal: Applied Economics* 6 (2):195–230.

Dupas, Pascaline, Marcel Fafchamps, and Laura Hernandez-Nunez. 2024. "Keeping up appearances: An experimental investigation of relative rank signaling." Tech. rep., National Bureau of Economic Research.

Dupas, Pascaline, Marcel Fafchamps, and Deivy Houeix. 2022. "Measuring relative poverty through peer rankings: Evidence from Côte d'Ivoire." Tech. rep., National Bureau of Economic Research.

Ecker, Olivier, Perrihan Al-Riffai, Clemens Breisinger, and Rawia El-Batrawy. 2016. *Nutrition and economic development: Exploring Egypt's exceptionalism and the role of food subsidies*. Intl Food Policy Res Inst.

Gadenne, Lucie, Samuel Norris, Monica Singhal, and Sandip Sukhtankar. 2021. "In-kind transfers as insurance." Tech. rep., National Bureau of Economic Research.

Gazeaud, Jules. 2020. "Proxy Means Testing vulnerability to measurement errors?" *The Journal of Development Studies* 56 (11):2113–2133.

Gentilini, Ugo, Maddalena Honorati, and Ruslan Yemtsov. 2014. "The State of Social Safety Nets 2014." URL http://documents.worldbank.org/curated/en/302571468320707386/The-state-of-social-safety-nets-2014.

Gutner, Tamar. 2002. "The political economy of food subsidy reform: the case of Egypt." *Food Policy* 27 (5):455–476. URL https://www.sciencedirect.com/science/article/pii/S0306919202000490.

Hanna, Rema and Benjamin A Olken. 2018. "Universal basic incomes versus targeted transfers: Anti-poverty programs in developing countries." *Journal of Economic Perspectives* 32 (4):201–226.

Haushofer, Johannes, Paul Niehaus, Carlos Paramo, Edward Miguel, and Michael Walker. 2025. "Targeting impact versus deprivation." *American Economic Review* 115 (6):1936–1974.

Hidrobo, Melissa, John Hoddinott, Amber Peterman, Amy Margolies, and Vanessa Moreira. 2014. "Cash, food, or vouchers? Evidence from a randomized experiment in northern Ecuador." *Journal of development Economics* 107:144–156.

Jayne, Thomas S, John Strauss, Takashi Yamano, and Daniel Molla. 2002. "Targeting of food aid in rural Ethiopia: chronic need or inertia?" *Journal of Development Economics* 68 (2):247–288.

Jensen, Robert T and Nolan H Miller. 2008. "Giffen behavior and subsistence consumption." *American Economic Review* 98 (4):1553–1577.

Karlan, Dean and Bram Thuysbaert. 2019. "Targeting ultra-poor households in Honduras and Peru." *The World Bank Economic Review* 33 (1):63–94.

Kennedy, Gina, Terri Ballard, and Marie Claude Dop. 2013. *Guidelines for Measuring Household and Individual Dietary Diversity*. Food and Agriculture Organization of the United Nations. URL https://openknowledge.fao.org/handle/20.500.14283/i1983e.

Lain, Jonathan, Sharad Tandon, and Tara Vishwanath. 2024. "How Much Does the Food Insecurity Experience Scale Overlap with Poor Food Consumption and Monetary Poverty? Evidence from West Africa." *The World Bank Economic Review* 38 (2):422–442.

Lyons, Angela C, Alejandro Montoya Castano, Josephine Kass-Hanna, Yifang Zhang, and Aiman Soliman. 2025. "A machine learning approach to assessing multidimensional poverty and targeting assistance among forcibly displaced populations." *World Development* 192:107013.

McKenzie, David. 2012. "Beyond baseline and follow-up: The case for more T in experiments." *Journal of Development Economics* 99 (2):210–221.

Premand, Patrick and Pascale Schnitzer. 2021. "Efficiency, legitimacy, and impacts of targeting methods: Evidence from an experiment in Niger." *The World Bank Economic Review* 35 (4):892–920.

Reuters. 2023. "Egypt's Inflation Quickens to Record 38% in September." *Reuters* Available at: https://www.reuters.com/world/africa/egypts-inflation-quickens-record-380-september-2023-10-10/ (Accessed: October 30th, 2023).

Ruel, Marie T and Harold Alderman. 2013. "Nutrition-sensitive interventions and programmes: how can they help to accelerate progress in improving maternal and child nutrition?" *The lancet* 382 (9891):536–551.

Saint Ville, Arlette, June Yee Tsun Po, Akankasha Sen, Anh Bui, and Hugo Melgar-Quiñonez. 2019. "Food security and the Food Insecurity Experience Scale (FIES): ensuring progress by 2030." URL https://rdcu.be/dF3PF.

Stoeffler, Quentin, Bradford Mills, and Carlo Del Ninno. 2016. "Reaching the poor: Cash transfer program targeting in Cameroon." *World Development* 83:244–263.

Sumarto, Sudarno, Elan Satriawan, Benjamin A Olken, Abhijit Banerjee, Achmad Tohari, Vivi Alatas, and Rema Hanna. 2025. "Community Targeting at Scale." Tech. rep., National Bureau of Economic Research.

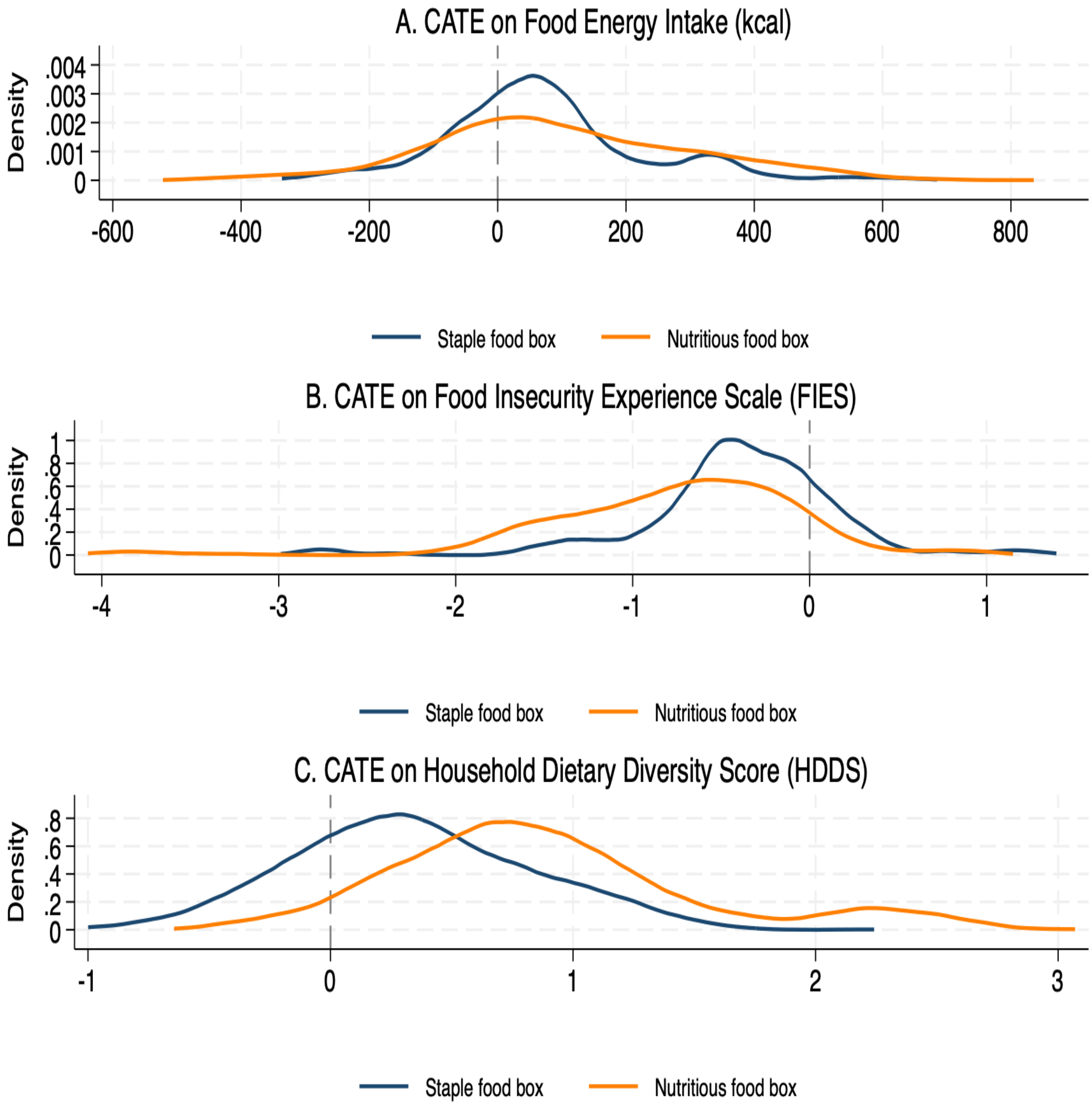
Tawfik, Afaf A, Emily T Hanna, and Awatif M Abdel-Maksoud. 2015. "Anemia and iron deficiency anemia in Egypt." *IOSR J Pharm* 5 (4):30–34.

Trachtman, Carly, Yudistira Permana, and Gumilang Sahadewo. 2022. "How much do our neighbors really know? The limits of community-based targeting." *University of California, Berkeley Working Paper* .

UNICEF, WHO, and the World Bank Group. 2021. "Levels and Trends in Child Malnutrition." .

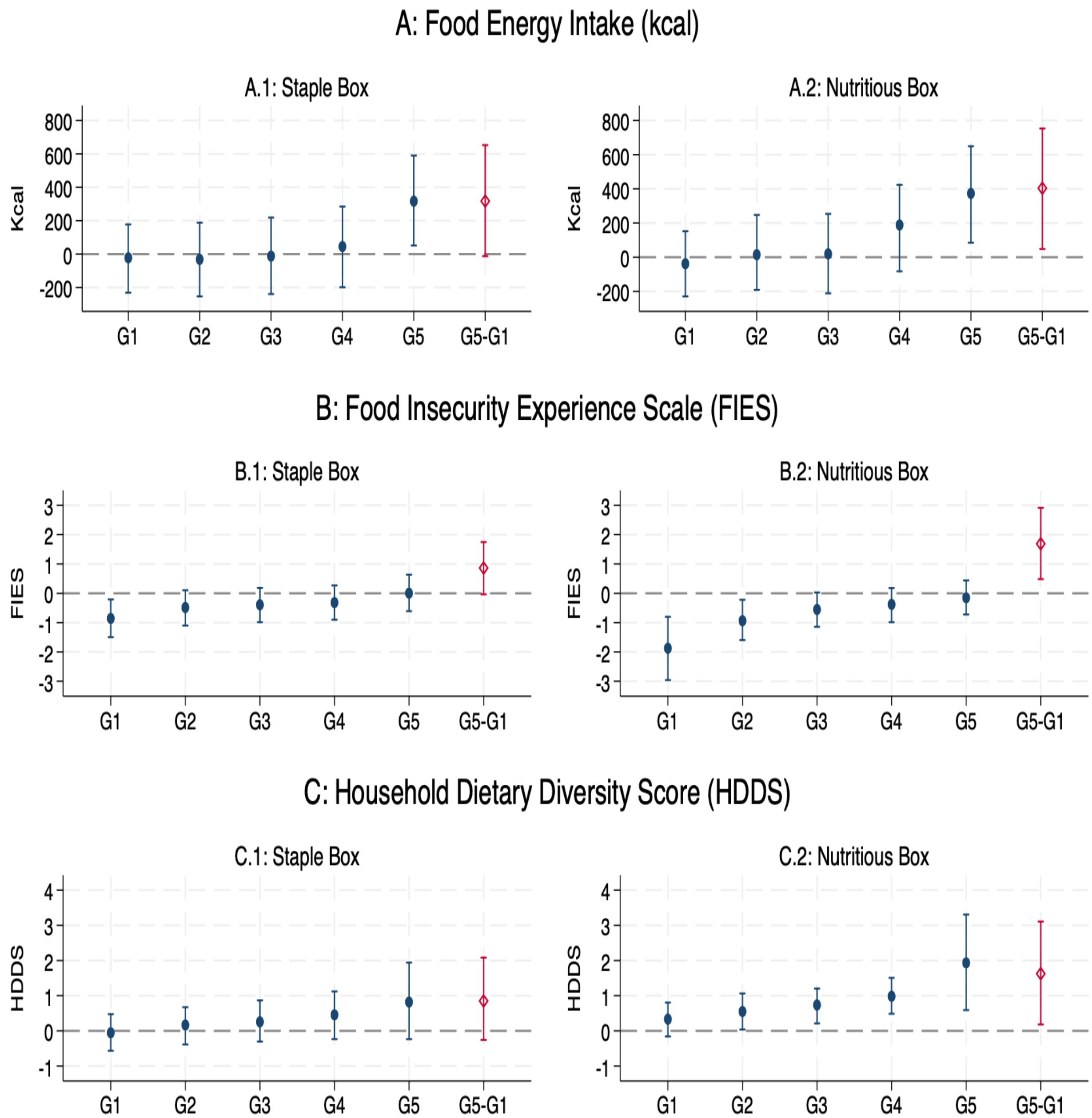
Figures and Tables

Figure 1: Kernel Density of Predicted Treatment Effects for the Main Outcomes by Treatment Group



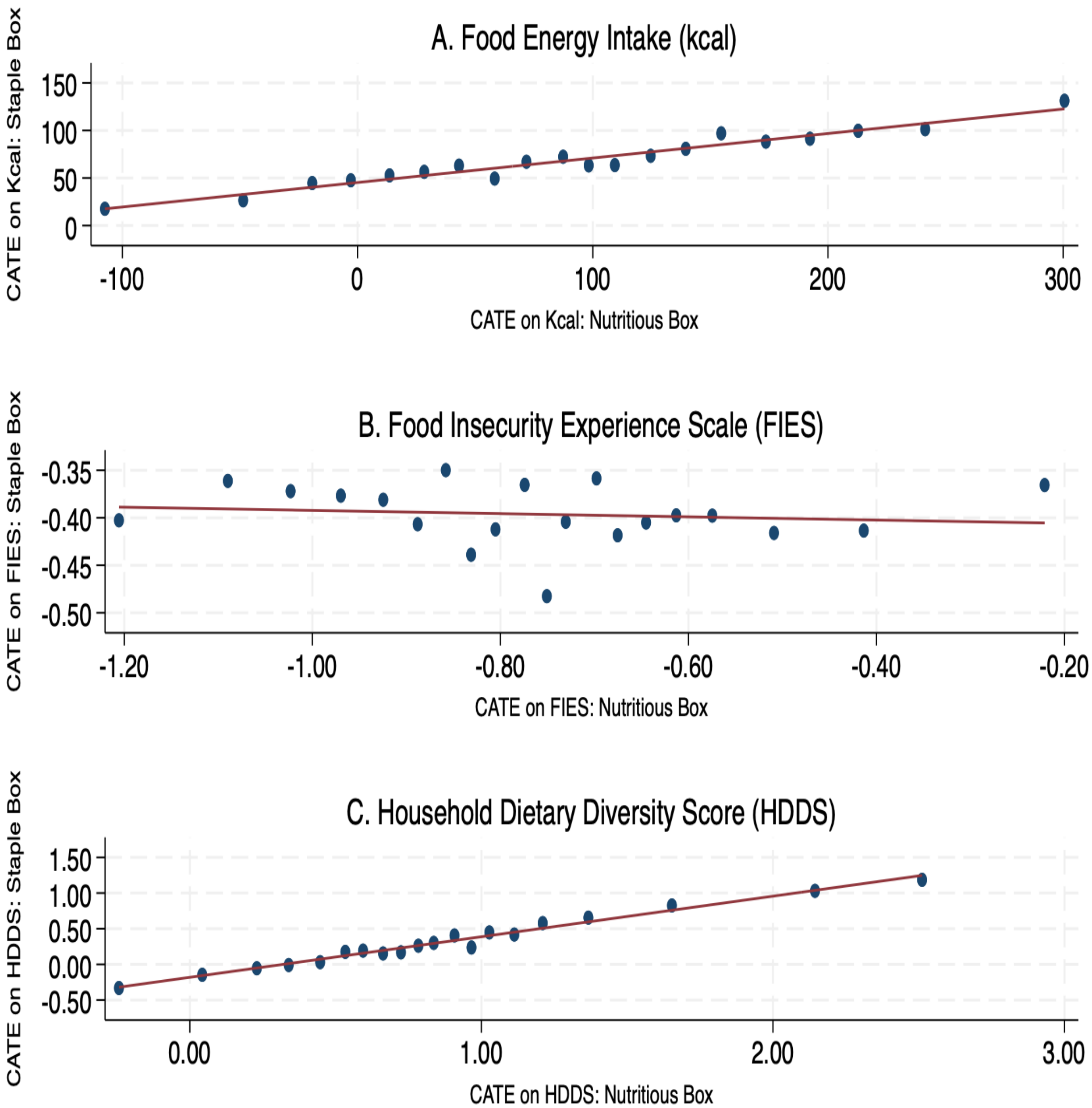
Notes: This figure presents Kernel density estimates of the distribution of the predicted conditional average treatment effect (CATE) using the best machine learning method (see Table 4 for the best method for each outcome). Panels A, B, and C, show the CATE on the outcomes: household head's food energy intake (kcal), household-level food insecurity experience scale (FIES), and household dietary diversity score (HDDS), respectively. The dark blue line shows the distribution of the CATE for the staple food box treatment. The orange line shows the distribution of the CATE for the nutritious food box treatment.

Figure 2: Sorted Group Average Treatment Effects (GATES) for the Main Outcomes by Treatment Group



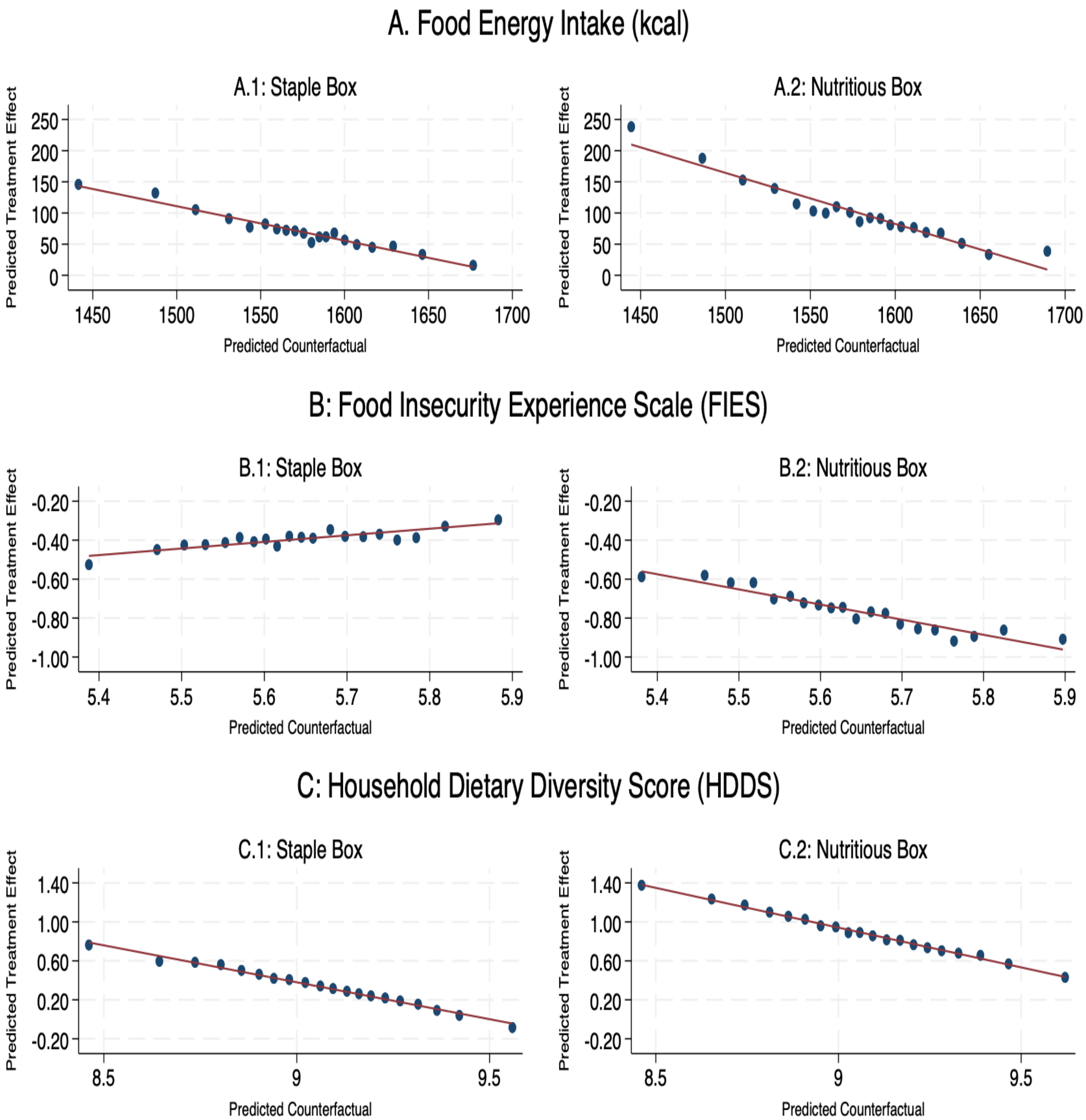
Notes: This figure summarizes results on estimating ordered group average treatment effects (GATES) for the main outcomes, as presented in Table 4. The outcome variables in panels A, B, and C are the household head's food energy intake (kcal), household-level food insecurity experience scale (FIES), and household dietary diversity score (HDDS), respectively. Confidence intervals are at the 90% level.

Figure 3: Relationship between the Predicted Impact of Staple Food Box and Nutritious Food Box



Notes: This figure presents binscatter plots on the correlation between the predicted conditional average treatment effect (CATE) for the staple food box and the nutritious food box for each of the three main outcome variables. The outcome variables in panels A, B, and C are the household head's food energy intake (kcal), household-level food insecurity experience scale (FIES), and household dietary diversity score (HDDS), respectively.

Figure 4: Relationship between Predicted Treatment Effect and Predicted Counterfactual



Notes: This figure presents binscatter plots on the correlation between the predicted conditional average treatment effect (CATE) and the predicted counterfactual as estimated by baseline conditional average (BCA). The outcome variables in panels A, B, and C are the household head's food energy intake (kcal), household-level food insecurity experience scale (FIES), and household dietary diversity score (HDDS), respectively. The treatment interventions in the sub-panels 1 and 2 are the staple food box and the nutritious food box treatments, respectively. Table 7 presents the results of this correlation analysis in details.

Table 1: Conceptual Framework: Targeting Criteria

Panel A: Staple Food Box

	Baseline Status	**Income Elasticity of Demand for Calories**	**Extra-/ Inframarginal Calories**	**Impact on Calorie Consumption**	**Targeting For Calorie Sufficiency**	**Targeting For Perceived Food Security**
A.1	Low calories	Low elasticity	Extramarginal	Possible high impact*	Possible high priority	Priority if **also** increases security
A.2	Low calories	High elasticity	Inframarginal	High impact	High Priority (Cash could replace)	Priority if **also** increases security
A.3	High calories	Low elasticity	Inframarginal	Probable low impact**	Probable low priority	Priority **only** if increases security
A.4	High calories	High elasticity	Inframarginal	High impact	Low priority (Risk over-consumption)	Priority **only** if increases security

Panel B: Nutritious Food Box

	Baseline Status	**Income Elasticity of Demand for Diverse Diet**	**Extra-/ Inframarginal Food Items**	**Impact on Dietary Diversity**	**Targeting For Dietary Diversity**	**Targeting For Perceived Food Security**
B.1	Low dietary diversity	Low elasticity	Extramarginal	Possible high impact*	Possible high priority	Priority if **also** increases security
B.2	Low dietary diversity	High elasticity	Inframarginal	High impact	High Priority (Cash could replace)	Priority if **also** increases security
B.3	Moderate dietary diversity	Low elasticity	Extramarginal	Possible high impact*	Possible high priority	Priority if **also** increases security
B.4	Moderate dietary diversity	High elasticity	Inframarginal	High impact	High priority (Cash could replace)	Priority if **also** increases security
B.5	High dietary diversity	Low elasticity	Inframarginal	No impact	Exclude	Unlikely to increase security

* Households with low demand for extramarginal food items may forgo consumption adjustments if side-selling or exchange is feasible. Alternatively, demand may shift as habits form, rendering the transfer inframarginal over time.

** Households with low demand for additional inframarginal food items may still consume them if food purchasing patterns are “sticky” or if the food delivery service successfully shifts preferences.

Table 2: Contents of Staple and Nutritious Food Boxes

Staple Food Box	**Nutritious Food Box**
Rice: 3 kilograms	Rice: 3 kilograms
Macaroni: 1 kilogram	Macaroni: 1 kilogram
Green or white beans: 500 grams	Green or white beans: 1 kilogram
Fava beans: 800 grams	Fava beans: 1 kilogram
Cheese (plant-based fat): 500 grams	Cheese (animal-based fat): 500 grams
Sugar: 1 kilogram	Wheat: 1 kilogram
Oil: 600 grams	Milk powder: 400 grams
Tomato paste: 400 grams	Dried fruits: 200 grams
Salt: 400 grams	Spices: 200 grams
	Red meat: 1 kilogram
	Yellow lentils: 1 kilogram
	Vermicelli: 800 grams
	Dates: 1 kilogram
	Molasses: 350 grams
	Starch: 100 grams
	Vanilla powder: 4 grams
	Cacao: 15 grams

Table 3: Impact on Food Energy Intake, Food Insecurity Experience (FIES), and Household Dietary Diversity (HDDS)

	(1) Calorie Intake (kcal)	(2) Calorie Intake (kcal % of RDA)	(3) Gap in Calorie Intake (kcal less than RDA)	(4) Food Insecurity Experience (0 - 8)	(5) Household Dietary Diversity (0 - 12)
Staple food box	67.58	3.32	-37.29	-0.33**	0.33*
	(50.31)	(2.47)	(31.81)	(0.14)	(0.19)
Nutritious food box	116.66**	5.73**	-53.64*	-0.68***	0.90***
	(55.46)	(2.72)	(31.58)	(0.16)	(0.18)
Baseline value of the outcome	0.15***	0.14***	0.13***	0.31***	0.25***
	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)
Strata FE	Yes	Yes	Yes	Yes	Yes
p-val: nutritious = staple	0.39	0.39	0.59	0.04	0.00
p-val: nutritious = 2 staple	0.84	0.84	0.70	0.41	0.93
R-squared	0.129	0.129	0.134	0.233	0.187
Control group mean	1580.63	77.62	551.17	5.66	9.06
Number of observations	2,258	2,258	2,258	2,277	2,277

Notes: The outcome variable in column (1) is food energy intake (kcal) for the female head of household that is estimated using a 24-hour recall survey. The outcome variable in column (2) is the household head's food energy intake as a percentage of recommended dietary allowance (RDA). The outcome variable in column (3) is the gap between the household head's actual food energy intake (kcal) and RDA. The outcome variable in column (4) is a household-level Food Insecurity Experience Scale (FIES), which is a standard measure of food insecurity developed and verified by FAO. The outcome variable in column (5) is the household dietary diversity score (HDDS), estimated using a seven-day recall of food groups consumed by all household members. Standard errors in parentheses are clustered at the community level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Heterogeneity in Treatment Effects on Food Energy Intake (Kcal), Food Insecurity Experience (FIES), and Household Dietary Diversity (HDDS)

	(1) Kcal (Staple Box)	(2) Kcal (Nutritious Box)	(3) FIES (Staple Box)	(4) FIES (Nutritious Box)	(5) HDDS (Staple Box)	(6) HDDS (Nutritious Box)
Panel A: Best Linear Predictor (BLP) Estimates						
ATE (β_1)	57.98	118.95	-0.40	-0.77	0.34	0.90
	(-53.29, 169.24)	(-4.63, 233.95)	(-0.68, -0.14)	(-1.11, -0.43)	(-0.04, 0.71)	(0.55, 1.25)
	[0.31]	[0.06]	[0.00]	[0.00]	[0.08]	[0.00]
HET (β_2)	0.53	0.51	0.49	0.86	0.45	0.73
	(0.03, 0.97)	(0.09, 0.90)	(0.11, 0.83)	(0.40, 1.28)	(-0.13, 1.06)	(0.07, 1.42)
	[0.03]	[0.02]	[0.01]	[0.00]	[0.13]	[0.03]
Best ML method	Elastic Net	Elastic Net	Elastic Net	Elastic Net	Elastic Net	Elastic Net
Panel B: Sorted Group Average Treatment Effects (GATES)						
Group 5 (Highest 20%)	316.69	373.08	0.00	-0.15	0.82	1.93
	(51.23, 590.13)	(84.56, 649.04)	(-0.61, 0.64)	(-0.72, 0.44)	(-0.24, 1.94)	(0.59, 3.30)
	[0.02]	[0.01]	[0.98]	[0.62]	[0.14]	[0.00]
Group 1 (Lowest 20%)	-22.20	-38.20	-0.86	-1.87	-0.05	0.33
	(-231.12, 177.59)	(-229.66, 151.74)	(-1.50, -0.21)	(-2.96, -0.80)	(-0.57, 0.47)	(-0.16, 0.80)
	[0.85]	[0.70]	[0.01]	[0.00]	[0.84]	[0.18]
Group 5 - Group 1	317.28	403.71	0.86	1.69	0.85	1.62
	(-12.41, 652.47)	(47.71, 753.00)	(-0.04, 1.75)	(0.48, 2.91)	(-0.26, 2.08)	(0.18, 3.11)
	[0.06]	[0.03]	[0.06]	[0.01]	[0.14]	[0.03]
Best ML method	Elastic Net	Elastic Net	Elastic Net	Elastic Net	Elastic Net	Elastic Net

Notes: This table presents results on a) estimating the best linear predictor (BLP) of the conditional average treatment effects (CATE); and b) estimating ordered group average treatment effects (GATES) following the approach of Chernozhukov et al. (2023). The outcome variable in columns 1-2 is food energy intake (kcal). In columns 3-4, the outcome variable is the food insecurity scale (FIES), and in columns 5-6, it is the household dietary diversity score (HDDS). Results from the best machine learning method (i.e., the algorithm with the highest prediction power) for each outcome are reported. Reported values are the median of predicted treatment effects over 100 splits. 90% Confidence intervals are in parentheses. P-values in square brackets.

Table 5: Heterogeneity in Treatment Effects: Impacts of the Nutritious Food Box Relative to the Staple Food Box

	(1) Kcal	(2) FIES	(3) HDDS
Panel A: Best Linear Predictor (BLP) Estimates			
ATE (β_1)	42.22	-0.35	0.56
	(-80.38, 173.23)	(-0.70, 0.02)	(0.29, 0.84)
	[0.48]	[0.04]	[0.00]
HET (β_2)	0.28	0.64	0.62
	(-0.16, 0.71)	(0.38, 0.91)	(0.13, 1.13)
	[0.18]	[0.00]	[0.01]
Best ML method			
Panel B: Sorted Group Average Treatment Effects (GATES)			
Group 2 (Highest 50%)	99.83	0.23	0.89
	(-61.39, 260.34)	(-0.15, 0.63)	(0.46, 1.30)
	[0.21]	[0.23]	[0.00]
Group 1 (Lowest 50%)	-7.22	-0.97	0.24
	(-167.24, 145.01)	(-1.52, -0.40)	(-0.08, 0.58)
	[0.92]	[0.00]	[0.14]
Group 2 - Group 1	102.67	1.17	0.64
	(-121.07, 321.40)	(0.54, 1.80)	(0.13, 1.16)
	[0.35]	[0.00]	[0.02]
Best ML method			

Notes: This table examines the heterogeneity in impacts of the nutritious food box relative to the staple food box intervention. It reports (a) the best linear predictor (BLP) of conditional average treatment effects (CATE), and (b) group average treatment effects (GATES) estimated following Chernozhukov et al. (2023). For GATES, we use two groups to reflect a budget-neutral comparison: the nutritious box costs twice as much as the staple box, so the same budget would cover half as many households. Outcomes are: (1) food energy intake (kcal), (2) food insecurity scale (FIES), and (3) household dietary diversity score (HDDS). Results are based on the best-performing machine learning algorithm (i.e., the algorithm with the highest prediction power) for each outcome. Reported values are the median treatment effect across 100 random splits. 90% confidence intervals are in parentheses; p-values are in square brackets.

Table 6: Impact on FIES and HDDS by Quantile of Predicted Impact on Calorie Intake

	Staple Food Box		Nutritious Food Box	
	(1)	(2)	(3)	(4)
	FIES	HDDS	FIES	HDDS
Group 1 (0 - 20%)	-0.33	0.42*	-0.15	0.65***
	(0.36)	(0.25)	(0.25)	(0.17)
Group 2 (20 - 40%)	-0.12	0.47**	-0.62**	0.64***
	(0.20)	(0.19)	(0.24)	(0.19)
Group 3 (40 - 60%)	-0.25	-0.13	-0.76***	0.76***
	(0.21)	(0.23)	(0.26)	(0.19)
Group 4 (60 - 80%)	-0.44*	-0.00	-1.21***	0.93***
	(0.24)	(0.20)	(0.31)	(0.27)
Group 5 (80 - 100%)	-0.82***	0.89	-1.05***	1.57***
	(0.31)	(0.71)	(0.37)	(0.58)
p-value all interactions equal	0.03	0.35	0.00	0.00
p-value treatxG1 = treatxG5	0.30	0.53	0.05	0.13
Mean for control in G1	5.69	9.20	5.25	9.34
Mean for control in G2	5.48	9.00	5.67	8.97
Mean for control in G3	5.54	9.26	5.51	9.13
Mean for control in G4	5.66	9.51	5.76	9.18
Mean for control in G5	5.88	8.43	6.10	8.71
Number of observations	1,503	1,503	1,522	1,522

Notes: This table presents estimates of treatment effects on the food insecurity experience (FIES) and household dietary diversity (HDDS), by quantile of the predicted impacts on calorie intake. Estimation follows a specification similar to equation 4, using predicted treatment effects on calorie intake to define quantiles, while replacing the dependent variable with either FIES or HDDS. At the bottom of the table, we report tests of equality in treatment effects across all quantiles, as well as between the bottom (G1) and top (G5) quantiles. Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7: Correlation Between Predicted Treatment Effect and Predicted Counterfactual

	(1) CATE: Kcal Intake (Staple Box)	(2) CATE: Kcal Intake (Nutritious Box)	(3) CATE: FIES (Staple Box)	(4) CATE: FIES (Nutritious Box)	(5) CATE: HDDS (Staple Box)	(6) CATE: HDDS (Nutritious Box)
BCA: Kcal Intake	-0.55***	-0.82***				
	(0.02)	(0.05)				
BCA: FIES			0.34***	-0.78***		
			(0.05)	(0.05)		
BCA: HDDS					-0.76***	-0.82***
					(0.01)	(0.01)
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.928	0.797	0.868	0.903	0.936	0.968
CATE Mean	70.63	100.71	-0.39	-0.76	0.34	0.88
BCA Mean	1573.02	1577.77	5.64	5.64	9.06	9.07
Number of observations	1,503	1,522	1,503	1,522	1,503	1,522

Notes: In each column, the dependent variable is the predicted conditional average treatment effect (CATE) and the independent variable is the predicted baseline conditional average (BCA), which represents an estimate of the predicted counterfactual in the absence of treatment. CATE and BCA are estimated using the machine learning algorithm Elastic Net. Columns (1) and (2) focus on the outcome variable: household head's food energy intake (kcal) for the staple food box and the nutritious food box treatments, respectively. Columns (3) and (4) focus on the outcome variable: household-level food insecurity experience scale (FIES) for the staple food box and the nutritious food box treatments, respectively. Columns (5) and (6) focus on the outcome variable: household dietary diversity score (HDDS) for the staple food box and the nutritious food box treatments, respectively. Standard errors in parentheses are clustered at the community level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 8: Difference Between Characteristics of Households With the Highest vs the Lowest Predicted Treatment Effects on Food Energy Intake (kcal)

	Staple Box			Nutritious Box		
	(1)	(2)	(3)	(4)	(5)	(6)
	20% Highest	20% Lowest	Difference	20% Highest	20% Lowest	Difference
Household Head Characteristics						
Age: < 40 years old	0.52	0.56	-0.04 [0.56]	0.47	0.59	-0.13 [0.02]
Widowed/ divorced/ separated	0.89	0.90	-0.01 [0.85]	0.85	0.93	-0.08 [0.03]
Education: never attended school	0.28	0.36	-0.09 [0.09]	0.39	0.33	0.05 [0.31]
Education: primary or preparatory	0.21	0.22	-0.01 [0.87]	0.29	0.17	0.12 [0.01]
Education: secondary & above	0.49	0.38	0.10 [0.06]	0.28	0.47	-0.19 [0.00]
Income from work	0.27	0.32	-0.05 [0.31]	0.28	0.28	-0.01 [0.87]
Household Characteristics						
Household size	3.66	3.50	0.17 [0.23]	3.67	3.55	0.10 [0.49]
Dependency ratio	1.28	1.39	-0.10 [0.43]	1.24	1.39	-0.17 [0.18]
Number of members enrolled in education	1.57	1.61	0.01 [0.87]	1.61	1.64	-0.04 [0.78]
Number of members never attended school	0.55	0.54	0.03 [0.76]	0.74	0.53	0.20 [0.03]
Number of unemployed members	1.34	1.15	0.19 [0.06]	1.32	1.28	0.04 [0.65]
Number of members with no pension	0.54	0.56	-0.03 [0.76]	0.64	0.47	0.17 [0.05]
Receives Takaful or Karama transfers	0.33	0.26	0.07 [0.15]	0.35	0.28	0.06 [0.20]
Has a ration card	0.87	0.69	0.19 [0.00]	0.86	0.72	0.14 [0.00]
Receives informal financial assistance	0.84	0.80	0.03 [0.42]	0.80	0.79	0.00 [0.98]
Dwelling Characteristics						
Number of members per room	1.40	1.36	0.03 [0.63]	1.47	1.36	0.12 [0.16]
Rural house	0.15	0.26	-0.10 [0.03]	0.29	0.15	0.14 [0.00]

Continued on next page

Table 8: Difference Between Characteristics of Households With the Highest vs the Lowest Predicted Treatment Effects on Food Energy Intake (kcal) (Continued)

	Staple Box			**Nutritious Box**		
	(1)	(2)	(3)	(4)	(5)	(6)
	20% Highest	20% Lowest	Difference	20% Highest	20% Lowest	Difference
Rented housing unit	0.13	0.28	-0.15	0.10	0.33	-0.23
			[0.00]			[0.00]
Roof material: cement	0.80	0.74	0.06	0.77	0.73	0.06
			[0.19]			[0.22]
Roof material: wood	0.08	0.20	-0.12	0.10	0.20	-0.09
			[0.00]			[0.03]
No private bathroom	0.15	0.06	0.09	0.19	0.06	0.12
			[0.01]			[0.00]
Modern toilet with flush	0.58	0.62	-0.01	0.45	0.66	-0.22
			[0.90]			[0.00]
Monthly expenditure on electricity	120.96	124.68	-3.65	118.46	136.19	-21.43
			[0.75]			[0.07]
Asset Ownership						
Regular washing machine	0.42	0.50	-0.08	0.50	0.39	0.11
			[0.17]			[0.03]
Half-automatic washing machine	0.37	0.37	-0.01	0.29	0.52	-0.23
			[0.81]			[0.00]
Full-automatic washing machine	0.15	0.04	0.11	0.14	0.04	0.11
			[0.00]			[0.00]
Internet access	0.18	0.10	0.09	0.10	0.12	-0.02
			[0.02]			[0.56]
Regular mobile phone	0.55	0.58	-0.03	0.65	0.49	0.17
			[0.60]			[0.00]
Smart phone	0.52	0.53	0.00	0.42	0.61	-0.18
			[0.91]			[0.00]

Notes: This table presents results from classification analysis (CLAN) for the outcome variable food energy intake (kcal) following the approach of Chernozhukov et al. (2023). That is, columns (1) and (4) show the median of each baseline covariate for the households in the quantile with the highest predicted group average treatment effects (δ_5). Columns (2) and (5) show the median of each baseline covariate for the households in the quantile with the lowest predicted group average treatment effects (δ_1). Columns (3) and (6) show the differences between the medians ($\delta_5 - \delta_1$). Columns 1-3 present this comparison for the staple food box treatment, while columns 4-6 present this comparison for the nutritious food box treatment. P-values of the differences are in brackets.

Table 9: Difference Between Characteristics of Households With the Highest vs the Lowest Predicted Treatment Effects on Food Insecuriety Experience Scale (FIES)

	Staple Box			Nutritious Box		
	(1)	(2)	(3)	(4)	(5)	(6)
	20% Highest	20% Lowest	Difference	20% Highest	20% Lowest	Difference
Household Head Characteristics						
Age < 40	0.57	0.49	0.07	0.59	0.47	0.13
			[0.20]			[0.02]
Widowed/ divorced/ separated	0.90	0.93	-0.03	0.92	0.87	0.05
			[0.36]			[0.13]
Education: never attended school	0.28	0.34	-0.06	0.50	0.24	0.25
			[0.28]			[0.00]
Education: primary or preparatory	0.33	0.17	0.17	0.24	0.16	0.08
			[0.00]			[0.07]
Education: secondary & above	0.36	0.47	-0.10	0.22	0.55	-0.34
			[0.08]			[0.00]
Income from work	0.32	0.21	0.11	0.24	0.37	-0.15
			[0.03]			[0.01]
Household Characteristics						
Household size	3.54	3.56	-0.04	3.73	3.60	0.12
			[0.77]			[0.36]
Dependency ratio	1.55	1.17	0.38	1.44	1.26	0.18
			[0.00]			[0.14]
Number of members enrolled in education	1.61	1.60	0.00	1.86	1.54	0.28
			[0.87]			[0.02]
Number of members never attended school	0.48	0.53	-0.04	0.72	0.59	0.13
			[0.59]			[0.15]
Number of unemployed members	1.14	1.41	-0.30	1.34	1.25	0.10
			[0.01]			[0.35]
Number of members with no pension	0.54	0.48	0.05	0.46	0.61	-0.16
			[0.51]			[0.06]
Receives Takaful or Karama transfers	0.27	0.28	-0.02	0.30	0.30	-0.00
			[0.80]			[0.93]
Has a ration card	0.81	0.80	0.01	0.81	0.82	-0.03
			[0.89]			[0.54]
Receives informal financial assistance	0.86	0.80	0.05	0.80	0.80	-0.00
			[0.17]			[0.93]
Dwelling Characteristics						
Number of members per room	1.37	1.40	-0.00	1.48	1.37	0.11
			[0.96]			[0.17]
Rural house	0.17	0.22	-0.04	0.26	0.15	0.11
			[0.35]			[0.01]

Continued on next page

Table 9: Difference Between Characteristics of Households With the Highest vs the Lowest Predicted Treatment Effects on Food Insecuriety Experience Scale (FIES) (Continued)

	Staple Box			Nutritious Box		
	(1)	(2)	(3)	(4)	(5)	(6)
	20% Highest	20% Lowest	Difference	20% Highest	20% Lowest	Difference
Rented housing unit	0.22	0.18	0.04	0.17	0.23	-0.05
			[0.39]			[0.22]
Roof material: cement	0.83	0.73	0.10	0.59	0.88	-0.29
			[0.03]			[0.00]
Roof material: wood	0.14	0.12	0.02	0.34	0.03	0.31
			[0.61]			[0.00]
No private bathroom	0.07	0.11	-0.05	0.18	0.08	0.10
			[0.20]			[0.01]
Modern toilet with flush	0.64	0.59	0.05	0.42	0.71	-0.29
			[0.41]			[0.00]
Monthly expenditure on electricity	126.51	126.32	0.23	113.42	139.18	-27.57
			[0.95]			[0.02]
Asset Ownership						
Regular washing machine	0.55	0.38	0.16	0.50	0.39	0.12
			[0.00]			[0.02]
Half-automatic washing machine	0.37	0.36	0.03	0.33	0.46	-0.13
			[0.55]			[0.02]
Full-automatic washing machine	0.04	0.20	-0.16	0.07	0.11	-0.04
			[0.00]			[0.23]
Internet access	0.14	0.12	0.02	0.09	0.13	-0.05
			[0.61]			[0.21]
Regular mobile phone	0.56	0.58	-0.02	0.57	0.58	-0.01
			[0.82]			[0.86]
Smart phone	0.56	0.51	0.05	0.49	0.50	-0.00
			[0.36]			[0.96]

Notes: This table presents results from classification analysis (CLAN) for the outcome variable food insecurity experience scale (FIES) following the approach of Chernozhukov et al. (2023). That is, columns (1) and (4) show the median of each baseline covariate for the households in the quantile with the highest predicted group average treatment effects (δ_5). Columns (2) and (5) show the median of each baseline covariate for the households in the quantile with the lowest predicted group average treatment effects (δ_1). Columns (3) and (6) show the differences between the medians ($\delta_5 - \delta_1$). Columns 1-3 present this comparison for the staple food box treatment, while columns 4-6 present this comparison for the nutritious food box treatment. P-values of the differences are in brackets. Given that a ***negative*** treatment effect on food **insecurity** experience is a desirable outcome, the results in this table should be interpreted with caution. In this case, columns (2) and (5) show the covariates of the most impacted households, while columns (1) and (4) show the covariates of the least impacted households.

Table 10: Difference Between Characteristics of Households With the Highest vs the Lowest Predicted Treatment Effects on Household Dietary Diversity Score (HDDS)

	Staple Box			Nutritious Box		
	(1)	(2)	(3)	(4)	(5)	(6)
	20% Highest	20% Lowest	Difference	20% Highest	20% Lowest	Difference
Household Head Characteristics						
Age < 40	0.53	0.50	0.02	0.55	0.49	0.04
			[0.73]			[0.46]
Widowed/ divorced/ separated	0.90	0.90	-0.01	0.89	0.89	-0.01
			[0.85]			[0.72]
Education: never attended school	0.27	0.40	-0.13	0.26	0.46	-0.19
			[0.02]			[0.00]
Education: primary or preparatory	0.17	0.29	-0.13	0.19	0.26	-0.07
			[0.01]			[0.12]
Education: secondary & above	0.55	0.27	0.27	0.53	0.24	0.29
			[0.00]			[0.00]
Income from work	0.30	0.27	0.03	0.35	0.28	0.06
			[0.53]			[0.24]
Household Characteristics						
Household size	3.28	3.93	-0.67	3.43	3.87	-0.47
			[0.00]			[0.00]
Dependency ratio	1.22	1.46	-0.27	1.25	1.36	-0.12
			[0.05]			[0.34]
Number of members enrolled in education	1.27	1.91	-0.63	1.41	1.93	-0.50
			[0.00]			[0.00]
Number of members never attended school	0.54	0.62	-0.08	0.58	0.69	-0.09
			[0.35]			[0.32]
Number of unemployed members	1.21	1.37	-0.18	1.22	1.41	-0.20
			[0.09]			[0.07]
Number of members with no pension	0.52	0.57	-0.05	0.57	0.56	0.00
			[0.62]			[0.92]
Receives Takaful or Karama transfers	0.29	0.28	0.01	0.30	0.32	-0.01
			[0.80]			[0.78]
Has a ration card	0.71	0.88	-0.18	0.72	0.90	-0.19
			[0.00]			[0.00]
Receives informal financial assistance	0.78	0.82	-0.05	0.81	0.75	0.05
			[0.28]			[0.27]
Dwelling Characteristics						
Number of members per room	1.30	1.52	-0.19	1.37	1.50	-0.12
			[0.02]			[0.18]
Rural house	0.18	0.21	-0.03	0.18	0.22	-0.04
			[0.47]			[0.35]

Continued on next page

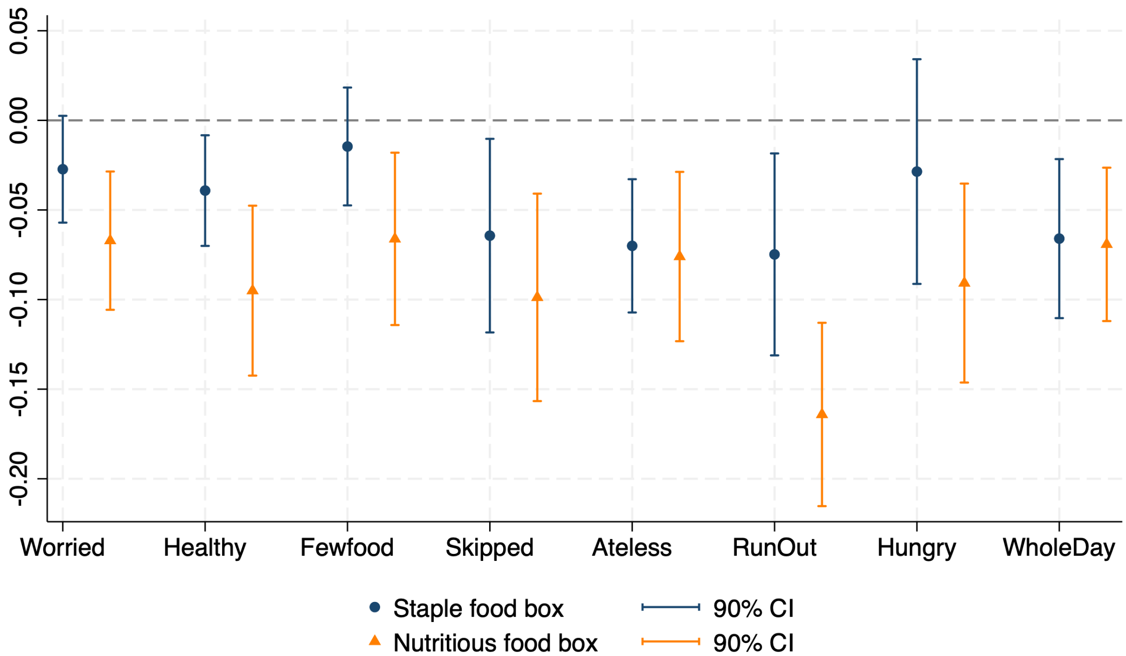
Table 10: Difference Between Characteristics of Households With the Highest vs the Lowest Predicted Treatment Effects on Household Dietary Diversity Score (HDDS) (Continued)

	Staple Box			Nutritious Box		
	(1)	(2)	(3)	(4)	(5)	(6)
	20% Highest	20% Lowest	Difference	20% Highest	20% Lowest	Difference
Rented housing unit	0.28	0.12	0.15	0.29	0.10	0.19
			[0.00]			[0.00]
Roof material: cement	0.79	0.77	0.03	0.79	0.59	0.21
			[0.57]			[0.00]
Roof material: wood	0.07	0.19	-0.12	0.04	0.34	-0.31
			[0.00]			[0.00]
No private bathroom	0.14	0.07	0.07	0.14	0.09	0.05
			[0.05]			[0.19]
Modern toilet with flush	0.62	0.62	0.01	0.64	0.49	0.15
			[0.82]			[0.01]
Monthly expenditure on electricity	106.39	138.04	-29.02	122.95	134.91	-13.37
			[0.01]			[0.30]
Asset Ownership						
Regular washing machine	0.39	0.50	-0.13	0.37	0.54	-0.16
			[0.03]			[0.00]
Half-automatic washing machine	0.40	0.40	0.01	0.46	0.35	0.12
			[0.91]			[0.02]
Full-automatic washing machine	0.12	0.07	0.06	0.10	0.07	0.04
			[0.06]			[0.27]
Internet access	0.18	0.09	0.09	0.15	0.07	0.06
			[0.02]			[0.04]
Regular mobile phone	0.60	0.55	0.05	0.54	0.62	-0.08
			[0.36]			[0.15]
Smart phone	0.46	0.59	-0.13	0.51	0.49	0.02
			[0.03]			[0.69]

Notes: This table presents results from classification analysis (CLAN) for the outcome variable household dietary diversity score (HDDS) following the approach of Chernozhukov et al. (2023). That is, columns (1) and (4) show the median of each baseline covariate for the households in the quantile with the highest predicted group average treatment effects (δ_5). Columns (2) and (5) show the median of each baseline covariate for the households in the quantile with the lowest predicted group average treatment effects (δ_1). Columns (3) and (6) show the differences between the medians ($\delta_5 - \delta_1$). Columns 1-3 present this comparison for the staple food box treatment, while columns 4-6 present this comparison for the nutritious food box treatment. P-values of the differences are in brackets.

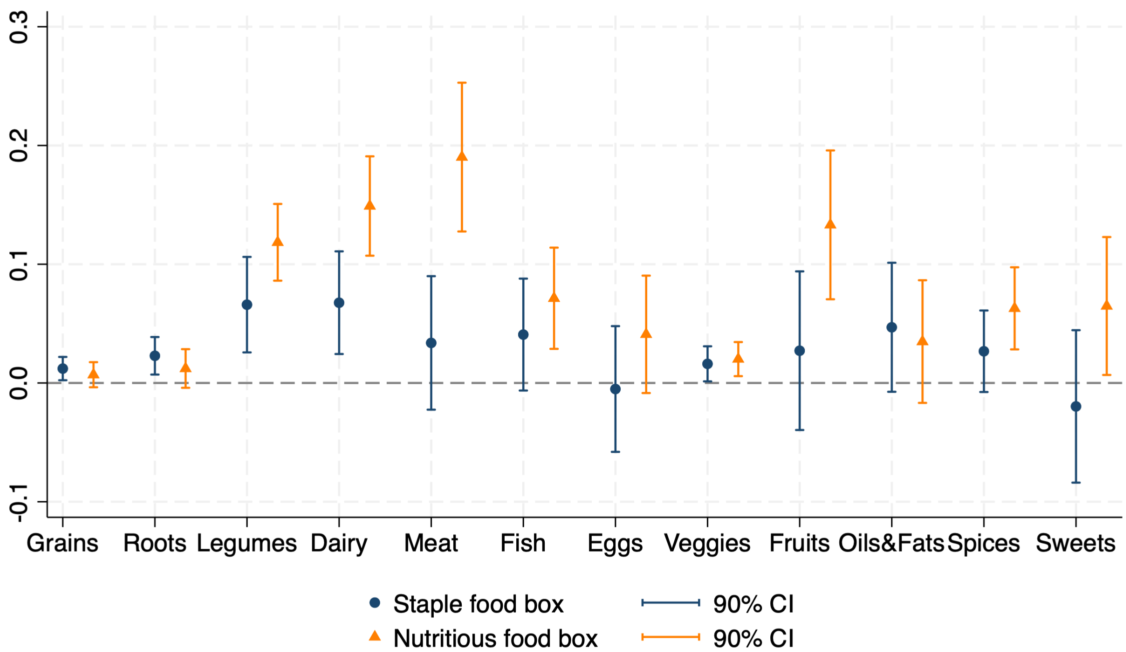
A Additional Figures and Tables

Figure A.1: Impact on Components of Food Insecurity Experience Scale (FIES)



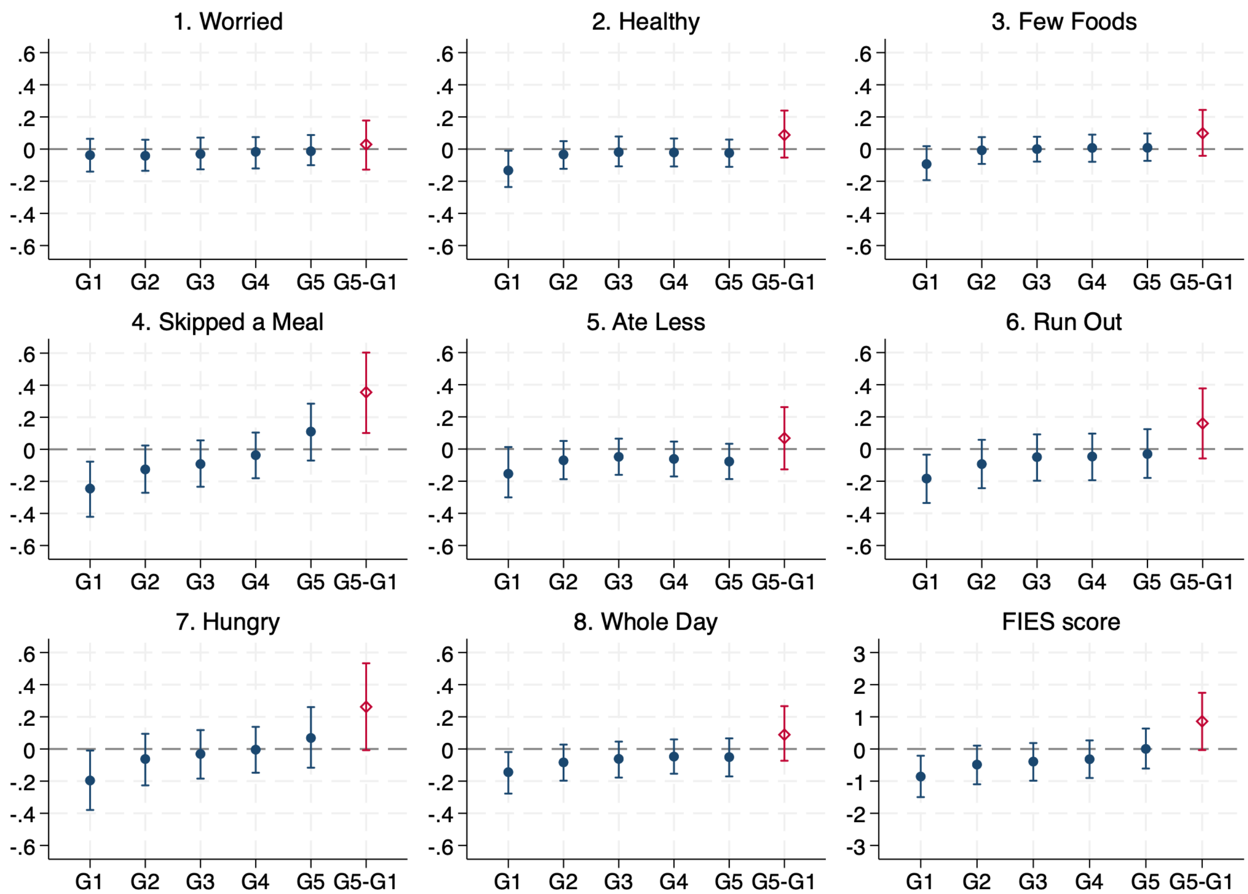
Notes: This figure shows the treatment effects on the household-level components of the household-level food insecurity experience scale (FIES). The dark blue lines are for the treatment effects of the staple food box, and the orange lines are for the treatment effects of the nutritious food box. Confidence intervals are at the 90% level. Table A.1 presents these results in details.

Figure A.2: Impact on Components of Household Dietary Diversity Score (HDDS)



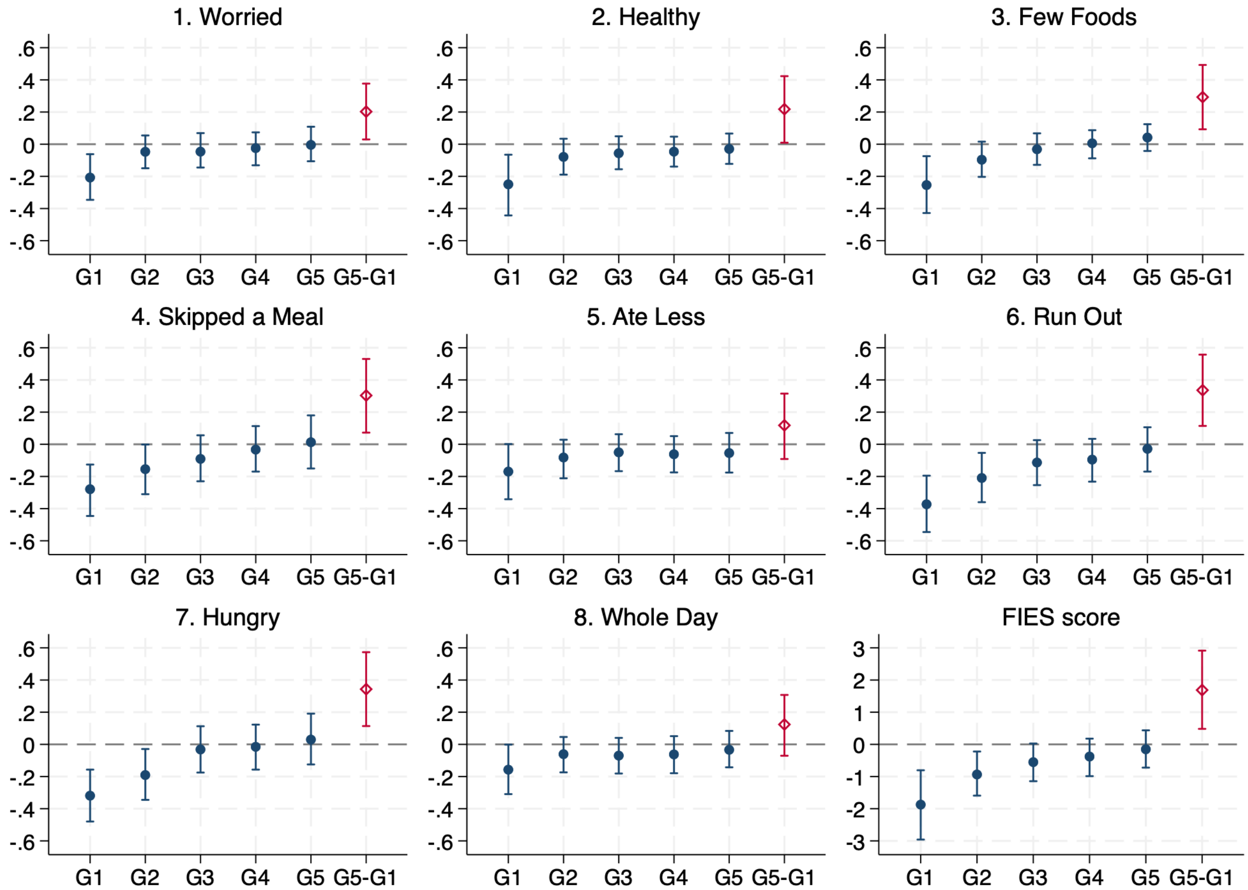
Notes: This figure shows the treatment effects on the components of the household dietary diversity score (HDDS). The dark blue lines are for the treatment effects of the staple food box, and the orange lines are for the treatment effects of the nutritious food box. Confidence intervals are at the 90% level. Table A.2 presents these results in details.

Figure A.3: Sorted Group Average Treatment Effects (GATES) for the Components of the FIES Score: **Staple** Food Box Treatment



Notes: This figure shows results on the estimated group average treatment effects (GATES) of the staple food box on the components of the household-level food insecurity experience scale (FIES). The outcome in panel (1) is an indicator for being worried about not having enough food due to lack of resources. The outcome in panel (2) is an indicator for not being able to eat healthy food due to lack of resources. The outcome in panel (3) is an indicator for eating only a few kinds of foods due to lack of resources. The outcome in panel (4) is an indicator for skipping a meal due to lack of resources. The outcome in panel (5) is an indicator for eating less due to lack of resources. The outcome in panel (6) is an indicator for running out of food due to lack of resources. The outcome in panel (7) is an indicator for being hungry and not eating due to lack of resources. The outcome in panel (8) is an indicator for spending a whole day without eating due to lack of resources. The final panel shows the GATES estimates for the overall FIES score. Confidence intervals are at the 90% level.

Figure A.4: Sorted Group Average Treatment Effects (GATES) for the Components of the FIES Score: **Nutritious** Food Box Treatment



Notes: This figure shows results on the estimated group average treatment effects (GATES) of the nutritious food box on the components of the household-level food insecurity experience scale (FIES). The outcome in panel (1) is an indicator for being worried about not having enough food due to lack of resources. The outcome in panel (2) is an indicator for not being able to eat healthy food due to lack of resources. The outcome in panel (3) is an indicator for eating only a few kinds of foods due to lack of resources. The outcome in panel (4) is an indicator for skipping a meal due to lack of resources. The outcome in panel (5) is an indicator for eating less due to lack of resources. The outcome in panel (6) is an indicator for running out of food due to lack of resources. The outcome in panel (7) is an indicator for being hungry and not eating due to lack of resources. The outcome in panel (8) is an indicator for spending a whole day without eating due to lack of resources. The final panel shows the GATES estimates for the overall FIES score. Confidence intervals are at the 90% level.

Table A.1: Impact on Components of Food Insecurity Experience Score (FIES)

	(1) Worried	(2) Healthy	(3) Fewfood	(4) Skipped	(5) Ateless	(6) RunOut	(7) Hungry	(8) WholeDay
Staple food box	-0.03	-0.04**	-0.01	-0.06*	-0.07***	-0.07**	-0.03	-0.07**
	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.03)	(0.04)	(0.03)
Nutritious food box	-0.07***	-0.10***	-0.07**	-0.10***	-0.08***	-0.16***	-0.09***	-0.07***
	(0.02)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
Baseline value of the outcome	0.17***	0.12***	0.11***	0.20***	0.17***	0.23***	0.22***	0.12***
	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
p-val: staple = nutritious	0.07	0.06	0.08	0.37	0.85	0.01	0.09	0.87
R-squared	0.104	0.087	0.082	0.124	0.097	0.194	0.177	0.115
Control village's mean	0.90	0.92	0.92	0.71	0.87	0.70	0.46	0.21
Number of observations	2,265	2,255	2,260	2,258	2,254	2,242	2,248	2,260

Notes: This table shows the impact on components of the standard food insecurity experience scale (FIES) as measured with reference to the respondent's experience of the past week. The outcome in column (1) is an indicator for being worried about not having enough food due to lack of resources. The outcome in column (2) is an indicator for not being able to eat healthy food due to lack of resources. The outcome in column (3) is an indicator for eating only a few kinds of foods due to lack of resources. The outcome in column (4) is an indicator for skipping a meal due to lack of resources. The outcome in column (5) is an indicator for eating less due to lack of resources. The outcome in column (6) is an indicator for running out of food due to lack of resources. The outcome in column (7) is an indicator for being hungry and not eating due to lack of resources. The outcome in column (8) is an indicator for spending a whole day without eating due to lack of resources. Standard errors in parentheses are clustered at the community level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The results shown in this table are summarized in Figure A.1.

Table A.2: Impact on Dietary Diversity by Food Group

	(1) Grains	(2) Roots	(3) Legumes & Seeds	(4) Milk & Dairy	(5) Meat	(6) Fish	(7) Eggs	(8) Fruits	(9) Veggies	(10) Oils & Fats	(11) Spices	(12) Sweets
Staple food box	0.01**	0.02**	0.07***	0.07**	0.03	0.04	-0.01	0.03	0.02*	0.05	0.03	-0.02
	(0.01)	(0.01)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.01)	(0.03)	(0.02)	(0.04)
Nutritious food box	0.01	0.01	0.12***	0.15***	0.19***	0.07***	0.04	0.13***	0.02**	0.03	0.06***	0.06*
	(0.01)	(0.01)	(0.02)	(0.03)	(0.04)	(0.03)	(0.03)	(0.04)	(0.01)	(0.03)	(0.02)	(0.04)
Baseline value of the outcome	0.01	0.04**	0.06***	0.13***	0.21***	0.16***	0.18***	0.09***	0.03	0.03	0.05	0.09***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.03)	(0.04)	(0.04)	(0.03)
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
p-val: staple = nutritious	0.17	0.12	0.01	0.00	0.00	0.23	0.12	0.01	0.40	0.56	0.03	0.04
R-squared	0.048	0.026	0.085	0.096	0.129	0.266	0.079	0.083	0.030	0.218	0.176	0.218
Control village's mean	0.99	0.96	0.84	0.73	0.51	0.29	0.53	0.65	0.97	0.92	0.91	0.77
Number of observations	2,276	2,277	2,275	2,277	2,277	2,277	2,276	2,269	2,275	2,276	2,266	2,266

Notes: This table shows the impact on components of the standard household dietary diversity score (HDDS) as measured with reference to food consumption over the past week. The outcome variables are as follows: column (1) foods made from grains, column (2) white roots and tubers, column (3) legumes nuts and seeds, column (4) milk and milk products, column (5) meat and organ meat, column (6) fish and seafood , column (7) eggs, column (8) fruits, column (9) vegetables, column (10) oil and fats, column (11) spices condiments and beverages, and column (12) sweets and sweetened beverages. Standard errors in parentheses are clustered at the community level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The results shown in this table are summarized in Figure A.2.

Table A.3: Impact on Food Expenditure (EGP/ week) by Food Group

	(1) Meat, poultry and fish	(2) Fruits	(3) Vegetables	(4) Legumes	(5) Eggs	(6) Dairy products	(7) Bread, flour, rice and pasta	(8) Ready-to-eat foods	(9) Snacks, juice, soda, and sweets	(10) Total Food Spending
Staple food box	5.98	0.86	1.84	-4.67**	-0.42	-4.82**	-9.05***	0.31	3.28	4.05
	(7.30)	(2.18)	(3.19)	(2.34)	(1.32)	(2.06)	(3.21)	(0.85)	(2.16)	(15.46)
Nutritious food box	2.13	4.36**	4.41	-11.65***	1.44	-3.00	-1.90	1.82**	6.32***	9.25
	(7.79)	(1.76)	(3.20)	(2.31)	(1.30)	(2.05)	(3.77)	(0.91)	(2.08)	(15.83)
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
p-val: staple = nutritious	0.60	0.09	0.44	0.00	0.17	0.40	0.06	0.13	0.22	0.74
R-squared	0.120	0.106	0.088	0.119	0.054	0.077	0.116	0.113	0.138	0.119
Control group mean	92.73	21.13	58.96	25.92	15.91	30.18	48.67	2.41	6.85	286.29
Number of households	2,200	2,212	2,194	2,195	2,217	2,218	2,187	2,180	2,148	2,277

Notes: This Table shows the impact on food expenditure in Egyptian pounds (EGP) over the past week on nine main food items (namely: column (1) meat, poultry and fish, column (2) fruits, column (3) vegetables, column (4) legumes, column (5) eggs, column (6) dairy products, column (7) bread, flour, rice and pasta, column (8) ready-to-eat foods, and column (9) sweets, soda, juice and snacks), and column (10) the total of food spending on all nine items added together. Standard errors in parentheses are clustered at the community level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.4: Correlation Between Predicted Treatment Effects from Staple and Nutritious Food Boxes

	(1) CATE: Kcal Intake (Staple Box)	(2) CATE: FIES (Staple Box)	(3) CATE: HDDS (Staple Box)
CATE: Kcal Intake (Nutritious Box)	0.26***		
	(0.02)		
CATE: FIES (Nutritious Box)		-0.02	
		(0.04)	
CATE: HDDS (Nutritious Box)			0.79***
			(0.02)
Strata FE	Yes	Yes	Yes
R-squared	0.927	0.864	0.934
Number of observations	750	750	750

Notes: This table demonstrates the relationship between predicted conditional average treatment effect (CATE) from the staple-heavy food box and that from the nutrition-sensitive food box among households in the control group. In column (1) the focus is on the CATE on calorie intake (kcal) by the household head. In column (2) the focus is on the CATE on household-level food insecurity experience scale (FIES). In column (3) the focus is on the CATE on household dietary diversity score (HDDS). Standard errors in parentheses are clustered at the community level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.5: Pairwise Correlation Between Predicted Impacts and Predicted Counterfactual

	(1) CATE FIES	(2) BCA FIES	(3) CATE HDDS	(4) BCA HDDS	(5) CATE Kcal Intake	(6) BCA Kcal Intake
Panel A: Staple Food Box						
(1) CATE FIES	1.000					
(2) BCA FIES	-0.165***	1.000				
(3) CATE HDDS	-0.433***	0.232***	1.000			
(4) BCA HDDS	0.195***	-0.418***	-0.779***	1.000		
(5) CATE Kcal Intake	-0.384***	0.154***	0.280***	-0.220***	1.000	
(6) BCA Kcal Intake	0.343***	-0.236***	-0.317***	0.317***	-0.566***	1.000
Panel B: Nutritious Food Box						
(1) CATE FIES	1.000					
(2) BCA FIES	-0.319***	1.000				
(3) CATE HDDS	-0.592***	0.393***	1.000			
(4) BCA HDDS	0.396***	-0.395***	-0.762***	1.000		
(5) CATE Kcal Intake	-0.162***	0.259***	0.278***	-0.131***	1.000	
(6) BCA Kcal Intake	0.046*	-0.282***	-0.307***	0.366***	-0.134***	1.000

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B Intuitive Illustration: Engel Curves and In-Kind Transfers

For further intuition on the nonlinearity and heterogeneity discussed in Section 2, this appendix presents a simple framework linking in-kind food transfers to household food demand. Stylized Engel curves illustrate how the composition of transfers shapes expected nutritional impacts as income levels and transfer types vary.

A Two-Good Engel Curve Framework

Suppose households allocate their food budget between *staples* (S)—such as grains and basic carbohydrates—and *diverse foods* (N), including fruits, vegetables, dairy, and proteins. The classic utility maximization problem is:

$$\max \;\; u(S, N)$$

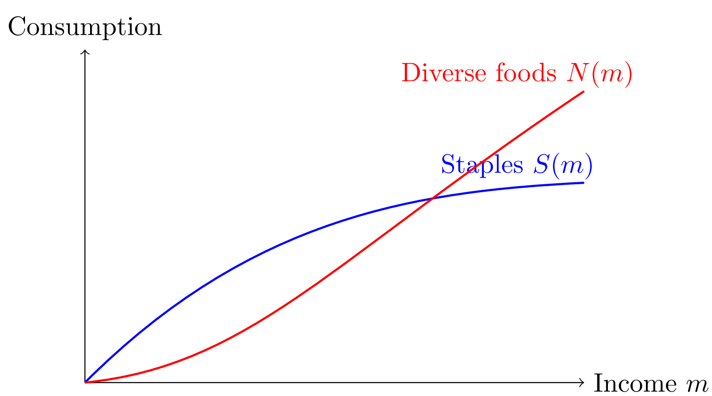
$$\text{s.t.}$$

$$p_S S + p_N N = m,$$

where m is total income allocated to food expenditure, and p_S, p_N are respective prices.

Engel curves describe how the consumption of staples and diverse foods evolves with income. Figure B.1 shows stylized Engel curves: at low incomes, almost all resources are devoted to staples, yielding a steep Engel curve for staples and a relatively flat curve for diverse foods. As income increases, staple consumption plateaus and spending shifts increasingly towards diverse foods, reflecting Engel's law and a nutritional transition from caloric sufficiency to improved dietary diversity.

Figure B.1: Engel Curves: Staples (blue) and Diverse Foods (red) as Income (m) Rises



Effects of In-Kind Transfers

In-kind food transfers can be interpreted as shifting a household's feasible bundle outward along the axes of staples or diverse foods. The nutritional and behavioral impact of such transfers depends upon their position on the recipient's Engel curve:

- **Staple-Heavy Transfers:** Giving additional staples increases S. For low-income households still on the steep portion of their staple Engel curve, this can meaningfully raise caloric intake (extramarginal effect). For better-off or more food-secure households, already near the saturation point for staples, such transfers are likely to simply replace cash expenditures (inframarginal effect), with limited incremental nutritional impact.

- **Nutrition-Sensitive Transfers:** Giving diverse foods directly increases N. Because such foods are ordinarily consumed in greater quantity only at higher incomes, these transfers may leapfrog the Engel curve, accelerating dietary diversity even among households who would otherwise remain staple-focused.

This logic explains why the composition of the transfer—the "stickiness" or lack of fungibility of specific foods—could matter for outcomes beyond a simple "income effect." Nutrition-sensitive transfers may catalyze a dietary transition that income alone would not achieve for the same household.

Nonlinear Effects and the Role of Food Aid Design

This stylized model clarifies the key theoretical points underpinning our empirical results. The impact of in-kind food aid is non-linear and depends critically on both composition and recipient characteristics. Staple-heavy transfers benefit the most calorie-deprived, but may exhibit diminishing returns for others. Nutrition-sensitive transfers can have a broader effect on dietary diversity and food security across the income distribution. Engel curves illustrate that the value of additional food depends not only on quantity (analogous to income) but also on composition. The type of food transferred determines whether a transfer is inframarginal or extramarginal for a given household, generating heterogeneous impacts across calorie intake and dietary diversity. This heterogeneity underpins the need for targeting strategies that reflect program objectives and beneficiary characteristics.

C Baseline Predictors

The baseline survey collected data on the variables that appear in the proxy mean test (PMT) used to target a large-scale cash transfer program in Egypt. This baseline data includes variables on household characteristics, dwelling characteristics, and asset ownership. From among these variables, we selected a subset of variables that exhibit nontrivial variation in our data, and can serve as an indicator for poverty or income level based on our knowledge of the local context. These covariates are:

1. Age of Household Head $<$ 40 years old
2. Household Head is widowed, divorced, or separated
3. Household Head never attended any type of education
4. Household Head attended primary or preparatory
5. Household Head attended secondary education or above
6. Household Head has income from work
7. Household size
8. Dependency ratio
9. Household receives Takaful or Karama cash transfer
10. Number of members without access to pension (neither have social security nor receive self-pension or Daman)
11. Number of household members who never attended school
12. Number of household members currently enrolled in education
13. Number of unemployed household members (cutoff age for unemployment is 15)
14. Household has a Ration card
15. Household receives financial assistance from any source (e.g., relatives, NGOs, etc.)
16. Number of household members per room
17. Rural house
18. Rented housing unit
19. Roof material: Cement brick/ concrete/ stones
20. Roof material: Wood
21. No private bathroom
22. Modern toilet with flush
23. Monthly expenditure on electricity
24. Internet access (or owns a router)
25. Regular mobile phone
26. Smart phone

27. Regular washing machine

28. Half-automatic washing machine

29. Full-automatic washing machine

We attempted to narrow down this list of predictors by running LASSO regressions using these 29 variables to predict our three main outcomes, so we can retain all variables selected by LASSO in any of these three models. However, the results of this exercise were not informative. For the first outcome variable, food energy intake, the LASSO regression did not retain any of the predictors listed above after including strata fixed effects. In contrast, for the second outcome on household dietary diversity score (HDDS), the LASSO regression retained almost all of the predictors listed above, even after including strata fixed effects as non-penalized regressors. For the third outcome, food insecurity experience scale (FIES), the LASSO regression retained 21 out of the 29 variables listed above, in addition to the strata fixed effects.

D Intent to Treat Estimates

We use a standard ANCOVA regression framework to estimate intent-to-treat (ITT) effects (McKenzie, 2012). Specifically, we estimate:

$$Y_{ijs} = \beta_1 Staple_j + \beta_2 Nutritious_j + \beta_3 BaselineY_{ijs} + \lambda_s + \epsilon_{ijs} \tag{5}$$

where Y_{ijs} is the outcome of interest (e.g., calorie intake, food insecurity experience scale (FIES), and household dietary diversity score (HDDS)) for household i in community j in strata s. $Staple_j$ and $Nutritious_j$ are binary variables for the staple food box and nutritious food box treatments, respectively. The omitted group is the control group communities. $BaselineY_{ijs}$ is the value of the outcome variable at baseline. λ_s represents strata fixed effects, since the treatment randomization is stratified at the governorate as well as rural/ urban level. ϵ_{ijs} is a random error term. We cluster standard errors at the community level, which is the level of randomization. A test for $\beta_1 = \beta_2$ reveals whether the nutritious food box has differential impacts on nutritional outcomes.

E Further Analysis of Calorie Intake: Baseline Stratification and Quantile Regression Insights

This section delves into the heterogeneity of treatment effects on calorie intake. By stratifying households into quantiles based on their *baseline* calorie consumption, we assess whether those with lower baseline intake derive greater benefits from the staple and nutritious food box interventions. Additionally, we employ quantile regression techniques to explore how treatment effects vary across the entire distribution of calorie intake at *midline*, providing a comprehensive understanding of the interventions' effectiveness across different segments of the population.

E.1 Baseline Calorie Intake and Treatment Effect Heterogeneity

We examine how baseline levels of calorie intake shape household responses to different food aid interventions. Our analysis reveals that staple food boxes have the greatest impact on calorie intake among households with the lowest baseline consumption, while nutritious food boxes are most effective for those with moderate baseline intake. These findings suggest that tailoring food assistance design to the nutritional status of recipients can improve impacts.

We begin by analyzing heterogeneity in treatment effects based on households' baseline calorie intake, revealing critical patterns in program efficacy. Panel A of Figure E.1 demonstrates a nonparametric visualization of the relationship between baseline and midline calorie intake across treatment arms. This figure shows that the staple food box exerts the greatest impact on enhancing calorie intake among households with the lowest baseline calorie intake (below 1,000 kcal/day). This finding highlights the prevalence of the scenario in our conceptual framework where households with insufficient baseline calories experience the highest impacts.

On the other hand, the nutritious food box treatment has highest impacts on calorie consumption for households with mid-range baseline calorie intake (1,000–2,000 kcal/day). As shown in Panel B of Figure E.1, the majority of sample households fall within the 750 to 2,500 kcal range. The attenuated effects at distributional extremes likely reflect behavioral responses: low-intake households may sell or under-utilize high-value nutritious items, while high-intake households may experience diminishing marginal returns.[15]

Next, we formally estimate treatment effect heterogeneity across quantiles of baseline calorie intake using the following regression model:

$$Y_{ijs} = \sum_{n=1}^{5} \alpha_n Staple_j * Qn + \sum_{n=1}^{5} \beta_n Nutritious_j * Qn + \sum_{n=2}^{5} \gamma_n Qn + \lambda_s + \varepsilon_{ijs} \tag{6}$$

where the outcome Y_{ijs} in this case is calorie intake, which is the only continuous variable among our primary outcomes. Qn denotes an indicator for the nth quantile of the distribution of calorie intake at baseline. The purpose of this analysis is to compare heterogeneity in treatment effects by baseline quantile of calorie intake with heterogeneity in predicted treatment effects, where the predictors are observable characteristics that can be used for targeting. We note that this approach differs from quantile regressions. We present results from quantile regressions in Section E.2.

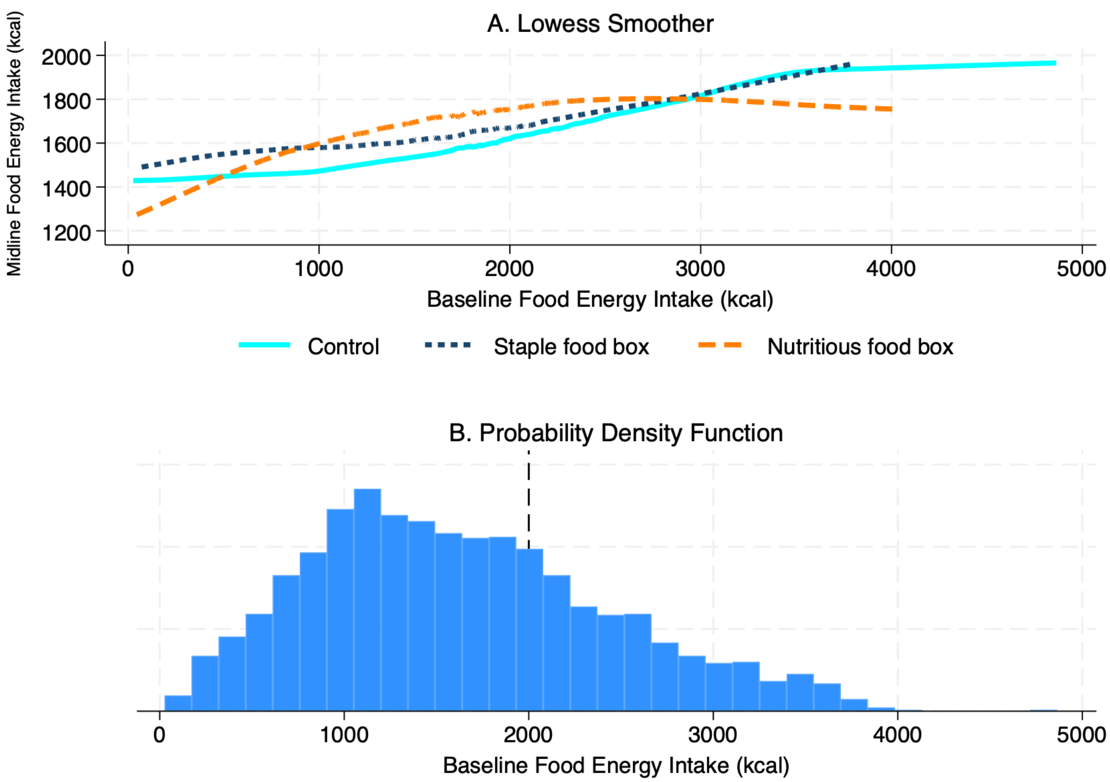
Table E.1 and Figure E.3 display results on heterogeneity in treatment effects on calorie intake across quantiles of baseline calorie intake. These results indicate that households in the third quantile of baseline calorie intake experienced the highest treatment effect from the nutritious box treatment,

[15] This interpretation is supported by Figure E.2, which shows a rightward shift in the mode of calorie intake distribution for nutritious box recipients, with overlapping tails across treatments.

resulting in an 11% increase relative to the recommended daily allowance (RDA) of calorie intake, as depicted in column (2) of Table E.1. However, even at this point, the difference between the impacts of the nutritious and staple food boxes on calorie intake is not statistically significant.

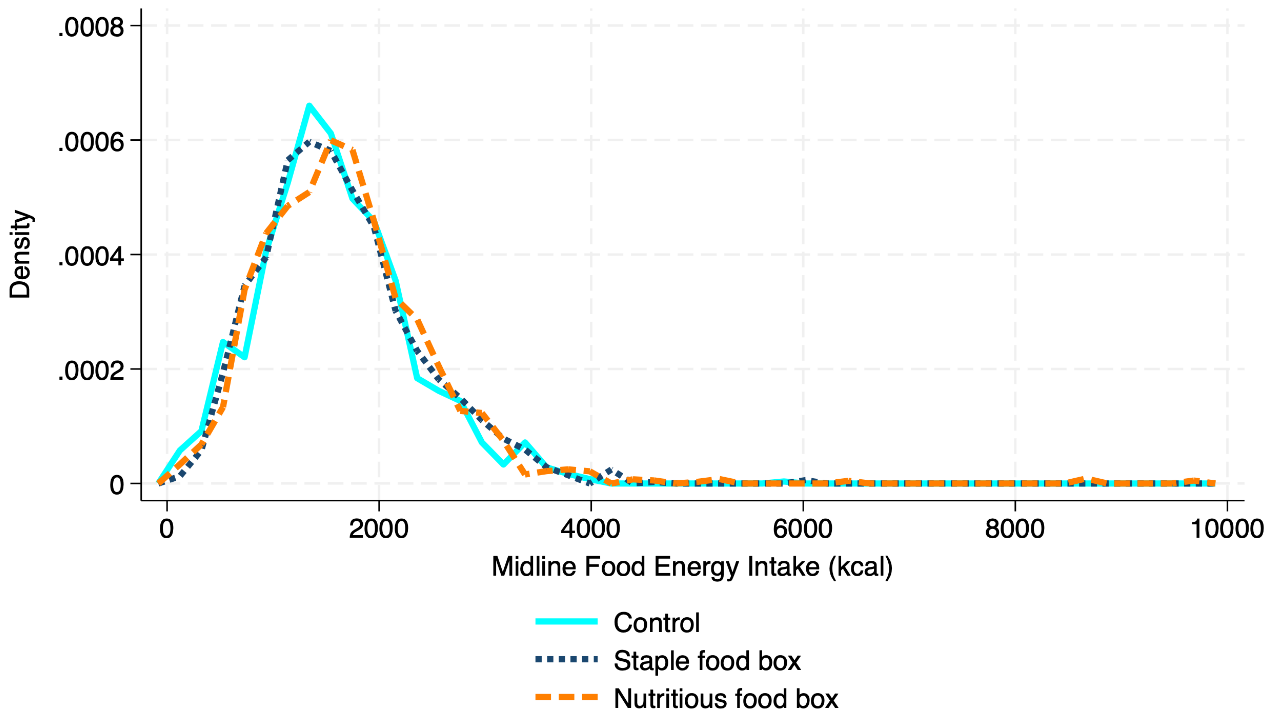
In conclusion, the pronounced impact of staple food box on households with low levels of calorie intake aligns with theoretical expectations of calorie deprivation driving immediate consumption. Conversely, the attenuated effects of nutritious boxes at distributional extremes highlight the complexity of optimizing nutrient-dense interventions.

Figure E.1: Relationship Between Baseline and Midline Food Energy Intake (kcal) by Treatment Status



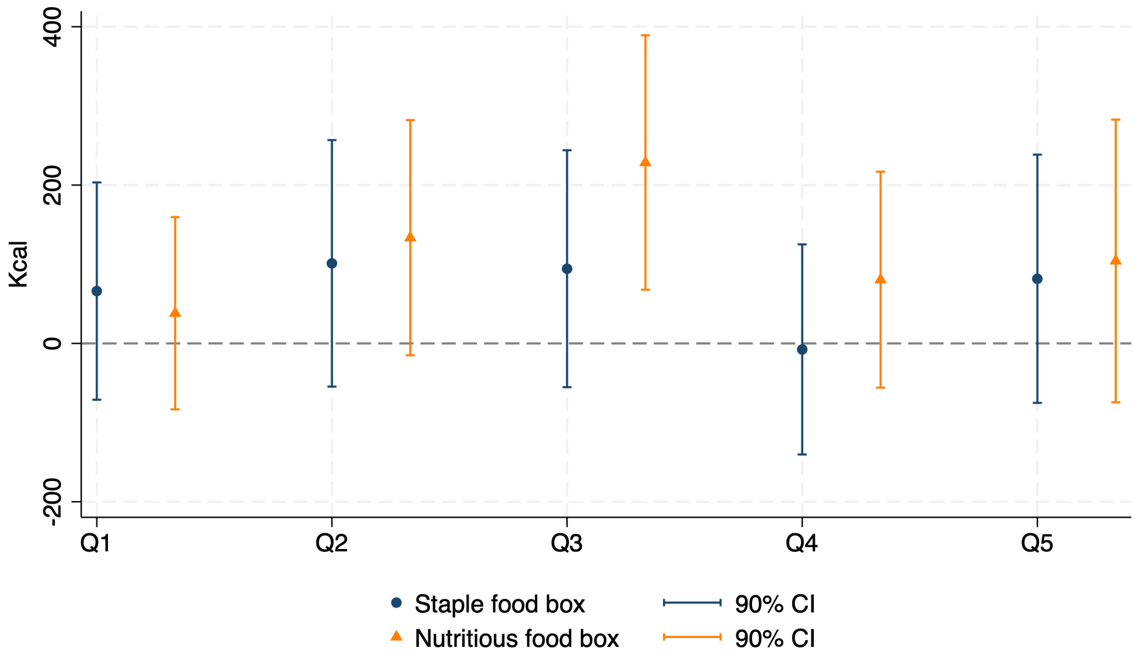
Notes: Panel A) plots a non-parametric relationship between the value of the household head's food energy intake (kcal) at baseline and that at midline, using LOWESS (Locally Weighted Scatterplot Smoothing) technique. The light blue, dark blue, and orange lines represent the control, staple food box treatment, and nutritious food box treatment subsamples, respectively. Panel B) plots a histogram showing the distribution of the household head's food energy intake (kcal) at baseline

Figure E.2: Kernel Density of Midline Food Energy Intake (kcal) by Treatment Group



Notes: This figure presents the Kernel density of the household head's food energy intake at midline by treatment arm. The light blue, dark blue, and orange lines show the distribution for the control, staple food box treatment, and nutritious food box treatment subsamples, respectively.

Figure E.3: Heterogeneity in Treatment Effects by Quantile of Baseline Food Energy Intake



Notes: This figure shows heterogeneity in treatment effects on food effects on food energy intake (kcal) by quantile of baseline food energy intake. Each point estimates represents a regression coefficient on the interaction between an indicator for nth quantile of baseline food energy intake and the corresponding treatment dummy as illustrated in equation 6. The quantiles of baseline calorie intake are: Q1 [26, 926), Q2 [926, 1315), Q3 [1315, 1765), Q4 [1765, 2333), Q5 [2333, 4863).

Table E.1: Heterogeneity in Treatment Effects by Quantile of Baseline Food Energy Intake

	(1) Food Energy Intake (kcal)	(2) Food Energy Intake (kcal%RDA)	(3) Gap in Food Energy Intake (kcal)
Staple box X Q1 baseline kcal	66.10	3.25	-30.98
	(83.43)	(4.10)	(56.50)
Staple box X Q2 baseline kcal	101.12	4.97	-70.26
	(94.67)	(4.65)	(61.47)
Staple box X Q3 baseline kcal	94.34	4.63	-62.40
	(90.96)	(4.47)	(66.87)
Staple box X Q4 baseline kcal	-7.60	-0.37	30.64
	(80.68)	(3.96)	(59.11)
Staple box X Q5 baseline kcal	81.65	4.01	-57.49
	(95.31)	(4.68)	(58.05)
Nutritious box X Q1 baseline kcal	38.11	1.87	-55.23
	(73.79)	(3.62)	(57.12)
Nutritious box X Q2 baseline kcal	133.55	6.56	-74.83
	(90.30)	(4.43)	(54.48)
Nutritious box X Q3 baseline kcal	228.52**	11.22**	-120.56*
	(97.69)	(4.80)	(62.51)
Nutritious box X Q4 baseline kcal	80.51	3.95	24.50
	(82.92)	(4.07)	(58.01)
Nutritious box X Q5 baseline kcal	104.20	5.12	-46.81
	(108.50)	(5.33)	(56.14)
Qn baseline kcal FE	Yes	Yes	Yes
Strata FE	Yes	Yes	Yes
p-val: staple*Q1 = nutritious*Q1	0.73	0.73	0.68
p-val: staple*Q2 = nutritious*Q2	0.74	0.74	0.93
p-val: staple*Q3 = nutritious*Q3	0.18	0.18	0.36
p-val: staple*Q4 = nutritious*Q4	0.30	0.30	0.91
p-val: staple*Q5 = nutritious*Q5	0.83	0.83	0.84
R-squared	0.130	0.130	0.137
Control group mean	1,580.63	77.62	551.17
Number of observations	2,258	2,258	2,258

Notes: The outcome variable in column (1) is food energy intake (kcal) for the female head of household that is estimated using a 24-hour recall survey. The outcome variable in column (2) is the household head's food energy intake as a percentage of recommended dietary allowance (RDA). The outcome variable in column (3) is the gap between the household head's actual food energy intake (kcal) and RDA. The quantiles of baseline calorie intake are: Q1 [26, 926), Q2 [926, 1315), Q3 [1315, 1765), Q4 [1765, 2333), Q5 [2333, 4863). Standard errors in parentheses are clustered at the community level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

E.2 Distributional Effects on Calorie Intake

We extend our analysis to investigate how treatment effects vary across the midline distribution of calorie intake using quantile regression. Unlike the baseline quantile approach in Section E.1, which stratifies households by pre-intervention nutritional status, this method estimates how treatments affect households at different points of the post-intervention calorie intake distribution. This dual lens allows us to disentangle whether program impacts reflect baseline deprivation or emerge dynamically from post-treatment behavioral changes.

We estimate quantile regressions, in which regression coefficients minimize the median absolute deviation at each quantile, q. For Equation 5, $\beta(q)$ coefficients will be estimated to minimize the following:

$$MAD = \frac{1}{n}\sum_{i=1}^{n}\theta_q\big|Y_{ijs} - (\beta_1 Staple_j + \beta_2 Nutritious_j + \beta_3 BaselineY_{ijs} + \lambda_s)\big| \tag{7}$$

where θ_q is a function of asymmetric weights that takes the form:

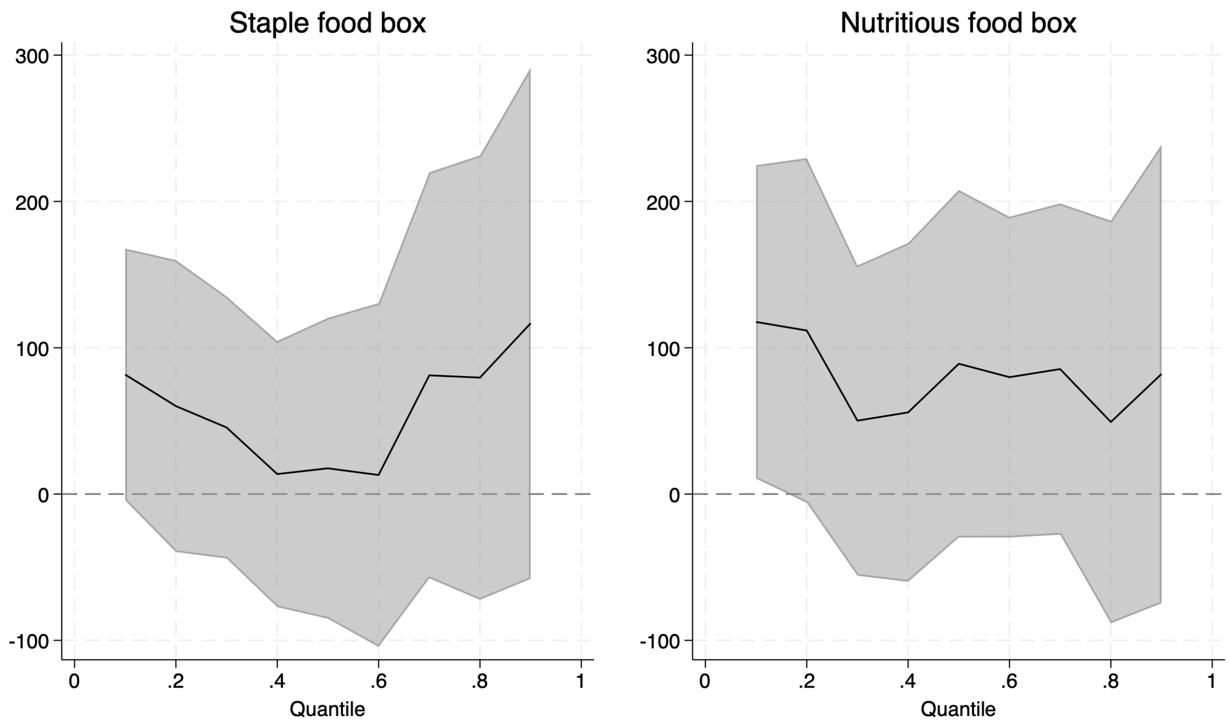
$$\theta_q = \begin{cases} q & if \quad \epsilon_{ijs} > 0 \\ (1-q) & if \quad \epsilon_{ijs} \leq 0 \end{cases}$$

ϵ_{ijs} represents the random error term as before.

Table E.2 and Figure E.4 reveal three key patterns. First, both interventions exhibit overlapping confidence intervals across much of the midline distribution, suggesting limited statistical significance in heterogeneous effects. Second, staple food boxes show relatively stronger impacts at the distributional extremes (e.g., the 10th and 90th percentiles), potentially reflecting catch-up consumption among the most deprived or surplus accumulation among better-off households. Third, nutritious boxes exhibit peak efficacy at the lowest quantile.

Therefore, while quantile regressions complement our baseline analysis by capturing post-treatment dynamics, their limited statistical precision underscores the practical challenges of real-time monitoring for unintended consequences in food aid programs.

Figure E.4: Quantile Treatment Effects on Food Energy Intake (kcal)



Notes: This figure plots results from the quantile regressions of the treatment effect on the household's food energy intake (kcal). Panels A and B show the results for the staple food box and nutritious food box treatments, respectively. The shaded regions represent 90% confidence intervals.

Table E.2: Impact on Food Energy Intake (kcal): Quantile regression

	(1) Q10	(2) Q25	(3) Q50	(4) Q75	(5) Q90
Staple food box	81.69	45.35	17.62	109.12	116.77
	(59.83)	(59.17)	(59.12)	(87.31)	(94.42)
Nutritious food box	117.65*	75.46	89.08	93.86	82.09
	(60.36)	(65.71)	(76.17)	(84.23)	(115.26)
Baseline value of the outcome	0.10***	0.12***	0.16***	0.16***	0.15***
	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)
Strata FE	Yes	Yes	Yes	Yes	Yes
p-value: staple = nutritious	0.57	0.66	0.30	0.87	0.78
Observations	2,258	2,258	2,258	2,258	2,258

Notes: Quantile regressions are estimated using STATA command *qrprocess*. Bootstrapped standard errors in parentheses are clustered at the community level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

F Machine Learning Analysis Implementation Steps

We implement a supervised machine learning (ML) approach to flexibly estimate the relationship between baseline household characteristics (Z) and potential nutritional outcomes under treatment ($D = 1$) or control ($D = 0$). The goal is to obtain individual-level predictions of conditional average treatment effect (CATE) and baseline conditional average (BCA) while minimizing overfitting and ensuring out-of-sample validity.

Algorithms Used

We apply a set of standard supervised learning methods that capture both linear and nonlinear relationships: random forest, elastic net, support vector machines, and gradient boosted trees. Using multiple algorithms reduces dependence on a single model's functional form and helps capture complex patterns in the data.

Repeated Cross-Fitting Procedure

To ensure robust and generalizable predictions, we adopt a repeated sample-splitting (cross-fitting) approach:

1. **Sample Splitting:** In each iteration, the data are randomly split into a training set and a validation set of approximately equal size (50–50).
2. **Model Training:** The training set is used to fit the ML models for $\widehat{E}[Y \mid D = 1, Z]$ and $\widehat{E}[Y \mid D = 0, Z]$ separately, using only baseline covariates Z.
3. **Out-of-Sample Prediction:** The trained models generate predicted treated and untreated outcomes for each observation in the validation set.
4. **CATE Estimation:** For each household i, the CATE is computed as:
$$\hat{\tau}(Z_i) = \widehat{E}[Y_i(1) \mid Z_i] - \widehat{E}[Y_i(0) \mid Z_i].$$
5. **Repetition and Aggregation:** Steps 1–4 are repeated at least 100 times, each with a new random split. The final predicted CATE for each household is the median across all repetitions.

Separate Estimation by Treatment Arm

The above procedure is implemented separately for each treatment contrast:

- Staple food box vs. control
- Nutritious food box vs. control

This ensures that the prediction models are tailored to the specific patterns of heterogeneity relevant to each program type.

Overfitting Mitigation and Validity

Two features of the procedure safeguard against overfitting: (i) each household's CATE is based solely on out-of-sample predictions, and (ii) results are aggregated over many random splits, reducing dependence on a single partition.

ALL IFPRI DISCUSSION PAPERS

All discussion papers are available **here**

They can be downloaded free of charge

INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE
www.ifpri.org

IFPRI HEADQUARTERS
1201 Eye Street, NW
Washington, DC 20005 USA
Tel.: +1-202-862-5600
Fax: +1-202-862-5606
Email: ifpri@cgiar.org