HiCN Households in Conflict Network
hicn.org

Conducting (Long-term) Impact Evaluations in Humanitarian and Conflict Settings: Evidence from a complex agricultural intervention in Syria

Aysegül Kayaoglu[1,2,3], Ghassan Baliki[1], Tilman Brück[1,4,5]

HiCN Working Paper 386

March 2023

Keywords: causal inference, matching, entropy balancing, conflict, food security, resilience, agricultural interventions

JEL classification: C18, C3, D1, I31, O13, Q1

Abstract
The number of vulnerable people in humanitarian emergencies worldwide is increasing due to the rising frequency and intensity of risk exposure. At the same time, most interventions in humanitarian emergency and conflict settings (HECS) are short-term in nature, as if people only require temporary help to overcome adversity. Yet there is an acute scarcity of rigorous impact evaluations in HECS testing if assistance works well (or at all). Moreover, the few available studies only cover a small range of countries and contexts. Furthermore, the knowledge gap concerning the long-term impacts of crisis interventions is even more pronounced. These gaps are primarily caused by the unavailability of (long-term) panel data in emergencies and by the challenges of constructing feasible counterfactuals. Our paper contributes to the literature in four ways. First, we review recent research on covariate balancing to assist researchers in conducting a rigorous impact evaluation in HECS with non-randomized treatment assignments and significant covariate imbalances between the treatment and control groups due to targeting. We thus suggest methods to overcome the challenges associated with conflict or humanitarian contexts. Second, employing a range of such methods for one case study, we offer rigorous evidence on the long-term causal impacts of agricultural interventions in a humanitarian crisis

[1] ISDC - International Security and Development Center, Germany
[2] Department of Economics, Istanbul Technical University, Türkiye
[3] IMIS, University of Osnabrück, Germany
[4] Thaer-Institute, Humboldt-University of Berlin, Germany
[5] Leibniz Institute of Vegetable and Ornamental Crops (IGZ), Germany

setting. Third, we show that agricultural or livestock interventions have different impacts in the long term, which implies that the combined interventions might have a more sustainable impact on households. In other words, our analyses demonstrate that short-term humanitarian assistance can indeed have long-term development impacts. Fourth, we offer innovative evidence for the case of Syria, using unique panel data with four waves of treated and untreated households, thus expanding the range of countries ever studied in the literature on humanitarian emergencies.

Acknowledgement
We would like to thank the tireless support of Melodie Al Daccache and Dorothee Weiffen with the data processing.

Funding
This work is supported by the Centre of Excellence for Development Impact and Learning (CEDIL), funded by UK Aid.

Conflict of interest
The authors declare no conflict of interest.

1. Introduction

The increased frequency and severity of risk exposure increase the population of vulnerable people worldwide. Two billion people live in fragile and conflict-affected areas (OECD, 2022), and over 220 million people in these settings experienced acute food insecurity in 2022 (WFP, 2022). Thus, strategies to enhance households' food security and resilience capacity are critical for enabling households to manage and cope with various shocks and stresses.

That said, many interventions in humanitarian emergency and conflict settings (HECS; hereafter) are focused on the short-term and assume that individuals need constant support during the emergency phase of a conflict. Thus, the long-term development impacts of humanitarian interventions are (implicitly) considered to be negligible. An important reason for this assumption is the lack of panel data availability in HECS to analyse the long-term dynamics of humanitarian interventions. Another reason is the difficulty of conducting rigorous impact evaluations in HECS, knowing that only 7% of impact evaluation studies were completed in fragile contexts (Moore et al., 2021). Puri et al. (2017) discuss in detail that impact evaluations in HECS are difficult due to various methodological, ethical, and practical challenges. These challenges include selection bias, information bias, contamination bias, non-random attrition and response, the need for rapid evaluations, the attribution problem, and (intentionally) harming vulnerable populations. Although methodological obstacles may be avoided by utilising randomisation to assign treatment and control groups, implementing Randomized Control Trials (RCTs) is especially problematic in HECS due to security, ethical, budgetary, and political considerations.

Furthermore, these existing impact evaluations primarily focus on a few fragile countries, neglecting the situation in many other countries. In their evidence map gap, Moore et al. (2021) find no rigorous impact evaluations, for example, in Djibouti, Central African Republic, Comoros, Congo, Eritrea, Afghanistan, Papua New Guinea, Tuvalu, Kiribati, Solomon Islands and Micronesia. The same evidence map also notes that only one food systems related impact evaluation was conducted in Syria, which dates back to 2008. When we check the evidence hub for impact evaluations[1] in Syria, we observe that there is also only one impact evaluation study in the agriculture, fishing and forestry sectors which showed that the adoption of zero tillage

[1] Please see https://developmentevidence.3ieimpact.org/search-result-details/impact-evaluation-repository/does-zero-tillage-improve-the-livelihoods-of-smallholder-cropping-farmers-/7447

technology increased the net crop income and per capita wheat consumption of treated households (El-Shater et al., 2015).

Thus, it is clear that one of these least studied humanitarian emergencies case is Syria, despite the fact that the civil war started in 2011 has led to the largest displacement crisis since World War II and almost 70% of the population in Syria (before the earthquake of early 2023) requiring humanitarian assistance (UNOCHA, 2022). In the early phases of humanitarian support to Syria, the focus was on providing in-kind emergency support. This early wave of emergency support included interventions such as the provisions of food and non-food items, shelter and health services. The following phase of emergency support included conditional and unconditional cash transfers, either in itself or in combination with the in-kind provisions. These emergency supports had a short-term targeting, such as reducing the immediate food insecurity of individuals. However, food insecurity is still very high even after more than a decade of emergency relief packages. This partly led to a transition towards developmental supports such as agricultural and livestock assistance to decrease food insecurity sustainably. However, given the evidence gaps, we lack sufficient understanding of how the impacts of these short-term or long-term oriented interventions differ, for example, in the emergency, rehabilitation, and peacebuilding phases of humanitarian emergency and conflict settings.

That said, and in light of the budget constraints of humanitarian stakeholders, there is a significant and growing need to find more effective ways to meet the needs of millions of conflict-affected individuals in Syria. This suggests that impact evaluations are critical for expanding the evidence base for better programming and targeting to reduce food insecurity in Syria. However, in the post-war period, no impact evaluation study focused on the impact of agricultural and livestock assistance. Therefore, more rigorous research is needed on the impact of both emergency and developmental supports.

Our paper contributes four ways to the literature on the impact evaluation in HECS. First, we synthesise the recent literature on covariate balancing to help researchers in HECS who face severe difficulties in conducting a rigorous impact evaluation to understand the available tools in proceeding with the impact evaluation, even in the case of non-randomized treatment assignment and high imbalances of covariates across treatment and control groups. Second, we provide the first available long-term causal evidence on the welfare impacts of agricultural interventions in humanitarian and conflict settings. Third, using a unique panel data of four waves, we provide evidence on whether agricultural and livestock assistance has any long-term developmental effects. And we show that agricultural or livestock interventions have different impacts in the long term, which implies that the combined interventions might have a more considerable and more sustainable impact on households. Fourth, we find and present long-term evidence from the case of Syria, which has been facing ongoing crises since the start of the civil war in 2011. Thus, we also contribute to the literature by expanding the range of countries ever studied for impact evaluations on humanitarian emergencies and conflict settings.

The structure of the article is as follows. Section 2 reviews the literature on the importance of long-term impact assessment of agricultural interventions in HECS. Section 3 explains the context of Syria, while Section 4 summarises the interventions we analyse. Section 5 brings the methodological discussion, and Section 6 provides the information about the data we analyse. Moreover, Section 7 presents the empirical findings. Finally, Section 8 concludes.

2. Literature

Conflict, climate change, rapid urbanisation and scarce natural resources are among the most pressing current challenges placing food systems under stress and contributing to food insecurity (Haddad et al., 2016; Willett et al., 2019). This situation, in return, puts an additional risk on the environment (UNEP, 2019; Springmann et al., 2018; UNSCN, 2019) that further affects food production negatively. Moreover, the necessity to share the already scarce natural resources creates tensions and conflict among people. It is estimated that around 700 million people live in extreme poverty (World Bank, 2016), and there is an urgent need to increase food production by 50% by 2050 (FAO, 2016). Besides, most people who face undernutrition live in low- and middle-income countries (WFP, 2015). Moreover, conflict and political instability increase food insecurity and undernutrition (Martin-Shields & Stojetz, 2019; Tranchant et al., 2021). In addition to weakening food production (George et al., 2021; Adelaja & George, 2019; Arias et al., 2019), armed conflicts also deteriorate health systems and destroy markets and institutions, which are also highly related to food security (Naude et al., 2023; Mercier et al., 2020; Justino, 2012). To deal with food shortages in conflict settings, households adopt harmful coping strategies, such as child labour (Churchill et al., 2022) or underage marriage (Bartels et al., 2018).

Poverty is a crucial causing factor in malnutrition and food insecurity. Relatedly, healthy and culturally appropriate meals are frequently more expensive and less accessible, particularly in low-income and rural settings (Darmon & Drewnowski, 2015; Dizon & Herforth, 2018; Development Initiative, 2020). This may cause significant socioeconomic gaps in nutrition and, as a result, non-communicable illnesses linked to diet (Darmon & Drewnowski, 2015). In addition to the deteriorating impacts of conflict on food security, an increasing number of studies show that it also affects the resilience of individuals, which is essential for their long-term welfare (Brück et al., 2019; Shemyakina, 2022). Resilience is defined as "the capacity of a household to bounce back to a previous level of well-being (for instance, food security) after a shock" in Resilience Index Measurement and Analysis (RIMA) (Alinovi et al., 2008) which is an econometric approach offered by FAO in 2008. Resilience Measurement Technical Working Group (RM-TWG) defines it as "a capacity that ensures stressors and shocks do not have long-lasting adverse development consequences". Analysing the long-term impact of the Naxal insurgency in India, Tranchant et al. (2014) found that individuals affected by the conflict had lower resilience capacities to cope with income shocks. Similarly, Brück et al. (2019) showed that the 2014 Gaza conflict reduced the resilience capacity of conflict victims by reducing their adaptive capacities. However, they also provided evidence that immediate support to those exposed to conflict can help restore their resilience.

Knowing that half of the world's poor will be living in HECS by the end of this decade (World Bank, 2021) and considering the spatial distribution of conflict and war in the world, it is, therefore, easy to understand the recent interest on the impact of interventions related to food security (Baliki et al., 2022b). Moreover, with the lower capacity of international funds than what is needed to fight poverty and food security, it is crucial to find sustainable interventions that cause long-term positive impacts, such as those that increase the resilience capacities of individuals and access to food in the long term.

Nevertheless, there is a dearth of evidence about the long-term impacts of input transfers or other agricultural interventions on individuals' food security and resilience, particularly in conflict settings. Among few, Jansen et al. (2022) measured the impact of beekeeping and entrepreneurship interventions in Tanzania and found that they reduced exposure to community violence. It was also found that there was an increased impact on young men's financial and social capital. Weiffen et al. (2022), using the same case study as ours, found that the provision of vegetable kits improved the short-term food security of female-headed households in Syria. There is also an increasing interest in the literature analysing the impact of interventions separately for men and women, particularly on women's empowerment. For example, recent studies found that agricultural technology training and improving the farmers' access to markets reduced women's decision-making power in the household in terms of the power of their views on the production and spending of income (Ntakyo & Van den Berg, 2022; Depenbusch et al., 2022; Baliki et al., 2022a; Gaworek-Michalczenia et al., 2022; Salazar et al., 2021).

3. The context

According to the Global Peace Index, Syria remains the world's third least peaceful country in 2022. In addition, the Syrian Observatory for Human Rights (SOHR) reports that 610,000 people have died in Syria since the start of the war[2]. One of the war's most severe and enduring effects is the displacement issue in Syria. According to UNOCHA, 12.3 million people have been compelled to leave their homes since the war began in 2011, and 6.7 million people are now displaced inside their nation.

The conflict severely affects the welfare and sustainable livelihoods of the country's remaining population. By 2021, Syria's economic production has decreased by 60%, and the value of its currency has fallen by 99%. Evidence from Syria's post-2011 agricultural production also demonstrates that vulnerable households –including those headed by women, unemployed young men, small-scale farmers who lost their assets due to the conflict, IDPs, or host families who must share limited natural resources with IDPs– are unable to obtain agricultural inputs (CFSAM, 2015, 2016; ECSWA 2016). The war also caused the loss of capital stock and tangible goods (ESCWA, 2016). Integration of value chains and resilience-building strategies is unavoidable in an environment where the agriculture industry and value chains are fragmented (SCPR, 2016). Moreover, the conflict was not the only factor reducing rainfed agricultural productivity; the recurrence of the drought was also a blow to smallholder farmers' output (Wendle, 2016).

---

[2] https://www.syriahr.com/en/243125/

Given the extreme food shortages and insecurity rates, the agricultural sector is crucial for livelihoods. In addition to the effects of a lengthy conflict, Syrians face the disastrous repercussions of climate-related shocks, with severe drought and flooding generating additional uncertainty. Syria is suffering from a severe and long-term drought, which has resulted in poor vegetative conditions and drier-than-normal precipitation seasons in 2022. Water shortages resulted in significant crop and economic losses, rising waterborne infections and malnutrition rates, displacements, and increased protection and gender-based violence, particularly for women and children (UNOCHA, 2022). Additionally, it is becoming more difficult for farmers to access, use, and get seeds; therefore, seed security is a crucial problem in Syria following the conflict (Bishaw et al., 2015).

As of December 2022, it is estimated that 15.3 million people, over 22.1 million total population in Syria, need humanitarian assistance while routine shortages of essential goods are common (UNOCHA, 2022). Moreover, 46% of those in need of assistance are children. That said, poverty rates jumped with the destruction of livelihoods, markets and institutions. According to World Vision, the economic cost of the war reached US$1.2 trillion by 2021. According to the UN Office for the Coordination of Humanitarian Affairs (UNOCHA), more than 90% of the population lives at the subsistence level. The World Food Programme (WFP) estimates that 12.4 million Syrians are food insecure, a startling rise of 3.1 million annually. More than 600,000 kids in Syria suffer from chronic undernutrition. Moreover, the population needing humanitarian assistance mostly lives in the country's northwestern region.

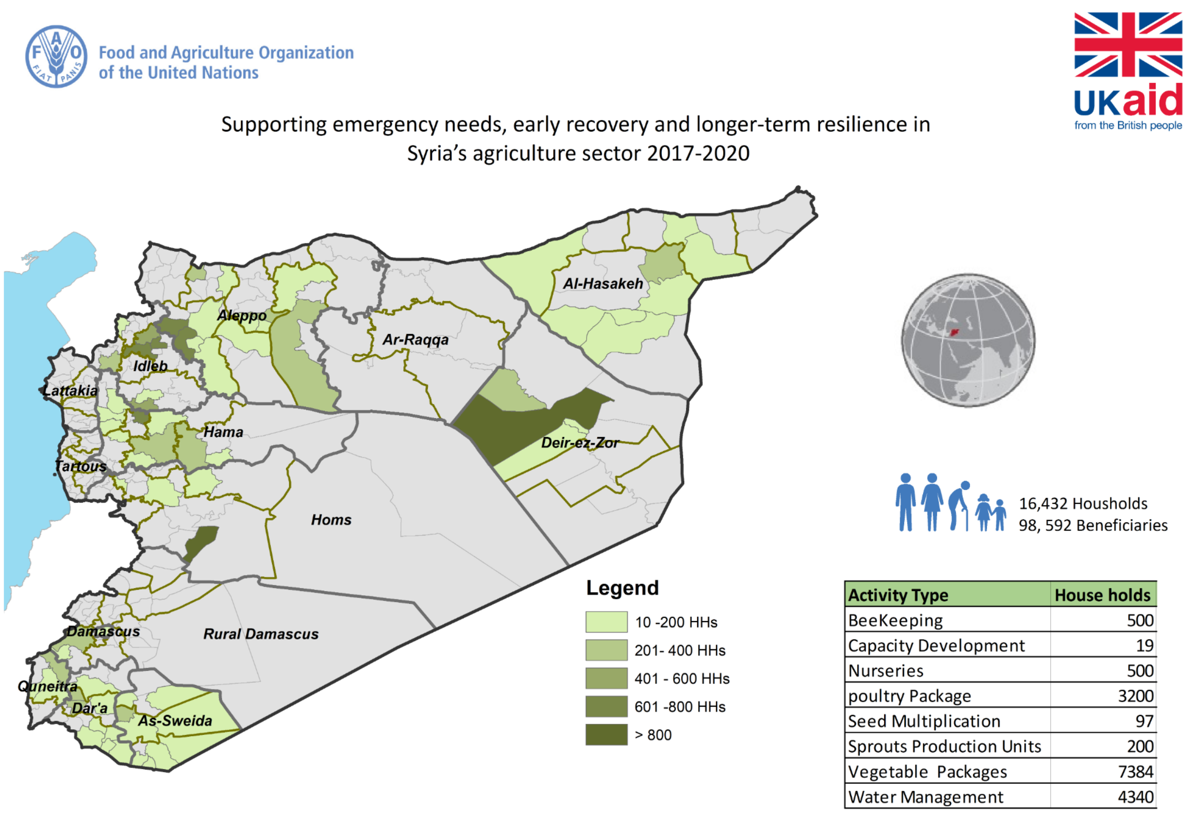
4. The intervention

Many interventions targeting sustainable agricultural and livestock production in low and middle-income countries aim to improve individuals' long-term resilience. This is especially crucial for the people living in HECS, as conflict has a deteriorating impact on assets, institutions and markets. The project "*Supporting emergency needs, early recovery, and long-term resilience in Syria's agriculture sector*" is an FAO initiative in Syria supported by FCDO (previously DFID). FAO Syria executed the initiative in nine governorates across Syria between October 2017 and May 2021, implementing several interventions which had three key objectives: to increase food availability for vulnerable households through improved smallholder production; to build sustainable access to productive assets, income and food supply; and to foster enabling environments for resilience building and recovery of the agricultural sector.

To that end, the Program offered emergency and resilience assistance. Emergency assistance comprises home agriculture and livestock inputs, whereas the resilience package includes community-wide irrigation system rehabilitation. Furthermore, it is critical to understand that all assistance supplied under this program is divided across Syria and different governorates. Moreover, each household was provided only one type of support under this program, and the interventions are at the household level. The theory of change posits that emergency assistance increases farmers' access to high-quality seeds and livestock so that (1) agricultural and livestock output can be improved in the short-run; (2) higher productivity levels, strengthened household-level food security, diverse income generation opportunities, and a decrease in the use of harmful coping strategies can be achieved in the medium-run and long-run. Regarding resilience, the effects mentioned above suggest a reduced vulnerability to shocks.

The following are the phases of interventions. Distribution of vegetable toolkits and some beekeeping support were part of the initial phase (September 2018 - June 2019). The distribution of poultry, irrigation rehabilitation, and the remaining beekeeping assistance was part of the second phase (July 2019 - March 2020). Finally, the creation of low tunnel nurseries and the delivery of vegetable seedlings, salt blocks, and livestock vaccinations were all part of the third phase (April 2020 - December 2020). ***Figure 1*** below shows the spatial distribution of all interventions across Syria.

Figure 1. The number and location of households reached under the FCDO programme



Source: FAO Syria

5. Methods

5.1. Potential Outcomes Framework and Causal Inference in Non-randomized Interventions

The potential outcomes framework has an essential assumption for causal inference, namely the unconfoundedness assumption, which is

$$(Y^1, Y^0) \perp D | X$$

, where D is the treatment status, X is a vector of covariates, Y^1 and Y^0 are, respectively, the potential outcomes when the treatment is taken or not. So, this assumption states that treatment assignment should be independent of covariates X. Or, we can state that the treatment assignment

D will be as good as random conditional on X. The unconfoundedness assumption is also called the 'ignorability' or 'selection on observables' assumption.

Arguably, achieving the ignorability assumption without randomisation (complete or fully blocked) is hard. However, randomised control trials (RCTs) are not always feasible, particularly in HECS, due to various ethical, methodological and practical challenges (Puri et al., 2017). Therefore, researchers use the appropriate quasi-experimental causal inference method to conduct a rigorous impact evaluation in such settings. However, quasi-experimental methods such as Regression Discontinuity Design (RDD), Difference-in-differences (DiD), and Instrumental Variables (IV) also rely on identification assumptions on unobservable factors. Thus, in observational data settings where the treatment assignment is not randomised, it is plausible to assume that unobserved individual factors play a role for some individuals receiving the treatment while others stay in the control group. Therefore, in observational studies, eliminating the selection bias is a crucial target for researchers who use various conditional causal inference methods such as weighting, matching, and sub-selection to ensure that the unconfoundedness assumption of the potential outcomes framework will hold or be justified. As the imbalance of observable and unobservable characteristics of treatment and control group individuals in interventions creates the problem of selection bias in causal inference, balancing of observable covariates is necessary because model dependence increases with the imbalance, which leads to biased estimates.

Therefore, one needs to adjust the covariates included in the model specification to find an intervention's (D) causal impact on the outcome variable (Y). In this balancing of confounders, it is essential to have the correct model specification to adjust all relevant confounders before the impact analysis. Thus, as we do in other quasi-experimental techniques such as instrumental variables methods and difference-in-differences, we need an identification assumption before proceeding with our impact evaluation. With matching, weighting or sub-classification, the idea is to have a balanced covariate distribution for treated and control groups. Balancing will also enable us to assume that, on average, unobservable characteristics would be similar. In other words, these aim to satisfy the conditional independence assumption, which tells us that the treatment assignment is as good as random conditional on some covariates. However, a researcher might not know the exact confounders that matter for the outcome. In cases where we have complete information about the estimation model, we might not have information about all the covariates in observational data. Thus, not only the observed covariates but also the relevant unobservable confounders matter for the identification strategy of causal inference. Moreover, an essential assumption for matching methods to provide unbiased estimates is that our estimation model has no unobserved confounders. Thus, the use of balancing the confounding variables with the difference-in-differences method is a good identification strategy because the latter will allow for the selection-on-unobservables. At the same time, the former considers the selection-on-observables.

Before comparing different covariate balancing for the impact evaluation, one must first discuss the conditions under which such methods would provide consistent estimates. If we use these balancing methods, we should also question their implications before using them blindly in any research setting. In other words, a researcher should consider how one group ends up being in the treatment group, and the other is in the control group, although they have similar characteristics (which we aim to achieve through matching, weighting or regression). If we cannot justify these reasons, then a matching strategy will still yield biased results due to differences of unobservable confounders. This is less of a concern if the treatment assignment is not decided by the treated units themselves so that there is no possibility of sorting. Thus, there are several reasons to use matching methods if it is clear that the unobservable characteristics of treatment receivers and control group members do not matter for their treatment status. Firstly, in settings where it is not the individuals but a different decision-maker who decides who will receive the treatment, we can assume that balancing covariates is useful in observational data. Moreover, it might be helpful in cases where both control and treatment group individuals were willing to or in need of treatment. However, operational capacities result in only a random portion of the potential treatment group. That said, if there are rather strict rules about who will get the treatment, researchers can use these rules for their identification strategy. Alternatively, it might be the case that decision-makers decide about the treatment status with a different potential outcome than the one analysed by the researcher, which might lead researchers to assume that unobservable characteristics of individuals on the latter should have negligible impact on the estimate (Imbens & Rubin, 2015).

5.2. Identification strategy for the long-term impact assessment

As is explained in the following section, we have unique panel data in a conflict setting. However, the baseline data (pre-treatment data) is only available for the 'vegetable kits' intervention. We have at least two waves of post-treatment data for the other interventions. Therefore, we employ different causal inference methods to analyse the long-term impact of interventions separately.

The difference-in-differences (DiD) framework is a widely used method in causal inference. However, it can only be used once panel or repeated cross-sectional data covering the pre- and post-treatment period is available. Thus, comparing the pre- and post-treatment trends of both the control and treatment groups provides researchers with the causal impact when time-invariant unobserved covariates invalidate the ignorability assumption. In other words, accounting for the time-invariant heterogeneity facilitates the causal inference (Angrist & Pischke, 2009). There are three critical assumptions of DiD.

The first assumption is SUTVA which implies that the outcome of each unit is independent of the other unit's treatment assignment status. The second is called no anticipation, which means that treated units should not know whether they will be treated or not, so they cannot manipulate their outcome. These two assumptions are not specific to DiD but apply to all methods that use the potential outcomes framework. The third assumption, however, is particular to the DiD and assumes parallel trends between treated and control groups. This assumption states that these parallel trends we observed in the pre-treatment period should have continued if there had been no treatment. In other words, parallel trends between treated and control units are assumed to exist in both pre- and post-treatment periods in the absence of treatment. We can test the pre-treatment trends if we have more than two pre-treatment periods in our data. However, we can never test what would have happened to post-treatment trends in the absence of treatment. That's why it is important to satisfy the parallel pre-trends so that our assumption of parallel post-trends in the absence of treatment becomes more plausible. Moreover, Kahn-Lang and Lang (2020) argue that it is not only the parallel trends but also similar levels that one needs to achieve an unbiased impact of an intervention using DiD. Thus, even in cases where no panel or repeated cross-sectional data is available for the pre-treatment period, researchers can first employ methods to balance the covariates and pre-treatment outcomes to be similar to each other and then follow with the DiD strategy. This will help researchers to find the hidden experiment within the observational data.

We do this in the long-term analysis of the ‘vegetable kits’ intervention in Syria, where we have only one pre-treatment wave but three post-treatment data periods. In order to account for the observed heterogeneity, we control for various household characteristics which should not be affected by the treatment. Thus, one could use first matching techniques such as nearest-neighbour matching, propensity score matching[3] or coarsened exact matching and then run the DiD model.

However, with the random pruning in propensity score matching or the possibility of making wrong designs for coarsened variables, researchers might end up having a smaller sample size, decreased power and bias. Besides, matching is an iterative method that leads researchers to continuously revisit their model specification if there are still imbalances after matching. Moreover, the recent literature on matching suggests that propensity score matching should not be used as a matching method but rather with inverse probability weighting (King and Nielsen, 2019). The balancing exercise we conducted using our panel data also showed that either propensity score matching or other distance matching methods could not satisfy the balance of covariates between treated and control units.

Another identification strategy is using kernel propensity-score matching DiD. Rosenbaum and Rubin (1983) showed that it is enough to condition the probability of being treated if strong ignorability assumption holds. In kernel propensity-score matching DiD, using the covariates unaffected by the treatment, one first calculates the kernel weights by estimating the propensity scores following Heckman et al. (1997, 1998). Kernel weights are then used in estimating the treatment effect[4]. This kernel propensity-score matching DiD method can also be used in repeated cross-sectional data (Blundell and Dias, 2009). One can also use only the observations in the common support to increase the internal validity of the average treatment effect. Importantly, we can also test if the outcome variable is orthogonal to the treatment indicator given the set of covariates, namely testing the balancing property.

---

[3] Propensity score estimation works as a way to fight the curse of dimensionality problem. However, the estimated one-dimensional propensity score that is used to cure this problem of matching creates bias because of random pruning.

[4] One can also use inverse probability weights which are differently estimated for each observation depending on their propensity score. Then these weights are used in the estimation of the impact. On the other hand, OLS regression conditional on Xs (covariates) gives more weight to observations with covariates with a larger variance of treatment probability.

That said, panel data and even repeated cross-sectional data are hard to collect in HECS. In cases where researchers have cross-sectional data, matching and weighting methods can still be used to balance the observable covariates between treatment and control group individuals. However, propensity score matching randomly prunes the observations and, therefore, increases bias. Moreover, other matching methods might greatly decrease our sample size if there were already large imbalances between treatment and control groups in the pre-treatment period, which would probably be the case in many HECS settings due to targeting the neediest populations. Therefore, entropy balancing (Hainmueller, 2012) or hierarchically regularised entropy balancing (Xu & Young, 2021), an extension of entropy balancing, can be used to enforce the balance across covariates through weighting. As seen below, entropy balancing and hierarchically regularised entropy balancing methods help us enforce balance even when no baseline data can be used in impact evaluations in HECS. The following section discusses the summary statistics of outcome variables for our long-term impact assessment, followed by the findings.

6. Data

Data is collected in four waves. First, baseline data is obtained in the fourth quarter of 2018, and then the following three post-treatment data are gathered in the first quarter of 2020, 2021 and the second quarter of 2022. However, the interventions were distributed in sequence, and this phased-in design was necessary, given the feedback the implementation team had received after the first wave of intervention distribution.

This type of implementation also resulted in changes and flexibility in our research design, as we had the baseline information only for households who received the vegetable kits intervention. For other households that received either beekeeping, poultry kits, livestock vaccination and salt blocks, agricultural tools and vegetable seedlings provision, we only have post-treatment information. ***Table 1*** below presents the availability of data for each intervention type. As can be seen, we have at least two waves of panel data for each intervention.

Table 1. Available Panel Data Waves for each Intervention Type

Intervention	Wave 1 (pre-treatment)	Wave 2 (post-treatment)	Wave 3 (post-treatment)	Wave 4 (post-treatment)
Vegetable kits	√	√	√	√
Poultry kits		√	√	√
Beekeeping		√	√	√
Agricultural tools and vegetable seedlings provision			√	√
Vaccination and salt blocks			√	√

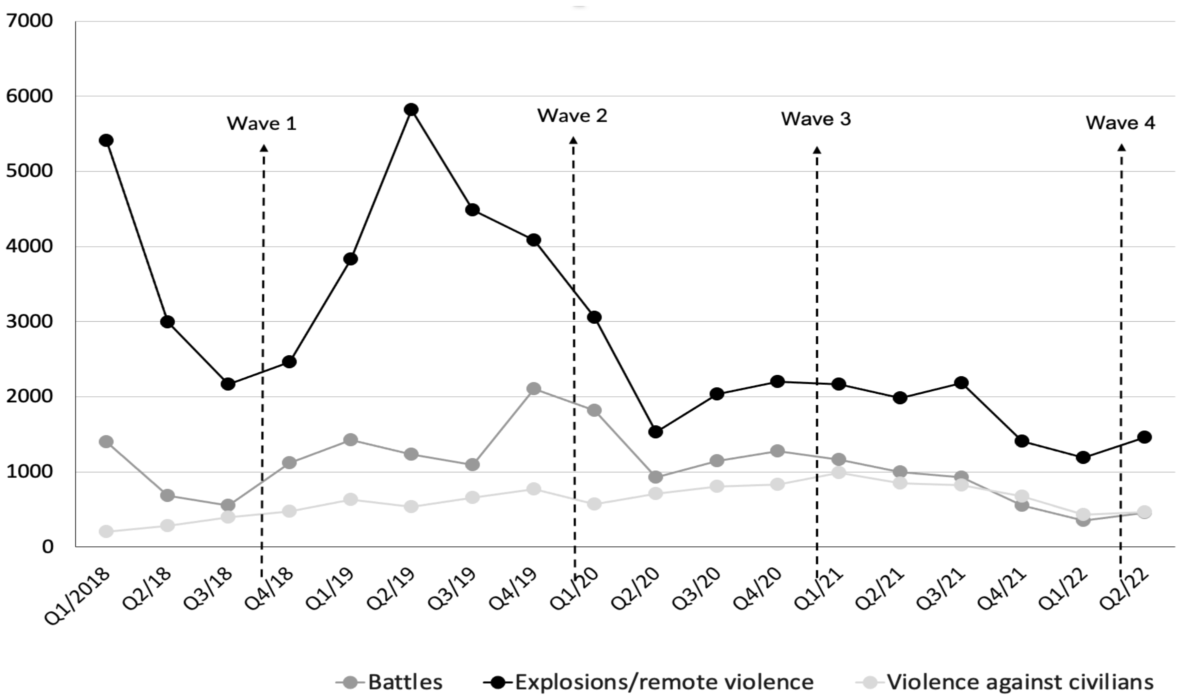
Moreover, the data is collected through PAPI (The Pen-and-Paper Personal Interview) by the local FAO teams and later transferred into the digital system. The uniqueness of data comes from its panel structure and the availability of a control group for all four waves, which can be used to balance the covariates between treated and control group households to increase the internal validity. Although pre-treatment information is not available for households receiving other than the vegetable kits intervention, we can still compare their characteristics at the same period such as Wave 3 and Wave 4. ***Table 2*** shows how households' main demographic and economic characteristics differ across control and intervention groups. We see that households that receive the poultry kits have the largest share of female household heads (67%), while those who benefited from livestock vaccination and salt blocks have the lowest share of households headed by a female. This considerable variation across different control and treatment-receiving households in terms of the gender of household head, as we observe from ***Table 2***, reflects significant differences in household head`s education level. On average, the education level of household heads is lower if there are more female-headed households in a group. For example, among households that received poultry kits, only 4.35% of them have a household head with a high school education, and almost half of the household heads are illiterate. Moreover, it is essential to note that households in each type of intervention have their family income partly from crop farming or livestock keeping. As expected, households that dominantly earn income from crop farming have, on average, higher shares of water constraints in irrigation.

Table 2. Comparison of different treatment groups with the control group, Wave 3

	Control group	Vegetable kits	Poultry kits	Beekeeping	Agricultural tools and vegetable seedlings provision	Vaccination and salt blocks
Female HHH (%)	.178	.406	.666	.316	.322	.096
Age HHH (mean)	51.394	51.568	49.391	46.621	49.355	48.152
Educ of HHH (mean)	1.964	1.758	1.580	2.114	1.967	1.764
1-Illiterate (%)	17.85	30.82	46.38	8.92	20.00	29.52
2-Below Highschool (%)	67.93	62.58	49.28	70.70	63.33	64.58
3-High school (%)	14.22	6.60	4.35	20.38	16.67	5.90
Income share: crop farming	57.929	65.415	67.159	21.316	59.111	20.188
Income share: herder	26.252	21.157	23.594	46.424	12.966	74.524
Income share: artisan	2.608	3.286	3.116	7.677	5.944	1.310
Having constraints of water in irrigation	30.02	34.91	31.88	12.66	14.44	5.90
# of observations	663	318	69	158	90	271

In addition to these imbalances of covariates across different groups of households, it should also be noted that violent attacks continued across Syria at different levels before, during and after data collection at different levels. ***Figure 2*** presents the total number of violent events during data collection. As can be seen, violent events were exceptionally high both before the first wave and between the first two waves.

Figure 2. The number of violent events in Syria



We analyze the long-term impact of the abovementioned agricultural and livestock interventions under such a humanitarian and conflict setting on household food security and resilience. We use several outcome variables to measure these. The first outcome variable is the household **food consumption scores (FCS)**, calculated using the information about the household's food consumption of 11 different food items over a seven-day recall period. It is a weighted average of the days a given household reported consuming these food items. The second outcome variable is the **household dietary diversity score (HDDS)** which is a proxy value from 1 to 12, showing how many different types of food items were consumed over the seven-day recall period without paying attention to the number of days they were eaten. Thirdly, we pay attention to a proxy for the **reduced coping strategy index (RCSI)**. Respondents were asked, "In the past 30 days, how many days has the family implemented one of the following strategies to deal with food shortages or lack of money to buy it?". Using the information provided for four different coping strategies, we calculated a weighted summation of days in the past 30 days, showing the reduced coping strategies for households as the value of the index gets larger. In addition to these important indices, we also estimated the impact of the intervention separately on **four types of coping strategies** for an over 30-day recall period, namely (1) relying on less preferred and less expensive food (i.e., cheaper, lower quality food), (2) borrowing food or relying on help from relative(s) or friend(s), (3) reducing the number of meals eaten in a day, and (4) limiting portion size at meals (i.e., less food per meal).

Table 3. Comparison of outcome variables (means) across different treatment groups and the control group, Wave 3

	Control group	Vegetable kits	Poultry kits	Beekeeping	Agricultural tools and vegetable seedlings provision	Vaccination and salt blocks
Food consumption scores (FCS)	55.626	56.633	52.362	52.082	62.172	59.637
Household dietary diversity score (HDDS)	7.530	7.028	6.391	8.261	7.755	7.875
Reduced coping strategy index (RCSI)	7.898	7.963	10.226	9.116	9.374	6.017
Relying on less preferred and less expensive food (over the last 30 days)	17.102	16.689	20	20.784	16.977	14.450
Borrowing food or relying on help (over the last 30 days)	1.617	1.544	2.043	3.728	.844	.841
Reducing the number of meals eaten in a day (over the last 30 days)	7.048	6.358	10.826	4.683	10.489	4.457
Limiting portion size at meals (over the last 30 days)	6.464	7.993	8.913	6.145	11.022	5.195

Table 3 presents the averages of outcome variables for each intervention-receiving and control group households. Unfortunately, the pre-treatment information about these outcome variables is only available for the vegetable kits receiving households. Therefore, to compare these outcome variables across different intervention groups, Table 3 calculates averages using the data in Wave 3, which is a post-treatment period for all types of interventions. This implies that the interventions have already affected the outcome variables if there were any impact. Still, it might be helpful to observe the differences across each group of households in the same period to understand if the average food security was similar across treatment groups when all have received different agricultural and livestock interventions. We observe that, in Wave 3, households who received poultry kits have higher reductions in their coping strategies and the lowest household diversity index compared to other households, on average. And there are significant differences in all outcome variables across control and treated households. However, we cannot know if there were larger differences or not in the pre-treatment period by looking at the outcome variables in Wave 3. Moreover, these are just the unconditional averages of outcome variables, so one needs to check

if the difference across groups is still statistically significant once we balance treated and control units and also condition on covariates and time-invariant spatial fixed effects.

Moreover, as already mentioned, we have pre-treatment information for the households who have received the vegetable kits. We analysed the impact on the outcome variables mentioned above using this panel data with baseline information. Moreover, we also evaluated the impact on several other variables about the households` resilience. Firstly, we analysed if the vegetable kits intervention had any impact on ***child labour*** and ***child marriage***. The questionnaire asked respondents if any children (under 16) started working because the family needed this coping strategy to deal with food shortages or lack of money to buy it. We created a dummy variable from responses to this question which is equal to 1 if the household had a working child and 0 if the household did not need to use this strategy even though there were children aged below 16. Secondly, we created a dummy variable for households that reported that a young girl (aged below 16) had to marry to ease the financial stress on the family. Secondly, we checked whether several other coping strategies had to be used by the family, which are (1) the ***sale of household assets***, (2) the ***sale of productive assets***, (3) the ***sale of food aid***, (4) ***sale of non-food humanitarian assistance*** and (5) ***taking credit to access food***. ***Table 4*** below compares these variables across control and treated households (those who received the vegetable kits) in the pre-treatment period. As can be seen, child marriage and child labour are significantly higher among the treated units. However, the sale of household assets and food aid is slightly lower in treated households in the pre-treatment period. It is important to note that these dummy variables are created by dropping the households who reported that they do not have the ability to use any of these coping strategies. Therefore, if a household does not have non-food humanitarian assistance, then it is dropped from the sample when we construct the dummy variable for the sale of non-food humanitarian aid. Finally, ***Table 4*** shows that the share of households that reported taking credit for food access is very high in treated and control units, above 80 per cent.

Table 4. Comparison of outcome variables (means) across treatment and control groups, pre-treatment

	Control		Treated	
	Mean	Sample size (Wave 1 only)	Mean	Sample size (Wave 1 only)
Child labor	.243	317	.366	93
Child marriage	.223	296	.301	93
Sale of household assets	.288	375	.400	115
Sale of productive assets	.386	368	.369	103
Sale of food aid	.325	274	.260	96
Sale of non-food humanitarian assistance	.183	235	.191	89
Take credit to access food	.840	481	.843	166

7. Long-term impacts of agricultural and livestock interventions

7.1. Panel data analysis for the "vegetable kits" intervention

For vegetable kit receivers, we have data for four waves, including one pre-treatment period information (baseline). ***Figure 3*** shows how the average food consumption score (FCS) changes in control and treatment groups across four waves of data. As can be seen, there is a significant gap in FCS in Wave 1 (pre-treatment period), which is almost closed right after the treatment intake and this small difference that is obtained in Wave 2 (post-treatment period) is maintained in later waves. This implies that the short-term change in FCS is also sustained in the long term. ***Figure 4*** presents the average treatment effect in all post-treatment periods, and, as we discussed, they are all positive and very close to each other in magnitude.

Figure 3. Trends of Food Consumption Score in control and treatment group

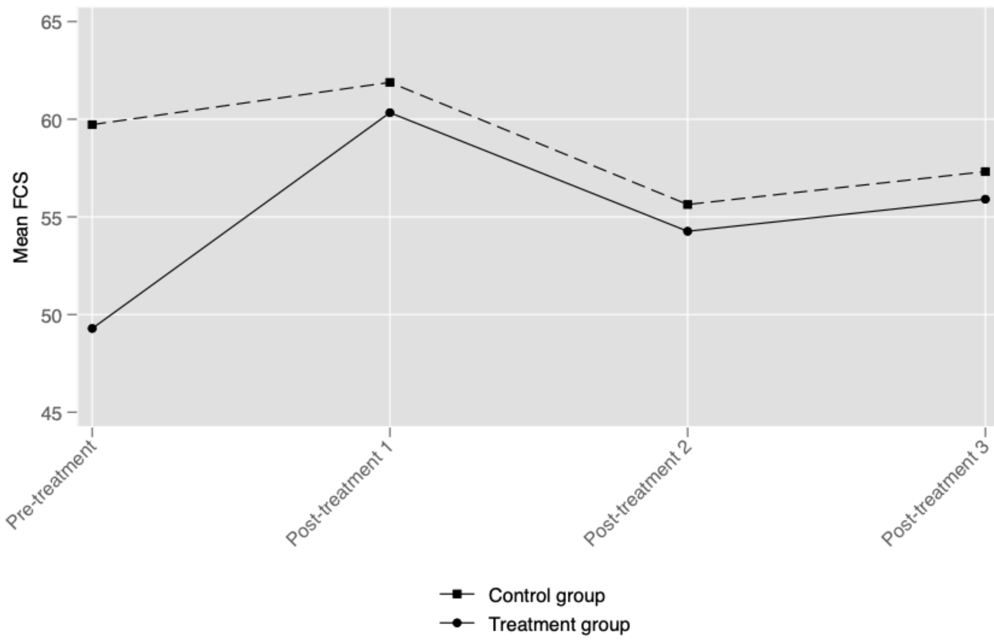
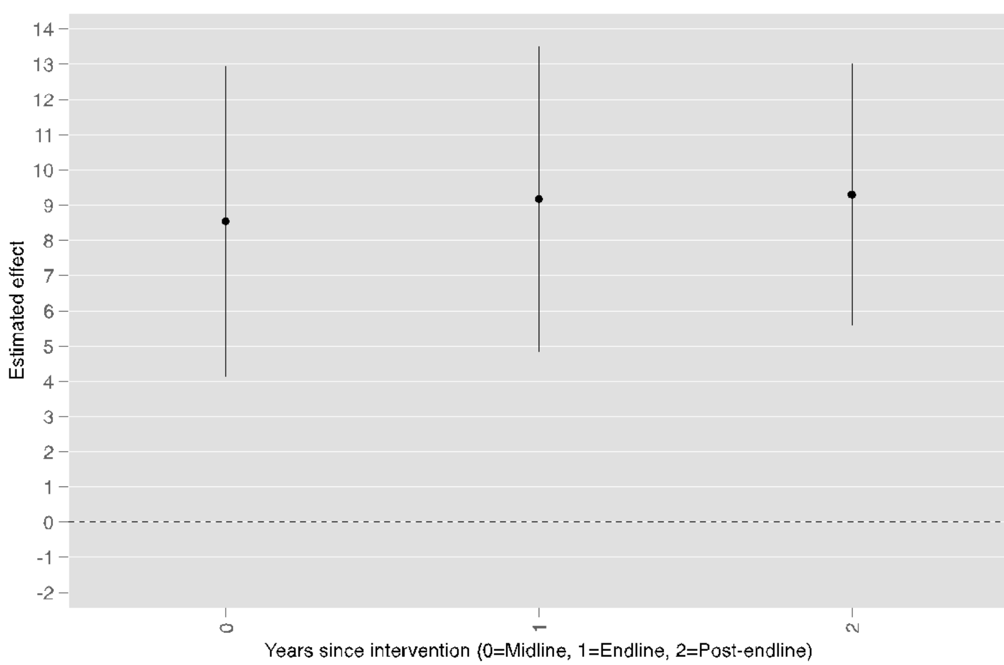


Figure 4. Average treatment effect in post-treatment periods



In addition to the graphical visualisation, we can also check the quantification of the average impact of the intervention, controlling for covariates. Employing the difference-in-differences method, ***Table 5*** shows that once we control for the gender, age and education of the household head and the sub-district fixed effects, the food consumption score in the treatment group is, on average, raised by around 9 points (corresponding to an 18% increase in the FCS observed in Wave 1 in the treatment group). Notably, the event-study coefficients show that this increase is kept even three years after the intervention and does not only belong to a short-run increase in the FCS.

Table 5. Average Treatment Effect of the Vegetable Kits Intervention on Food Consumption Score

	Diff-in-Diff (DiD)	Event-study Analysis
ATET	8.920***	
	(1.577)	
1-year ATET		8.534***
		(2.244)
2-year ATET		9.174***
		(2.201)
3-year ATET		9.295***
		(1.887)
Covariates	HHH gender HHH age HHH educ	HHH gender HHH age HHH educ
Subdistrict FE	Yes	Yes
# of Obs.	3,060	3,060
R-squared	0.285	0.5208

Notes. HH-clustered standard errors in parenthesis. ***p-value<0.001

However, in cases where there exist significant imbalances between control and treatment groups or, in other words, if randomisation in treatment assignment is not possible or not done perfectly in the application, then researchers should use other methods to make treatment and control groups comparable ex-post. There are different ways to check the balance in the data: statistical tests such as the Kolmogorov-Smirnov test statistic, which compares distributions for covariates across treatment and control groups, or t-test statistics comparing the means of covariates or simply using balance plots. ***Table 6*** shows statistically significant imbalances even at basic characteristics and the food consumption scores of the households in the pre-treatment period (Wave 1). This makes the difference-in-difference estimates questionable.

Table 6. Pre-treatment Balance in the Raw Data (only vegetable kits treatment and the control group in the panel data)

	Treated	Control	Diff
Female HHH (%)	.322	.155	.166***
Age HHH (mean)	49.293	49.132	.160
Educ of HHH (mean)	1.760	1.907	-.146***
1-Illiterate (%)	31.22	20.71	
2-Below Highschool	61.46	67.86	
3-High school	7.32	11.43	
FCS	49.239	59.997	-10.768***
No. of Obs.	205	560	-

We used various matching and weighting techniques to eliminate this imbalance and make treatment and control groups similar to each other to compare them. Using the same model specification as used above, ***Table 7*** presents the results. The coefficient in propensity score matching (PSM) DiD shows that the average treatment effect on the treated increases the FCS by 5.2 points. However, it decreases the sample size almost by half. Once we use the Coarsened Exact Matching (CEM), the decrease in the sample size is not that large, and the coefficient estimate is similar to the DiD estimate we found above. Finally, we also used the Kernel propensity-score matching (Kernel PSM) DiD, which uses propensity scores to calculate kernel weights which are then used in the estimation. Thus, there are no decreases in the sample size, and the coefficient estimation is larger than the CEM estimates.

Table 7. ATET using matching and weighting techniques

	PSM	CEM	Kernel PSM
ATET	5.165***	8.261***	9.692***
	(1.890)	(1.648)	(1.622)
Covariates used in matching/weighting	HHH gender HHH age HHH educ	HHH gender HHH age HHH educ	HHH gender HHH age HHH educ
Subdistrict FE	Yes	Yes	Yes
# of Obs.	1,576	2,876	3060
R-squared	0.288	0.291	0.300

Notes. HH-clustered standard errors in parenthesis. ***p-value<0.01

As discussed in the methodology section, matching is an iterative process and does not always guarantee the balance between the treatment and control groups. Moreover, it is subjective to the researchers to decide which covariates should be included in the matching procedure. Our simple exercise shows that satisfying balance is challenging even in a simple model case with a few covariates. ***Table 8, 9*** and ***10*** below present the balance test results between the treatment and control groups. Even in cases where the balance between covariates is satisfied such as in the Kernel PSM model, there are large differences across the ex-ante average food consumption scores.

Table 8. Pre-treatment Balance in the Propensity Score Matched Data (only vegetable kits treatment and the control group in the panel data)

	Treated	Control	Diff
Female HHH (%)	.322	.155	.147***
Age HHH (mean)	49.293	49.132	.160
Educ of HHH (%)	1.760	1.778	-.017
FCS	49.239	54.320	-5.090**
No. of Obs.	205	189	

Table 9. Pre-treatment Balance in the Coarsened Exact Matched Data (only vegetable kits treatment and the control group in the panel data)

	Treated	Control	Diff
Female HHH (%)	.319	.152	.167***
Age HHH (mean)	49.372	49.909	-.536
Educ of HHH (%)	1.755	1.856	-.100**
FCS	49.318	55.850	-6.531***
No. of Obs.	204	557	

Table 10. Pre-treatment Balance in the Kernel PSM Data (only vegetable kits treatment and the control group in the panel data)

	Treated	Control	Diff
Female HHH (%)	.322	.313	0.009
Age HHH (mean)	49.293	49.921	-0.503
Educ of HHH (%)	1.761	1.780	-0.019
FCS	49.229	60.146	-10.917***
No. of Obs.	205	560	

As the Kernel PSM DiD works best in balancing the covariates, we used it with additional covariates. ***Table 11*** presents the results for the food consumption score. In Model 2, we also use kernel weights created using the baseline data for the income shares of households from different economic activities in addition to gender, age, and education levels of household heads. As a result, the ATET slightly increases compared to Model 1. In Model 3, we also use the total number of shocks experienced by households one year before Wave 1(pre-treatment period) in our kernel weight estimations. All regressions include the sub-district fixed effects to control for any time-invariant but subdistrict-specific common shocks such as environmental or violence shocks. Our estimates in Model 3 show that the ATET is larger once we balance treatment and control groups according to the shocks they experienced. However, we need to keep in mind that sample size decreases Model 3 because of missing variables in the shocks variable. These estimates show that the increase in the food consumption score is estimated to be from 18.12% to 26.83%.

Table 11. ATET using Kernel Propensity Score Matching Difference-in-differences estimation

	Model 1	Model 2	Model 3
ATET	9.692***	10.053***	13.210***
	(1.622)	(1.629)	(1.773)
Covariates used in matching/weighting	HHH gender HHH age HHH educ	HHH gender HHH age HHH educ Share of income from: (1) Farming (2) Herding (3) Beekeeping (4) Labour (5) Artisanship	HHH gender HHH age HHH educ Share of income from: (1) Farming (2) Herding (3) Beekeeping (4) Labour (5) Artisanship Total number of shocks experienced in the past year
Subdistrict FE	Yes	Yes	Yes
# of Obs.	3060	3000	2172
R-squared	0.300	0.300	0.320

Notes. HH-clustered standard errors in parenthesis. ***p-value<0.01

However, even though we balance the covariates in the Kernel PSM DiD models, the significant imbalance in food consumption score in baseline might still cause bias in our estimates as it shows that treated and control groups were not indeed comparable to each other. This is expectedly the case if the interventions are targeted to the households who are particularly in need. Such targeting issues exist in humanitarian emergencies and conflict settings. Thus, to estimate the intervention's average treatment effect, one needs to find out the ‘hidden’ experiment from the observational data and start the impact evaluation with two groups with a similar pre-treatment trend of the outcome variable. However, even one period of pre-treatment data is scarce in HECS, let alone the multiple pre-treatment periods. And, in cases where we cannot check the trends between control and treated groups, it is plausible that the outcome variable has a similar distribution in the available pre-treatment period. The following section provides the estimations for the average treatment effect on the treated using a balancing method suggested recently in the literature.

7.2. Entropy Balancing as a way to find the hidden experiment in observational studies

Entropy balancing reweights the data from the control units to match a set of moments computed from the treated units' data (Hainmueller, 2012). In this section, the means and variances of specified covariates and the outcome variable in the treatment group data are calculated using the baseline data, and a set of entropy weights are computed such that the means and variances in the reweighted control group data match the means and variances from the treatment group by design. The advantage of entropy balancing is that it does not require researchers to use the iterative process of covariate balancing used in conventional matching methods such as Mahalanobis distance matching or coarsened exact matching. Moreover, the latter's sample size decreases in large proportions due to pruning, which is not necessarily the case for entropy balancing. Finally, targeting in HECS might result in an economically and statistically significant gap in the outcome variable. As shown in the examples in the previous section, this imbalance is not eliminated even if we use propensity score matching or other distance matching methods. Thus, entropy balancing helps us estimate the ATET even in cases where the conditional independence assumption of causal inference is violated.

Using our data with vegetable kits receiving households and control groups, ***Figure 5*** shows that weights estimated in the entropy balancing algorithm equalise the average food consumption score in the pre-treatment period. Similar to our main analysis, we observe a positive impact on FCS in the short term, which is sustained in the long term. ***Table 12*** summarises the results of DiD with entropy balancing. It shows that in Model 1, where we apply entropy balancing using the FCS and household head’s gender, age and education level in the baseline data, the long-term average treatment effect is 3.54, corresponding to about a 7.2% increase in the FCS at baseline. When we also balance treated and control group households according to the economic activity types, the impact is still positive but slightly smaller (about a 5% increase in FCS). Thus, agricultural emergency assistance had a positive effect both in the short-term and long term.

Figure 5. Trends between treated and untreated groups using the Entropy Balancing with the pre-treatment data

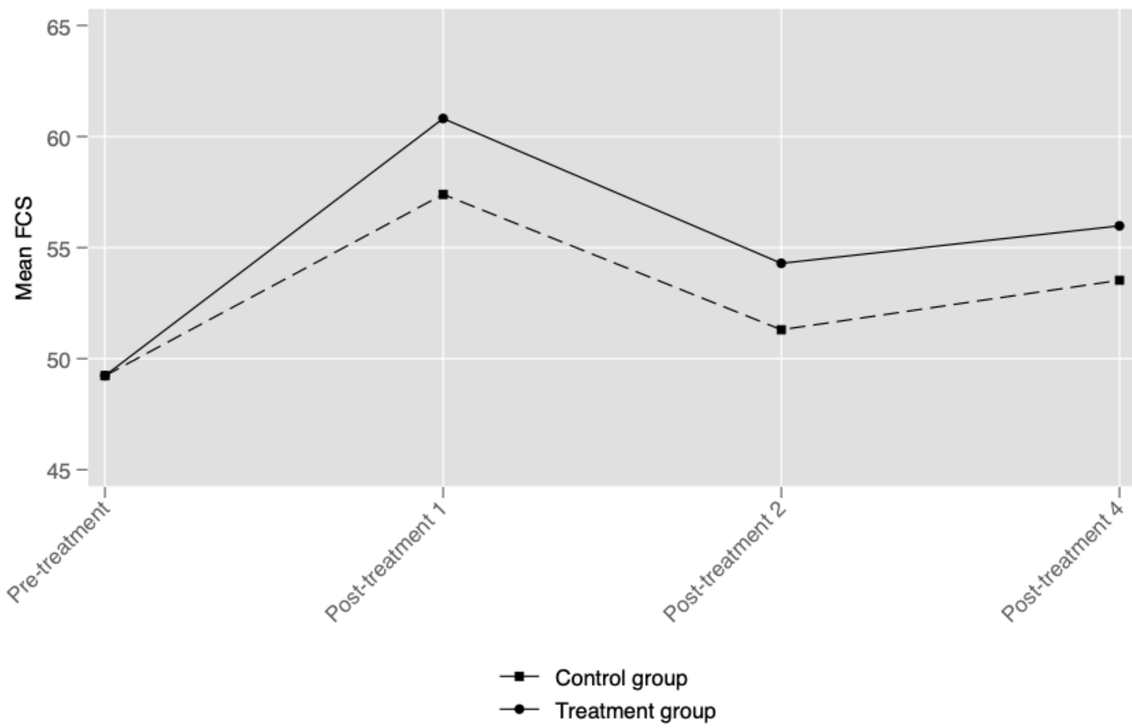


Table 12. Long-term ATE on Food Consumption Score (FCS) using Entropy Balancing

	(1)	(2)
ATE	3.539**	2.414*
	(1.594)	(1.489)
Variables used in entropy balancing	FCS HHH gender HHH age HHH educ	FCS HHH gender HHH age HHH educ Share of income from: (1) Farming (2) Herding (3) Beekeeping (4) Labour (5) Artisanship
Subdistrict FE	Yes	Yes
# of Obs.	3,060	3,060
R-squared	0.299	0.331

Notes. ***p-value<0.01, **p-value<0.05, *p-value<0.10

As we already discussed, in HECS, panel data is very scarce, and in rare situations where we have data, it is usually only from the post-treatment period. Thus, our panel data of four waves with pre-treatment information is unique in this setting. This implies that methods of matching and balancing with the pre-treatment data are only plausible in some cases. Using our panel data on vegetable kits intervention, we show below that using entropy balancing even in the post-treatment period gives us similar ATEs. We use entropy balancing weights estimated using the baseline data in the first two columns. However, we then dropped the baseline data and estimated the ATE using only the post-treatment data. We repeated the same exercise by estimating the entropy balancing weights using the Wave 2 covariates. As can be seen, the coefficient estimates in models with the same covariates are very similar. Entropy balancing helps us to find the ATE even in cases where there is no baseline data. However, it should be noted that we estimated the ATE using the post-treatment periods. Thus, it omits the short-term (between Wave 1 and Wave 2, in our case) impact we observed right after the intervention until Wave 2. The short-term positive effect on the food consumption score of treated households is not reversed in the long term in our case, and that's why the estimates in ***Table 13*** are also close to the ones in ***Table 12***.

Figure 6. Average food consumption score in post-treatment periods with Entropy Balancing weights

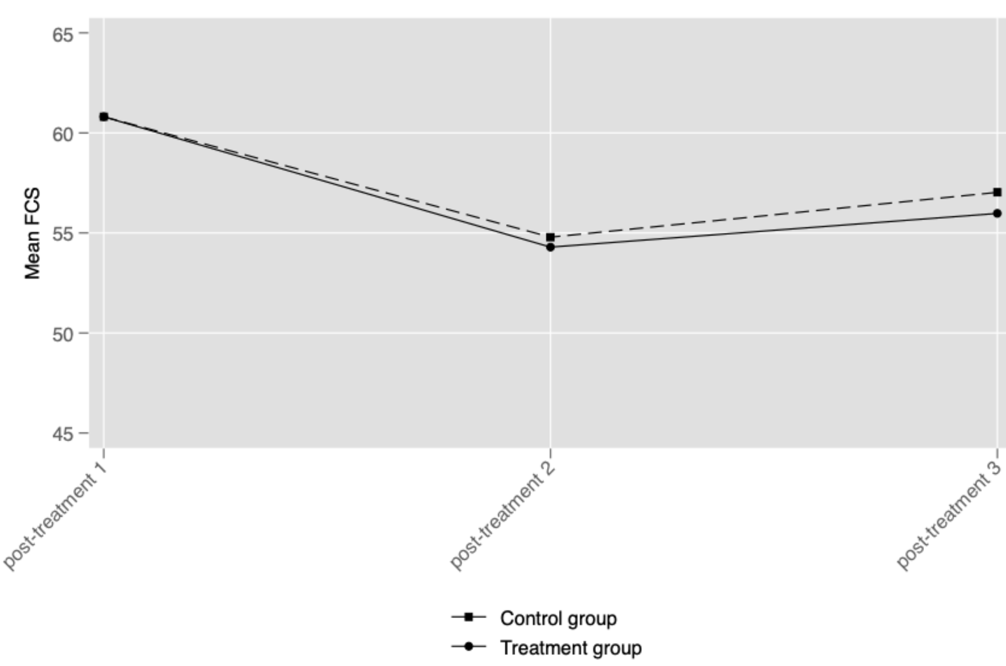


Table 13. ATET using Entropy Balancing WITHOUT BASELINE (pre-treatment) information

	Using Baseline (Wave 1) Entropy weights	Using Baseline (Wave 1) Entropy weights	Using Post-treatment (Wave 2) Entropy weights	Using Post-treatment (Wave 2) Entropy weights
ATET	2.346** (.926)	3.652*** (.938)	2.373*** (.934)	3.165*** (.922)
Variables used in entropy balancing	FCS HHH gender HHH age HHH educ	FCS HHH gender HHH age HHH educ Share of income from: (1) Farming (2) Herding (3) Beekeeping (4) Labour (5) Artisanship	FCS HHH gender HHH age HHH educ	FCS HHH gender HHH age HHH educ Share of income from: (1) Farming (2) Herding (3) Beekeeping (4) Labour (5) Artisanship
Subdistrict FE	Yes	Yes	Yes	Yes
# of Obs.	2,295	2,295	2,295	2,295
R-squared	0.343	0.379	0.338	0.368

Notes. ***p-value<0.01, **p-value<0.05, *p-value<0.10

In addition to analysing the impact of the 'vegetable kits' intervention on food consumption score, we also checked if the same intervention had any impact on the household dietary diversity scores (HDDS) and the reduced coping strategy index (RCSI). HDDS are calculated using the information about the household's food consumption of 11 different food groups. The higher the HDDS, the better the dietary diversity. RCSI, on the other hand, is calculated by weighing the number of days households used different coping strategies to deal with food shortages or lack of money to buy food in the last 30 days. Thus, a higher RCSI implies more days with coping strategies to deal with food shortages.

Figure** 7* and ***Figure 8 present the trends in HDDS and RCSI in treated (vegetable kit receivers) and control groups. We used entropy balancing weights calculated with the pre-treatment information on their outcome variables and other basic household characteristics described in ***Table 14***. We see that, on average, HDDS slightly decreases in the treated group. The most considerable difference between the treated and control groups is observed two years after the treatment, which then recovered. Although the ATE on HDDS was very small but negative, ***Figure 8*** shows that the weighted number of days that households had to employ coping strategies to deal with food shortages decreased with the treatment both in the short and long run.

Figure 7. HDDS trends between treated and untreated groups using the Entropy Balancing with the pre-treatment data

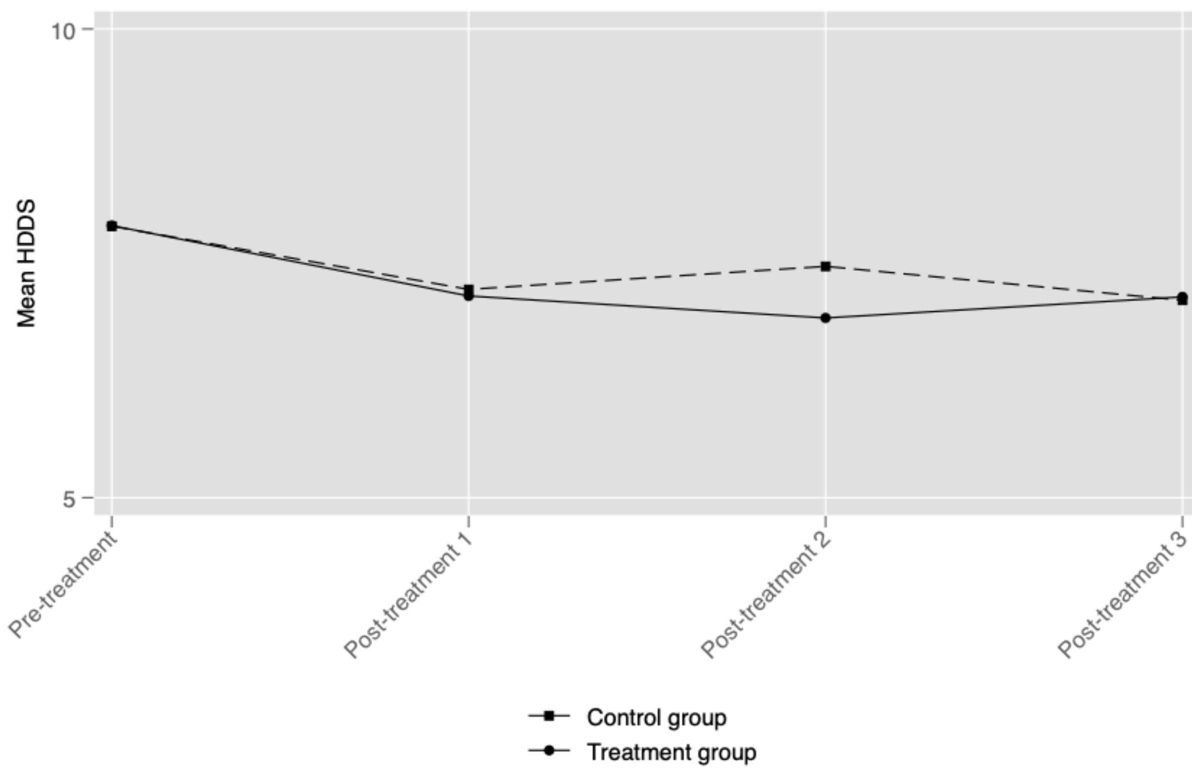


Table 14 presents the difference-in-differences coefficients with the weighted sample obtained with the entropy balancing algorithm. It shows that the HDDS of treated households decreases by .29. This is a considerably small decrease compared to the mean of 7.90 HDDS for the treated group in the baseline. When we estimate the ATE of vegetable kits intervention on the RCSI, we find that treated households, on average, had to employ coping strategies such as borrowing food or relying on friends and relatives 1.4 days lower in a month. Knowing that the average RCSI was 7.415 for the treated group in the baseline, this decrease in RCSI corresponds to an 18.4% improvement for the treated households.

Figure 8. RCSI trends between treated and untreated groups using the Entropy Balancing with the pre-treatment data

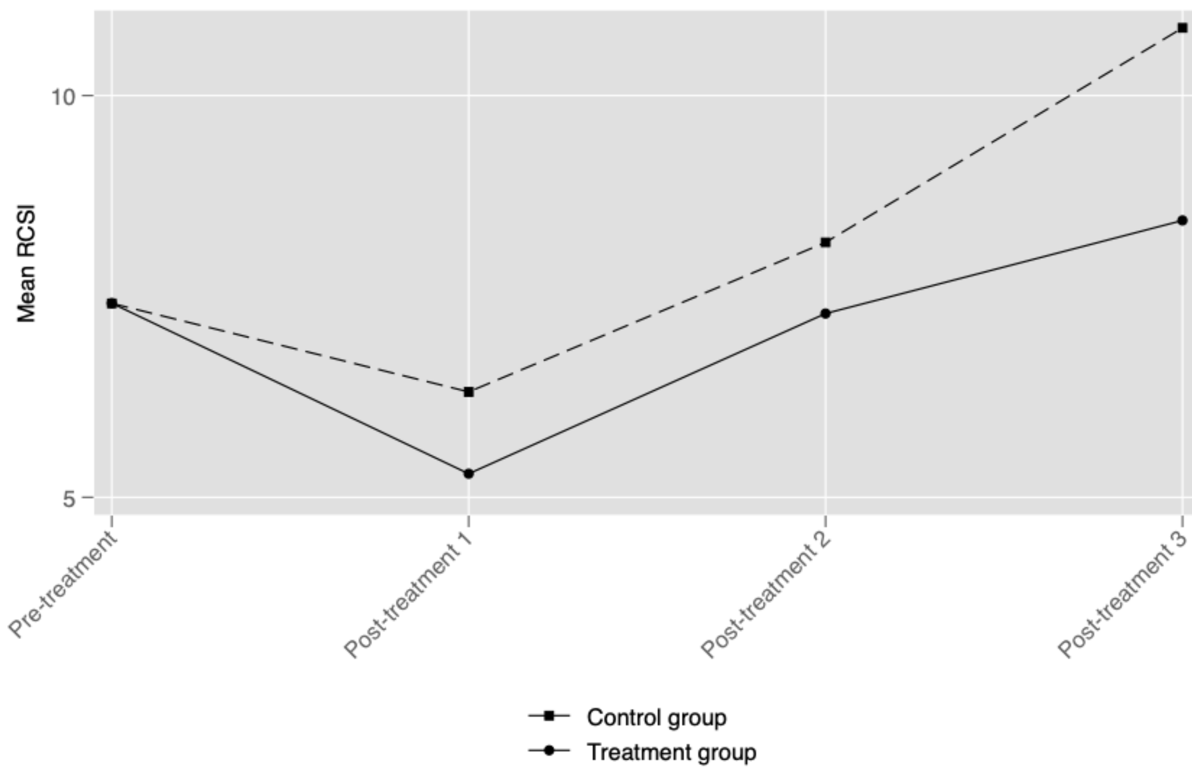


Table 14. Long-term Average Treatment Effect (ATE) on HDDS and RCSI using Entropy Balancing

	HDDS	HDDS	RCSI	RCSI
ATE	-.279* (.171)	-.289* (.171)	-1.400*** (.536)	-1.362** (.547)
Variables used in entropy balancing	RCSI HHH gender HHH age HHH educ	HDDS HHH gender HHH age HHH educ Share of income from: (1) Farming (2) Herding (3) Beekeeping (4) Labour (5) Artisanship	RCSI HHH gender HHH age HHH educ	RCSI HHH gender HHH age HHH educ Share of income from: (1) Farming (2) Herding (3) Labour (4) Artisanship
Subdistrict FE	Yes	Yes	Yes	Yes
# of Obs.	3,060	3,060	3,060	3,060
R-squared	0.257	0.289	0.232	0.251

Notes. HH-clustered standard errors in parenthesis. ***p-value<0.01

7.3. Long-term impact of vegetable kits intervention on household resilience

In addition to the long-term impact of vegetable kits intervention on food consumption score, HDDS and RCSI, we also analyse if the intervention affected the resilience of treated households by examining the use of undesirable coping strategies to access food. To do so, we analysed the four waves of panel data for vegetable kits intervention, where we also had the pre-treatment information. We used Kernel propensity score matching DiD as the method of identification instead of entropy balancing in this section. The reason is that our sample sizes were already small in the treatment groups because only some households in our sample could use coping strategies. For example, if no children are in a household, they cannot have child marriage or child labour, so they are dropped from the analysis. Or if a household did not receive any non-food humanitarian assistance, then they will not be able to sell them and so on. Thus, using entropy balancing with an already small sample decreased our sample sizes further to enforce the balance of covariates across treated and control units.

Using the Kernel PSM DiD method, ***Table 15*** below shows that both the probability of child labour and child marriage decreased in the treated households. Moreover, as seen in Table 16, we find that treated households also have a lower probability of taking credit to access food or selling household (productive) assets in cases of food shortages. Thus, our results show that short-term oriented agricultural input provision can also have long-term impacts on households' welfare. Finally, estimation coefficients presented in ***Table 17*** show that treated households' probability of selling food aid or other non-food humanitarian assistance in food shortages or lack of money to buy food decreased compared to the control group.

Table 15. Long-term Average Treatment Effect (ATE) of Vegetable Kits Intervention on Household Resilience using Kernel PSM Difference-in-differences

	Child Labor (Kernel PSM DiD)	Child Marriage (Kernel PSM DiD)
ATET	-0.156***	-0.130*
	(0.056)	(0.068)
Subdistrict FE	Yes	Yes
# of Obs.	1156	461
R-squared	0.340	0.380

*Notes. ***p-value<0.01, **p-value<0.05, *p-value<0.10; Variables used in calculating kernel weights: rCSI, HHH gender, HHH age, HHH educ, Child Labor, Child Marriage, Sale of HH Assets, Sale of Productive Assets, Sale of Food Aid.*

Table 16. Long-term Average Treatment Effect (ATE) of Vegetable Kits Intervention on Household Resilience using Kernel PSM Difference-in-differences

	Take Credit to access Food (Kernel PSM DiD)	Sale of HH Assets (Kernel PSM DiD)	Sale of Productive Assets (Kernel PSM DiD)
ATET	-0.172**	-0.217***	-0.302***
	(.067)	(.062)	(.057)
Subdistrict FE	Yes	Yes	Yes
# of Obs.	1373	593	1304
R-squared	0.31	0.400	0.140

*Notes. ***p-value<0.01, **p-value<0.05, *p-value<0.10; Variables used in kernel weights: rCSI, HHH gender, HHH age, HHH educ, Child Labor, Child Marriage, Sale of HH Assets, Sale of Productive Assets, Sale of Food Aid.*

Table 17. Long-term Average Treatment Effect (ATE) of Vegetable Kits Intervention on Household Resilience using Kernel PSM Difference-in-differences

	Sale of Food Aid (Kernel PSM DiD)	Sale of Humanitarian Assistance (Kernel PSM DiD)
ATET	-0.170***	-0.111**
	(.053)	(.046)
Subdistrict FE	Yes	Yes
# of Obs.	1097	1023
R-squared	0.17	0.24

*Notes. ***p-value<0.01, **p-value<0.05, *p-value<0.10; Variables used in kernel weights: rCSI, HHH gender, HHH age, HHH educ, Child Labor, Child Marriage, Sale of HH Assets, Sale of Productive Assets, Sale of Food Aid.*

7.4. Long-term impact evaluation using only post-treatment data

In the sub-sections above, we focused on the long-term impact of vegetable kit provision. The data we analysed had the pre-treatment information so that we could use several methodologies to test the average impact of the treatment. The FAO program has also provided other agricultural and livestock interventions. It is also interesting to know the impacts of these two groups of interventions to understand the effectiveness of the overall programming better.

There are six separate models in our estimations in this section. We first check the impact of interventions on the 'levels' of our outcomes. In this specification, we group all beneficiaries into one overarching 'treatment' group without paying attention to whether treated households have received agricultural support, such as the provision of vegetable seedlings, or livestock support, such as livestock vaccination. Second, we check the impact of all agricultural and livestock interventions separately. After the impact analyses at the levels of outcome variables in Wave 4, we then test the impact of receiving any intervention or only agricultural or livestock intervention on the change in our outcomes between Wave 3 and Wave 4, the last two waves in our panel data. Their results are presented in the last three columns of the tables below, which show the results for different outcome variables.

Our analysis used the entropy method to create balancing weights with the post-treatment data available for control and treatment group households. Balancing is done for the mean and variances of age, gender and education of household heads. As this is a cross-sectional analysis, we did not use the outcome variables in the entropy balancing algorithm. Moreover, we paid attention not to control for any variable which might have been affected by the intervention. The estimations still include all these variables in the regression model except the potential colliders. We also controlled for the total number of shocks households experienced for the last 1-year recall period, each household's treatment duration, and the interaction of the duration with the treatment type. Moreover, estimation models include sub-district fixed effects to control for any observable subdistrict-specific effects households might have, such as environmental shocks or the severity of the conflict.

Table 18. Long-term Analysis for the FCS with Entropy Balancing

	FCS	FCS	FCS	Diff FCS	Diff FCS	Diff FCS
Treatment	-2.498			4.714		
	(3.009)			(3.324)		
Agri. Treatment		10.682***			4.217	
		(3.725)			(4.887)	
Livestock Treatment			-10.995**			7.042
			(0.494)			(5.149)
Duration of treatment	Yes	Yes	Yes	Yes	Yes	Yes
Duration X treatment type	No	Yes	Yes	No	Yes	Yes
Covariates	Yes	Yes	Yes	Yes	Yes	Yes
Subdistrict FEs	Yes	Yes	Yes	Yes	Yes	Yes
# of Obs.	1,040	736	759	1,011	730	732
R-squared	0.450	0.475	0.475	0.391	0.394	0.433

Notes. Covariates are gender, age and education level of household heads, household size, number of shocks, income shares from different economic activities, land restriction dummy, and water constraint dummy. Diff FCS is the difference between Wave 4 and Wave 3 levels of FCSs.
**p-value<0.05, ***p-value<0.01

Table 18 shows no impact on the FCS when we compare all treated households with the control group households. However, columns 2 and 3 show that the null effect is because the negative effect of the livestock interventions cancels the positive impact of agricultural interventions. We find no significant difference when we look at the differences in these effects in the last two periods of our panel data, implying that the level effects we saw in Wave 4 were similar in the medium term.

The positive impact of agricultural interventions was also relevant for the household dietary diversity scores (HDDS), as seen in ***Table 19***. Column 2 shows that households who received agricultural interventions on average have consumed an additional type of food compared to control group households. Moreover, livestock interventions also positively impact HDDS in the last wave, giving us an overall positive impact of being in the treatment group. Furthermore, the last three columns in ***Table 19*** show that the HDDS of treated households also improved between Wave 3 and Wave 4. So, there were statistically significant and positive long-term effects of either having agricultural or livestock interventions on the HDDS.

Table 19. Long-term Analysis for the HDDS with Entropy Balancing

	HDDS	HDDS	HDDS	Diff HDDS	Diff HDDS	Diff HDDS
Treatment	.807*** (.296)			2.263*** (.413)		
Agri. Treatment		1.099*** (.400)			1.220** (.538)	
Livestock Treatment			1.100** (.499)			4.636*** (.758)
Duration of treatment	Yes	Yes	Yes	Yes	Yes	Yes
Duration X treatment type	No	Yes	Yes	No	Yes	Yes
Covariates	Yes	Yes	Yes	Yes	Yes	Yes
Subdistrict FEs	Yes	Yes	Yes	Yes	Yes	Yes
# of Obs.	1,040	736	759	1,011	730	732
R-squared	0.356	0.359	0.399	0.390	0.388	0.435

Notes. Covariates are gender, age and education level of household heads, household size, number of shocks, income shares from different economic activities, land restriction dummy, and water constraint dummy. Diff HDDS is the difference between Wave 4 and Wave 3 levels of HDDS. *p-value<0.10, **p-value<0.05, ***p-value<0.01

In ***Table 20,*** we show that treated households, on average, need to use coping strategies at a lower extent to deal with food shortages. When we look at the impact of agricultural and livestock interventions separately, we observe that households who received livestock interventions spent, on average, 8.7 days less in using coping strategies for food shortages in Wave 4. On the other hand, the decrease in RCSI is 3.3 days for households that received agricultural interventions. Moreover, we find a further reduction in RCSI from Wave 3 to Wave 4 for households who received livestock treatment.

Table 20. Long-term Analysis for the Reduced Coping Strategy Index with Entropy Balancing

	RCSI	RCSI	RCSI	Diff RCSI	Diff RCSI	Diff RCSI
Treatment	-2.604**			-.685		
	(1.062)			(1.346)		
Agri. Treatment		-3.267***			-2.346	
		(1.273)			(1.898)	
Livestock Treatment			-8.745***			-5.727**
			(2.048)			(2.332)
Duration of treatment	Yes	Yes	Yes	Yes	Yes	Yes
Duration X treatment type	No	Yes	Yes	No	Yes	Yes
Covariates	Yes	Yes	Yes	Yes	Yes	Yes
Subdistrict FEs	Yes	Yes	Yes	Yes	Yes	Yes
# of Obs.	1,060	751	772	1,060	751	772
R-squared	0.448	0.504	0.487	0.442	0.501	0.461

Notes. Covariates are gender, age and education level of household heads, household size, number of shocks, income shares from different economic activities, land restriction dummy, and water constraint dummy. Diff RCSI is the difference between Wave 4 and Wave 3 levels of RCSI. *p-value<0.10, **p-value<0.05, ***p-value<0.01

In addition to analysing the impact on the RCSI, we also checked the impact of each intervention type separately on the components of RCSI. The results are presented in the tables below. Firstly, we found that livestock interventions, on average, have a significant and considerable impact on the days spent relying on less preferred and less expensive food. ***Table 21*** displays that those households which received agricultural interventions had to rely on less preferred and less costly food in case of food shortages by 17.17 days lower than the control group households. On the other hand, when we analyse the impact of intervention types on food borrowing, as ***Table 22*** shows, we find that agricultural intervention receivers have a lower number of days borrowing food compared to the control group household. However, livestock intervention has no statistically significant impact on food borrowing.

Table 21. Long-term Analysis for Relying on less preferred and less expensive food (LPLE) with Entropy Balancing

	LPLE	LPLE	LPLE	Diff LPLE	Diff LPLE	Diff LPLE
Treatment	-4.654***			-3.255		
	(1.588)			(2.051)		
Agri. Treatment		-.588			1.305	
		(1.619)			(2.457)	
Livestock Treatment			-17.166***			-15.379***
			(2.936)			(3.654)
Duration of treatment	Yes	Yes	Yes	Yes	Yes	Yes
Duration X treatment type	No	Yes	Yes	No	Yes	Yes
Covariates	Yes	Yes	Yes	Yes	Yes	Yes
Subdistrict FEs	Yes	Yes	Yes	Yes	Yes	Yes
# of Obs.	1,060	751	772	1,060	751	772
R-squared	0.411	0.426	0.434	0.516	0.569	0.530

Notes. Covariates are gender, age and education level of household heads, household size, number of shocks, income shares from different economic activities, land restriction dummy, and water constraint dummy. Diff LPLE is the difference between Wave 4 and Wave 3 levels of LPLE. *p-value<0.10, **p-value<0.05, ***p-value<0.01

Table 22. Long-term Analysis for borrowing food or relying on help from relatives or friends (# of days) using Entropy Balancing

	BF	BF	BF	Diff BF	Diff BF	Diff BF
Treatment	-2.610***			-1.272		
	(.923)			(1.122)		
Agri. Treatment		-4.572***			-4.710***	
		(1.392)			(1.597)	
Livestock Treatment			-1.131			1.852
			(1.720)			(2.148)
Duration of treatment	Yes	Yes	Yes	Yes	Yes	Yes
Duration X treatment type	No	Yes	Yes	No	Yes	Yes
Covariates	Yes	Yes	Yes	Yes	Yes	Yes
Subdistrict FEs	Yes	Yes	Yes	Yes	Yes	Yes
# of Obs.	1,060	751	772	1,060	751	772
R-squared	0.258	0.309	0.261	0.187	0.249	0.178

Notes. Covariates are gender, age and education level of household heads, household size, number of shocks, income shares from different economic activities, land restriction dummy, and water constraint dummy. Diff BF is the difference between Wave 4 and Wave 3 levels of BF. *p-value<0.10, **p-value<0.05, ***p-value<0.01

Furthermore, we show in ***Table 23*** that livestock interventions also decreased the number of days households had to reduce the number of meals eaten in a day by 8.2 days on average compared to control group households in Wave 4. However, we found no impact of agricultural interventions on this outcome variable either in Wave 3 or the change of its level from 2021 to 2022. Relatedly, as ***Table 24*** shows, livestock intervention receivers limit their portion sizes, on average, 9.8 days less in Wave 4 compared to the control group households. Again, agricultural interventions are not found to have a statistically significant impact on either reduction in the number of meals or portion size.

Table 23. Long-term Analysis for Reducing the number of meals eaten in a day (Number of days) using Entropy Balancing

	RNM	RNM	RNM	Diff RNM	Diff RNM	Diff RNM
Treatment	.231 (1.562)			3.110 (2.180)		
Agri. Treatment		-.136 (1.741)			2.593 (3.053)	
Livestock Treatment			-8.234** (3.218)			-3.664 (3.973)
Duration of treatment	Yes	Yes	Yes	Yes	Yes	Yes
Duration X treatment type	No	Yes	Yes	No	Yes	Yes
Covariates	Yes	Yes	Yes	Yes	Yes	Yes
Subdistrict FEs	Yes	Yes	Yes	Yes	Yes	Yes
# of Obs.	1,060	751	772	1,060	751	772
R-squared	0.478	0.558	0.473	0.500	0.512	0.524

Notes. Covariates are gender, age and education level of household heads, household size, number of shocks, income shares from different economic activities, land restriction dummy, and water constraint dummy. Diff RNM is the difference between Wave 4 and Wave 3 levels of RNM. *p-value<0.10, **p-value<0.05, ***p-value<0.01

Table 24. Long-term Analysis for Limiting portion size at meals (Number of days) using Entropy Balancing

	LPS	LPS	LPS	Diff LPS	Diff LPS	Diff LPS
Treatment	-1.266			-.246		
	(1.721)			(2.159)		
Agri. Treatment		-4.531			-4.532	
		(2.839)			(2.839)	
Livestock Treatment			-9.820***			-9.208**
			(3.229)			(3.756)
Duration of treatment	Yes	Yes	Yes	Yes	Yes	Yes
Duration X treatment type	No	Yes	Yes	No	Yes	Yes
Covariates	Yes	Yes	Yes	Yes	Yes	Yes
Subdistrict FEs	Yes	Yes	Yes	Yes	Yes	Yes
# of Obs.	1,060	751	772	1,060	751	772
R-squared	0.445	0.467	0.477	0.407	0.467	0.438

Notes. Covariates are gender, age and education level of household heads, household size, number of shocks, income shares from different economic activities, land restriction dummy, and water constraint dummy. Diff LPS is the difference between Wave 4 and Wave 3 levels of LPS.
*p-value<0.10, **p-value<0.05, ***p-value<0.01

8. Conclusions

In this paper, we first review recent research on covariate balancing to guide researchers who face difficulties in conducting rigorous impact evaluations in HECS with non-randomized treatment assignments and significant covariate imbalances between treatment and control groups due to targeting. Then, as a result, we propose and test methods for overcoming the challenges of causal identification in conflict or humanitarian contexts.

We then apply these methods, providing novel evidence on agricultural interventions' causal, long-term impact in a humanitarian crisis setting. Employing quasi-experimental methods with the proper weighting algorithms, we demonstrate that there are long-term impacts of humanitarian interventions provided under the so-called SEEDS program in Syria. We show that both agricultural and livestock interventions impacted food consumption scores, household dietary diversity and reduced coping strategies for the treated households. The panel data analysis reveals that the agricultural interventions' initial impacts were quite significant and did not keep increasing in the subsequent periods. Still, the difference in outcomes between treated and control group families was sustained at the same levels three years after the end of the humanitarian intervention.

Moreover, we find that agricultural interventions have a stronger long-term impact on the welfare of households by reducing their vulnerabilities to food insecurity. We also show that providing vegetable kits in Syria decreased the probability of child labour and child marriage in treated households three years after the programme ended. It also reduced the use of other harmful coping strategies, such as selling productive assets to access food. In other words, our findings show that short-term humanitarian assistance can have long-term positive development impacts.

Furthermore, by testing the long-term effects of livestock and agricultural interventions separately, we provided evidence that agricultural or livestock interventions have different long-term impacts, implying that combined interventions may have more significant and longer-term impacts on households.

Our study also reveals novel findings from post-war Syria, where no information exists on the impact of agricultural interventions on household resilience and welfare. We broaden the range of countries examined in the literature on humanitarian settings by using unique panel data with four waves of both treated and untreated households, providing rigorous evidence where it did not exist before.

Overall, our findings suggest that employing advanced econometric techniques can reveal how humanitarian interventions create high short-term impacts, as expected. However, their long-term impact should not be underestimated in planning and implementation. We find that the humanitarian intervention reduced households' vulnerability to shocks in the future. These changes were strong despite worsening weather, macroeconomic and health conditions in Syria during the same period. We also provided evidence that agricultural and livestock interventions have different effects on households` coping strategies in case of food shortage. As we have seen in our case, small farmers diversify their economic activity between crop farming and herding. Thus, humanitarian interventions in conflict settings can increase the welfare of households when combinations of input provisions are provided to households instead of focusing only on one aspect of the economic activity they are involved in.

References

Adelaja, A., & George, J. (2019). Effects of conflict on agriculture: Evidence from the Boko Haram insurgency. *World Development*, 117, 184-195.

Alinovi, L., Mane, E., & Romano, D. (2008). Towards the measurement of household resilience to food insecurity: Applying a model to Palestinian household data. In FAO (Ed.), Deriving food security information from national household budget surveys. Experiences, achievement, challenges (pp. 137–152). Rome: FAO. Retrieved from http://www.fao.org/3/i0430e/i0430e00.htm [Google Scholar]

Angrist, J. D., and J.-S. Pischke. (2009). Mostly Harmless Econometrics: An Empiricist's Companion. Princeton, NJ: *Princeton University Press.*

Arias, M. A., Ibáñez, A. M., & Zambrano, A. (2019). Agricultural production amid conflict: Separating the effects of conflict into shocks and uncertainty. *World Development*, 119, 165-184.

Baliki, G., Schreinemachers, P., Brück, T., & Uddin, N. M. (2022a). Impacts of a home garden intervention in Bangladesh after one, three and six years. *Agriculture and Food Security*, 11(1), 1-9.

Baliki, G., Kayaoglu., A., Daccache, M., Weiffen, D., & T. Brück (2022b). Novel Approaches for Impact Evaluations in Humanitarian Emergency and Conflict Settings. *ISDC Evidence Brief*, August 2022.

Bartels, S. A., Michael, S., Roupetz, S., Garbern, S., Kilzar, L., Bergquist, H., ... & Bunting, A. (2018). Making sense of child, early and forced marriage among Syrian refugee girls: a mixed methods study in Lebanon. *BMJ Global Health,* 3(1), e000509.

Bishaw, Z., Struik, P. C., & G Van Gastel, A. J. (2015). Wheat and barley seed system in Syria: How diverse are wheat and barley varieties and landraces from farmer's fields?. *International Journal of Plant Production*, *9*(1), 117-150.

Blundell, R., & M. C. Dias. (2009). Alternative approaches to evaluation in empirical microeconomics. *Journal of Human Resources,* 44, 565–640.

Brück, T., d'Errico, M., & Pietrelli, R. (2019). The effects of violent conflict on household resilience and food security: Evidence from the 2014 Gaza conflict. *World Development*, 119, 203-223.

Churchill, S. A., Smyth, R., & Trinh, T. A. (2022). The intergenerational impacts of war: Bombings and child labour in Vietnam. *The Journal of Development Studies*, 1-17.

Darmon, N.. & Drewnowski, A. (2015). Contribution of food prices and diet cost to socioeconomic disparities in diet quality and health: a systematic review and analysis. *Nutrition Reviews*, 73(10), 643-660.

de Brauw, A., Moursi, M., and Munhaua, A.B. (2019). Vitamin A intakes remain higher among intervention participants 3 years after a biofortification intervention in Mozambique. *The British Journal of Nutrition*, 122(10), 1175-1181.

Depenbusch, L., Schreinemachers, P., Brown, S., & Roothaert, R. (2022). Impact and distributional effects of a home garden and nutrition intervention in Cambodia. *Food Security,* 14(4), 865-881. *doi:10.1007/s12571-021-01235-y.*

Development Initiatives (2020). 2020 Global Nutrition Report: Action on equity to end malnutrition. Bristol, UK; *Development initiatives.*

Dizon, F., & Herforth, A. (2018). The cost of nutritious food in South Asia. *Policy Research Working Paper - World Bank*, 8557, 24.

Doocy, S., Tappis, H., Lyles, E., Witiw, J., & Aken, V. (2017). Emergency food assistance in northern Syria: an evaluation of transfer programs in Idleb Governorate. *Food and Nutrition Bulletin*, *38*(2), 240-259.

ECSWA (2016). Syria at war: Five years on. https://archive.unescwa.org/sites/www.unescwa.org/files/publications/files/syria-war-five-years-english_1.pdf

El-Shater, T., Yigezu, Y.A., Mugera, A., Piggin, C., Haddad, A., Khalil, Y., Loss, S. & Aw-Hassan, A., (2016). Does zero tillage improve the livelihoods of smallholder cropping farmers? *Journal of Agricultural Economics*, *67*(1), 154-172.

Food and Agriculture Organization of the United Nations, (2016). Improving Food Systems for Nutrition and Health. *Food and Agricultural Organization of the United Nations*, Rome.

Gaworek-Michalczenia, M. F., Sallu, S. M., Di Gregorio, M., Doggart, N., & Mbogo, J. (2022). Evaluating the impact of adaptation interventions on vulnerability and livelihood resilience. *Climate and Development*, 14(10), 867-883. *doi:10.1080/17565529.2021.2018987.*

George, J., Adelaja, A., & Awokuse, T. O. (2021). The agricultural impacts of armed conflicts: the case of Fulani militia. *European Review of Agricultural Economics*, 48(3), 538-572.

Ghosh, S., Pellett, P. L., Aw-Hassan, A., Mouneime, Y., Smriga, M., & Scrimshaw, N. S. (2008). Impact of lysine-fortified wheat flour on morbidity and immunologic variables among members of rural families in northwest Syria. *Food and Nutrition Bulletin*, *29*(3), 163-171.

Haddad, L., Hawkes, C., Webb, P., Thomas, S., Beddington, J., Waage, J., & Flynn, D. (2016). A new global research agenda for food. *Nature*, 540(7631), 30–32.

Hainmueller, J. (2012). Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies. *Political Analysis* 20 (1): 25–46.

Heckman, J. J., H. Ichimura, and P. E. Todd. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. Review of Economic Studies 64: 605–654.

Heckman, J. J., H. Ichimura, & P. E. Todd. (1998). Matching as an econometric evaluation estimator. Review of Economic Studies 65: 261–294.

Imbens, G. W., & Rubin, D. B. (2015). Causal inference in statistics, social, and biomedical sciences. Cambridge University Press.

Jansen, L., Outwater, A. H., Lowery Wilson, M., Iseselo, M. K., & Bärnighausen, T. (2022). A controlled pilot intervention on community violence prevention, financial and social capital generation in Dar Es Salaam, Tanzania. *BMC Public Health,* 22(1), 335.

Justino, P. (2012). War and poverty. In M. Garfinkel & S. Skaperdas (Eds.), Handbook of the economics of peace and conflict. Oxford University Press.

Kahn-Lang, A., & Lang, K. (2020). The promise and pitfalls of differences-indifferences: Reflections on 16 and pregnant and other applications. *Journal of Business & Economic Statistics*, 38(3), 613–620.

King, G., & Nielsen, R. (2019). Why propensity scores should not be used for matching. *Political Analysis*, *27*(4), 435-454.

Martin-Shields, C. P., & Stojetz, W. (2019). Food security and conflict: Empirical challenges and future opportunities for research and policy making on food security and conflict. *World Development*, 119, 150-164.

Mercier, M., Ngenzebuke, R. L., & Verwimp, P. (2020). Violence exposure and poverty: Evidence from the Burundi civil war. *Journal of Comparative Economics*, 48(4), 822-840.

Moore, N, Lane, C, Storhaug, I, Franich, A, Rolker, H, Furgeson, J, Sparling, T & Snilstveit, B. (2021). The effects of food systems interventions on food security and nutrition outcomes in low- and middle-income countries, 3ie *Evidence Gap Map Report 16.* New Delhi: International Initiative for Impact Evaluation (3ie). Available at: https://doi.org/10.23846/EGM016

Naudé, W., E. Amorós & T. Brück (2023). State-Based Armed Conflict and Entrepreneurship: Empirical Evidence. Edward Elgar Handbook of Research on Entrepreneurship and Conflict, *forthcoming.*

Ntakyo, P. R., & Van Den Berg, M. (2022). The unintended side-effects of a major development strategy: commercialization of smallholder production and women empowerment in Uganda. *The Journal of Development Studies*, 58(8), 1605-1626. *doi:10.1080/00220388.2022.2032671*.

OECD (2022). *States of Fragility 2022*, OECD Publishing, Paris, https://doi.org/10.1787/c7fedf5e-en.

Puri, J., Aladysheva, A., Iversen, V., Ghorpade, Y. & Brück, T. (2017). Can Rigorous Impact Evaluations Improve Humanitarian Assistance? Journal of Development Effectiveness, 9(4), 519-42.

Salazar, L., Palacios, A. C., Selvaraj, M., & Montenegro, F. (2021). Using satellite images to measure crop productivity: Long-term impact assessment of a randomized technology adoption program in the Dominican Republic (No. IDB-WP-01234). IDB Working Paper Series. *doi:10.22004/ag.econ.315408?utm_source=ideas.*

SCPR (2016). Forced Dispersion A Demographic Report on Human Status in Syria. Syrian Centre for Policy Research. https://www.scpr-syria.org/forced-dispersion-syrian-human-status-the-demographic-report-2016/

Shemyakina, O. (2022). War, conflict, and food insecurity. *Annual Review of Resource Economics*, *14*, 313-332.

Springmann, M., Clark, M., Mason-D'Croz, D., Wiebe, K., Bodirsky, B. L., Lassaletta, L., ... & Willett, W. (2018). Options for keeping the food system within environmental limits. *Nature*, *562*(7728), 519-525.

Swinburn, B., Dominich, C. H., Vandevjivere, S. (2014). Benchmarking food environments: experts' assessments of policy gaps and priorities for the New Zealand government. Auckland: University of Auckland.

Tranchant, J.-P., Justino, P., & Mueller, C. (2014). Political violence, drought and child malnutrition: Empirical evidence from Andhra Pradesh, India. *HiCN Working Paper 17*3.

Tranchant, J. P., Gelli, A., & Masset, E. (2021). Impact of conflict-related violence and presence of armed groups on food security: Evidence from longitudinal analysis in Mali. *Intl Food Policy Res Inst,* 2095.

UNEP (2019). Collaborative Framework for Food Systems Transformation. A multistakeholder pathway for a sustainable food system

UNICEF (2019). Children, food and nutrition. The state of the world children 2019. https://www.unicef.org/media/63016/file/SOWC-2019.pdf

UNOCHA (2022). 2022 Humanitarian Needs Overview: Syrian Arab Republic (February 2022). https://www.humanitarianresponse.info/en/operations/whole-of-syria/document/2022-humanitarian-needs-overview-syrian-arab-republic

UNSCN (2019). Food environments: Where people meet the food system

Weiffen, D., Baliki, G., & Brück, T. (2022). Violent conflict moderates food security impacts of agricultural asset transfers in Syria: A heterogeneity analysis using machine learning. *Households in Conflict Network, 381.*

Wendle, J. (2016). Syria's climate refugees. *Scientific American, 314(3), 50-55.*

Willett, W., Rockström, J., Loken, B., Springmann, M., Lang, T., Vermeulen, S., & Murray, C. J. L. (2019). Food in the Anthropocene: the EAT-Lancet Commission on healthy diets from sustainable food systems. Lancet (London, England). https://doi.org/10.1016/S0140-6736(18)31788-4

World Bank (2016). World development report 2016: Digital dividends. Washington, D.C.: World Bank Group.

WFP, F. (2022). Hunger Hotspots. FAO-WFP early warnings on acute food insecurity: October 2022 to January 2023 Outlook.

Xu, Yiqing and Yang, Eddie, Hierarchically Regularized Entropy Balancing (March 19, 2021). Available at SSRN: https://ssrn.com/abstract=3807620 or http://dx.doi.org/10.2139/ssrn.3807620