



HAL
open science

Changing Dietary Habits Early in Life: A Field Experiment with Low-Income Families

Michèle Belot, Noémi Berlin, Jonathan James, Valeria Skafida

► **To cite this version:**

Michèle Belot, Noémi Berlin, Jonathan James, Valeria Skafida. Changing Dietary Habits Early in Life: A Field Experiment with Low-Income Families. *Journal of Political Economy Microeconomics*, 2024, 2 (4), pp.000-000. 10.1086/732126 . hal-04647948

HAL Id: hal-04647948

<https://hal.science/hal-04647948v1>

Submitted on 30 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Changing Dietary Habits Early in Life: A Field Experiment with Low-Income Families

Michèle Belot¹, Noémi Berlin², Jonathan James³, and Valerie Skafida⁴

¹Cornell University

²CNRS, EconomiX, University of Paris Nanterre

³University of Bath

⁴University of Edinburgh

Abstract

We evaluate the malleability of dietary habits early on in life with two interventions targeting low-income families with very young children. The interventions were conducted over twelve weeks and evaluated with a field experiment involving 285 families. Families in the treatment groups were asked to either prepare healthy meals with ingredients delivered to their door, or to avoid snacking and eat at regular times. We find that both interventions decreased children's Body Mass Index (BMI) relative to the control group. The effect persisted for three years for the first intervention, but faded away for the second.

JEL Classification: I12, I14, I18

Keywords: Dietary Habits, Early Life, Healthy Eating, Field Experiments, Obesity

Acknowledgments: This project was funded by the European Union's Seventh Framework Programme, under grant agreement no. 607310. The experiment is pre-registered under trial no. AEARCTR-0003281. We thank the editor and two anonymous reviewers for their valuable comments. We also thank local authorities, early life centres, and schools in Edinburgh and Colchester; Ilenia Paciarotti, Hazel Blakemore and Gaja Lescinar for their help with the design and implementation of the interventions, Philipp Kircher, Andrew Kelloe, Lise van Wijngaarden, Kritsina Enger, Hasna Fadhilah, Dragos Gruia, Yvonne Laird, Agnès Charpin, Chiara Serra, Zoey Verdun and Miguel Alquezar for their help with assistance during the study, ISER and the EssexLab for allowing us to use their facilities, and nutritionists from Alison Clark Health and Nutrition. We thank Gareth Leng and John Menzies for their help with the blood samples collection. We thank Joe Doyle, Gareth Leng, John Menzies and Philipp Kircher for very useful comments. We also thank the participants of the seminars at Aberdeen, Amsterdam, ALISS, Bath, Bolzano, Copenhagen, Essex, GATE (Lyon), Glasgow, Gothenburg, INRAE, LIRAES, Lucerne, Mainz, Nottingham, Paris 5, Paris School of Economics, Stellenbosch, Stirling, UCL, Utrecht, Hamburg; and participants to the ESRC workshop in Health Economics 2018, Hamburg workshop in Risky Behaviors 2019, ESA 2016, ESPE 2018, EALE 2018, Royal Economic Association 2019, Australian Health and Human Capital Seminar Series 2021. We thank Christine Schmidt for proofreading the paper. An earlier version of the paper was circulated with the title "The Formation and Malleability of Dietary Habits: A Field Experiment with low-income Families". The paper was edited by Anna Dreber.

1 Introduction

Poor diet is a major issue in most developed and developing countries. While there are many policies targeting diets such as taxes and information campaigns¹ and, more recently, a series of interventions based on insights from behavioural economics, the evidence on medium or long-term effects of interventions remains limited. The most promising interventions with a potential for long-term effects appear to be those targeting children (Belot and James, 2022), in line with evidence on the effectiveness of early-life interventions in other domains of life (Baulos and Heckman, 2022). In the context of diet specifically, the nutrition literature has documented the persistence of diet from childhood to later stages in life (Birch, 1999), suggesting that habits form early. Another reason for early intervention relates to the challenges of reversing obesity later on in life. Recent research in biology has shown that the number of adipocytes (“fat cells”) - a strong predictor of obesity in adulthood - is set during childhood and early adolescence (Spalding et al., 2008). For all these reasons, intervening in childhood is likely to be more effective.

In this study, we propose and evaluate two comprehensive interventions, each lasting for three months, aimed at changing eating habits. The interventions target low-income families with young children (below the age of six) and involve the entire family. The first intervention (*Meal treatment*) consists of providing families with ingredients and recipes, directly at home, to prepare five meals per week over a period of twelve weeks. This intervention aims at relaxing constraints families may face in accessing healthy foods and thereby echoes the suggestion by Ludwig et al. (2011) to study the problem of “food deserts”.² The second intervention (*Snack treatment*) targets snacking and erratic eating habits. This treatment is inspired by evidence that suggests that a rising proportion of calories in children’s diets is coming from snack foods, which are in turn associated with higher sugar and fat consumption (St-Onge et al., 2003; Bellisle, 2014). Additionally, studies have shown meal irregularity to be linked to poor dietary outcomes (Laska et al., 2015; Leech et al., 2015). Again for twelve weeks, families are instructed to eat three meals a day, at regular times, avoiding eating

¹Capacci et al. (2012) and Frisvold and Lensing (2020) propose recent reviews of relevant studies.

²Ludwig et al. (2011) propose the recruitment of low-income families to be randomly assigned to receive free weekly delivery of fresh fruits and vegetables to their homes, in order to investigate the role of food deserts on obesity.

(snacking) between meals.

We evaluate these interventions with a randomized controlled trial (RCT) on a sample of 285 low-income families located in two different locations in the United Kingdom: Colchester (England) and Edinburgh (Scotland). We focus on low-income families because of the well-documented evidence of a strong socio-economic gradient in chronic diseases and obesity (Dalstra et al., 2005). We use different measures to evaluate the impact of the interventions: some objective, such as body mass index (BMI), others based on incentivized choices, and finally, methods used in nutrition research based on in-depth surveys of participants. We collect information before, immediately after, and once a year for three years after the interventions.

We find evidence of significant sustained changes in children's BMI (z-score) following the *Meal* treatment. Children in both treatment groups appear to have moved down in the distribution of BMI. These effects are large (between 0.2 and 0.34 standard deviations) and are sustained in the longer run for the *Meal* treatment, while they attenuate over time for the *Snack* treatment and are no longer precisely estimated after the initial post period but remain negative. Parallel to these changes in BMI, we find evidence of significant changes in diet (measured with different methods). The changes are more sustained in the *Meal* treatment, in line with the results on BMI. In particular, we find consistent evidence of a reduction in the intake of added sugar in treated children, this is more pronounced in the *Meal* treatment than in the *Snack* treatment. Using an incentivized measure of dietary choices, we find that the parents of children in the *Meal* treatment purchase less sugary foods post-intervention.

Notably, we find no evidence that the interventions altered preferences in favor of healthier foods among children. Measures of food preferences indicate that children in the treated groups seem to like sweets more after the intervention, which goes against predictions of habit formation models (Becker and Murphy, 1988). However we do find that the interventions affected the purchasing behavior of parents, who were less likely to purchase unhealthy items and foods high in sugar in particular. Thus, the most plausible mechanism for the effects we observe is that parents reduced access to unhealthy foods, in particular sweets, and that this effectively limited weight gain in the short and medium run. We also find that the *Meal*

protocol was easier to follow than the *Snack* protocol, which could have also contributed to why we observe long-term effects for the *Meal* treatment and not for the *Snack* treatment.

The study focuses primarily on children. However, parents were also treated, so we collected information on them as well. We find little evidence of changes in dietary habits among adults. Neither subjective nor objective measures show evidence of improvement in the healthiness of their diet or their preferences for healthier foods, in the short or longer run. These findings reinforce the conclusion in the previous paragraph, which is that parents changed what they purchased for their children but not for themselves. Overall, our results confirm that interventions later on in life are much more challenging.

Our study makes several contributions to the existing literature (discussed in Section 2). First, the study focuses on families with very young children (aged between 2 and 6 at the beginning of the study). The interventions also target the home setting rather than schools, and the parents are treated along with their children. To our knowledge, this is the first experimental evaluation of interventions targeting children and their parents at home.³ Second, we evaluate effects on the whole diet rather than on isolated single purchase/eating decisions. Most studies rely on partial measures of dietary choices, such as the consumption of specific items. Downs and Loewenstein (2011) identify this as a key shortcoming of existing studies, writing that “the true success of such measures will remain unclear until researchers are able to measure an individual’s total food intake – not only calories at a single meal or in a single episode of snacking.” Third, we evaluate effects on a much longer horizon than other studies. We follow the families for *three* consecutive years. Most studies are typically short-run or, if they include a follow-up at a later point in time, the horizon considered is usually shorter than a year. Fourth, the scope of the interventions tested is unprecedented. Particularly for the *Meal* treatment, we would argue that it is among the most invasive intervention designs to be implemented with families over a temporary period.

The remainder of the paper is organized as follows: Section 2 reviews the recent relevant literature. Section 3 presents the experimental design, Section 4 describes the main outcome measures and Section 5 presents the analysis and results. In Section 6, we discuss possible

³Charness et al. (2023) find in a school-based experiment that involving the parents is key to the long-term success of an intervention aiming at encouraging healthy choices at school. The interventions tested are one-shot, they show that using a grading system for informing children about the healthiness of foods is effective in encouraging healthy choices, with effects carrying on weeks after the information was given.

mechanisms behind the effects we find. In Section 7, we discuss the robustness of the results, and in Section 8 the external validity. Section 9 concludes with a discussion of our findings.

2 Related Literature

The paper fits within the recent literature testing and evaluating interventions aimed at fostering healthier habits. Charness and Gneezy (2009) were the first to show that an intervention of a temporary nature (presumably long enough to form new habits) could lead to a sustained change in a lifestyle behavior - in their case, exercise. In their study, students were incentivized to exercise for one month. They found that these students were then more likely to exercise, even when the incentives were removed.

The idea of using temporary interventions to encourage healthier habits has also found encouraging support in the domain of nutrition. Several studies (Just and Price, 2013; List and Samek, 2015; Belot et al., 2016; Loewenstein et al., 2016) have shown promising evidence that children’s dietary choices can be influenced over sustained periods. However, the evidence relates to isolated single eating decisions, and the evidence on long-run effects is very limited. Also, these studies do not evaluate the short- or long-run effects on the overall diet or other relevant measures such as weight or body mass index (BMI).

There are very few studies focusing on very young children (below the age of 6). A recent study by Griffith et al. (2018) studies the effects of a fruit and vegetables voucher scheme targeting low-income families with children below the age of 3 in the UK. They find a significant impact on purchases of fruit and vegetables for eligible families.⁴ Daniels et al. (2012) conduct an RCT to evaluate an educational intervention targeting parents of infants and aiming to encourage healthy feeding practices. They find that the intervention had an impact on the BMI of the treated children measured 6 months after the intervention. Neither study investigates effects in the longer-term though.

More broadly and not targeting children specifically, Cawley et al. (2016) and List et al. (2022) conduct field experiments in collaboration with a supermarket and test different types of interventions targeting the prices of nutritious and less nutritious foods (subsidy,

⁴They use as a comparison group families with a pregnant mother or with children slightly older than 3.

taxes, information) to increase the consumption of the former. These experiments show that these interventions affect purchasing behavior (low-income families purchased more of both nutritious and less-nutritious food under the subsidy framing in Cawley et al. (2016) and that incentives can lead to sustained changes in the purchase of fruit and vegetables even when the incentives are removed, suggesting habit formation has taken place List and Samek (2015). However, overall spending in the store involved in the experiment was low, suggesting that most other food purchases were taking place elsewhere. Brownback et al. (2019) also target grocery shoppers (with a low-income) and examine the effects of combining food subsidies with other behavioral interventions (giving shoppers greater agency in the choice of subsidies and introducing a waiting period before the shopping trip). They find that these behavioral interventions substantially increase the effectiveness of food subsidies. While providing encouraging evidence, these studies focus on purchasing behavior rather than consumption, they only have a partial picture of what people buy or eat, and they do not evaluate effects in the longer-term either.

Our paper also contributes to a recent literature on the effects of the quality of food environments one is exposed to. Allcott et al. (2019) use a structural model to simulate the effects of a hypothetical experiment where low-income households would have access to the same products and prices as high-income households. They find that this would reduce nutritional inequality by only approximately 10 percent. This is because the demand for healthy foods is lower among low-income households. Here we provide experimental evidence where families are actively encouraged to try different foods, and the goal is therefore to evaluate whether it is possible to affect the demand.

Our study also relates to recent work in Economics on the formation of preferences, and the role of parents in the formation of fundamental economic preferences such as risk and trust attitudes (Zumbuehl et al., 2021) and in early childhood development (Baulos and Heckman, 2022). The general finding is that early interventions in a range of domains are more effective than later interventions, and that parental involvement tends to be associated with the development of more favorable preferences, that is, traits that are correlated with better life outcomes. Our study findings are in line with these general lessons.

Finally, our study also relates to the recent literature on paternalism. Ambuehl et al.

(2021) study what motivates choice architects when designing a choice environment.⁵ To this date, there is very little work on understanding parental attitudes towards paternalistic interventions. There is evidence that people are willing to restrict other people’s choices for their own good (Krawczyk and Wozny, 2017) and may be more patient for others than for themselves and act as “projective paternalists” who seek to align others’ choices with their own aspirations (Shapiro, 2010). Our study shows that parents are willing to impose certain choices on their kids without necessarily adopting these choices themselves.

3 Experimental Design

The experiment was reviewed and approved by the Ethics Committee at the University of Edinburgh and was pre-registered on the AEA RCT Registry.⁶ Full details of the experimental design and deviations from the pre-analysis plan are provided in the online Appendix Section A.3. All the relevant documentation, protocols and survey instruments are accessible in the following [Dropbox folder](#).

3.1 Selection of Experimental Sample

We are interested in evaluating two interventions targeting young children (eating solid foods) from low-income families in the UK. As mentioned earlier, the evidence suggests a negative correlation between socioeconomic status and the risk of childhood obesity. Following the suggestion in List (2020), we describe the process by which we selected our sample and present summary statistics on key variables comparing our target population (low-income families in the UK) with our experimental sample.

First, we selected two areas in the UK to conduct the study: Edinburgh (Scotland) and Colchester (England). These two locations were chosen because of practical reasons (we had access to University facilities where families could come - we had access to a computer lab

⁵We thank an anonymous referee for suggesting this connection.

⁶Details of the registered trials can be found on AEA RCT Registry: Belot, Michèle et al. 2018. “The formation and malleability of dietary habits: A Field Experiment with low-income families.” AEA RCT Registry. October 04. <https://doi.org/10.1257/rct.3281-1.0>. Pre-registration was not common practice at the start of the data collection, so the experiment was registered after the start of the experiment but before the end of data collection. The trial was however described in detail in the ethical approval application at the University of Edinburgh, which was approved before the start of the experiment. The main deviation from the original plan is in the fractions allocated to each treatment, which we discuss below.

to conduct surveys, facilities where children and adults could be measured and, in the case of Edinburgh, facilities where blood samples could be collected.)

Second, families were required to satisfy the following criteria to participate in the study: have a child aged between 2 and 6 years old at the start of the study, own a fridge and a hob (cooktop), live in Edinburgh or Colchester, have a household income below £26,426 for Scotland, and £26,600 for England.⁷ Families with severe chronic health issues were excluded. These criteria are likely to matter for the generalizability of our findings. Not all low-income families in the UK have access to a fridge and a hob. Recent research revealed that 1 in 7 low-income households (with household incomes equal to or below £35,000) live without at least one essential household item (washing machine, fridge, cooker, or freezer).⁸

Various recruiting methods were used, such as adverts in public spaces, community centers, nurseries, buses, and shopping malls; letters sent to school principals, advertisements on radio, and individual letters targeting low-income areas. Samples of our recruitment materials (leaflet and poster) are shown in Appendix Figure A.1a and Figure A.1b. Families needed to sign up for the study by registering on a dedicated website or calling a dedicated phone number.

For ethical reasons, participants needed to provide informed consent before being included in the study. To mitigate potential self-selection concerns, we were careful in not disclosing our research question and providing only general information about the study - such as the study being related to health and lifestyle choices, and the study duration of three years. Also, participants received monetary compensation commensurate to the time required to spend on visiting the facilities (between £350 and £400 for six visits of around two hours), which should mitigate potential self selection effects. This said, we can of course not claim that our sample of families is not self-selected in characteristics that could matter for the effects of the treatments, such as their motivation and openness to change.

Our baseline sample consists of 285 families (91 families in Edinburgh and 194 in Colchester)⁹. Appendix Table A.7 compares our experimental sample with the targeted population

⁷These thresholds correspond to the median individual income at the time. Since a large fraction of our families are single-headed households, we used a measure based on individual income.

⁸See report by the National UK Charity Turn2us "Living Without - The Scale and Impact of Appliance Poverty" available on their website <https://www.turn2us.org.uk>.

⁹Originally 377 families registered for the study but 285 came to the initial sessions. 11 pregnant women

on several key observable characteristics. The table shows that our experimental samples are on average slightly better off than the targeted population. As mentioned earlier, we excluded families who do not own a cooker or a fridge. The statistics we have for Edinburgh and Colchester in the representative surveys do not exclude such families, so this could explain the differences. And, of course, this also means that our findings will not extrapolate to populations who may not have access to basic cooking equipment.

3.2 Experimental Interventions

We implemented two protocols, both for 12 weeks. These protocols were implemented in a natural manner: families were asked to follow these protocols as part of their normal life. While these protocols aimed at introducing significant changes in dietary patterns, they allowed for some flexibility on the side of participants. As recent work by Beshears et al. (2021) shows, interventions that impose rigid constraints may in fact be counterproductive.

One of the key parameters here is the length of intervention. We settled on 12 weeks. This choice was motivated by earlier studies evaluating interventions encouraging children to eat healthier foods (see Belot and James (2022) for a review of relevant studies). The interventions tested usually last between two and four weeks. Because the long term effects of these interventions have been mixed, we extended the length of the intervention to the maximum feasible. The constraints were related to holiday periods in the UK, which would have perhaps reduced compliance.

In Section A.3.4 of the Appendix we provide details explaining our design choices.

Treatment 1 - “Meal Treatment”

The first treatment, referred to as “Meal Treatment”, consists of providing (free of charge) ingredients and recipe booklets every week, for 12 weeks, directly at participants’ homes for five main meals for the *whole* family.¹⁰ The main goal here is to get families to try new

were excluded from the analysis of adults. There were 6 pairs of twins at initial session so a maximum of 291 children were included in the analysis. It is often the case that for some measurements in children, we have a few observations missing. Note that we collected information only from one child in the family, the youngest child between the age of 2 and 6. When the youngest children were twins, both children were included in the study.

¹⁰Families could select between a regular or vegetarian food basket. For a list of recipes by week see Appendix Tables A.3 and A.4.

recipes and be exposed to new foods for a period of 12 weeks.

The protocol has been designed to ensure convenience and limit non-financial costs that could be important obstacles in adopting a healthy diet: Families do not have to plan for these meals, i.e. they do not have to search for suitable recipes, organize the shopping, etc. The food is delivered at home, and families receive a weekly booklet of recipes for the ingredients delivered. The recipes have been chosen by a nutritionist for their low cost and simplicity of execution, which ensured that the food families were exposed to would be part of the usual British cuisine.¹¹ Most of them were borrowed from a government website associated with the largest UK public health campaign around healthy eating at the time (“Change4Life”). This protocol has also been deliberately chosen over protocols that would impose more constraints on families on all meals and food consumed, to ensure its feasibility.

Also, by providing the food for free, we aimed to maximize compliance during the 12 weeks of treatment. The costs of the meals have, however, been calibrated to the average weekly budget of low-SES families in the UK so families can afford to continue buying the ingredients and recipes once the interventions are over.¹²

Of course, the subsidizing of food may have spillover effects on other expenditures through income and substitution effects. This is a key reason why collecting measures of the whole diet and of BMI is important.

Treatment 2 - “Snack Treatment”

The second treatment, referred to as “*Snack Treatment*”, consists of regulating eating patterns and avoiding unhealthy snacking, again for 12 consecutive weeks. Adults in the family were asked to eat three meals per day at regular times (selected by participants) and to

¹¹An isocaloric comparison (fixed at 365 calories, the average calories of the meals) between the recommended nutritional guidelines and our recipes shows that our recipes are overall consistent with the recommendations, and are lower than the maximum thresholds on sugar and fat (and saturated fat), compensating for these calories via higher carbohydrate and protein contents. This can be seen in Appendix Table A.10, by comparing the second and the third columns.

¹²Based on ONS household expenditure data for 2015 (see <https://tinyurl.com/Familyfooddatasets> and <https://tinyurl.com/householdexpend>), average food expenditure for the first five income deciles was £28.81 per person per week (average taking into account both expenditure for adults and children). Our sample has an average household composition of 1.65 adults and 1.9 children per household. Based on this we calculate that our families would spend on average £102.28 per week on all in-house food and non-alcoholic drink expenditure. No data exists to capture the amount of this spent on the evening meal, though we hypothesize that this will be circa one-third of this budget. Based on this assumption, we calculate that £34.09 per week is spent on seven evening meals, and thus £24.35 is spent on five evening meals per week on average. This final figure was our benchmark for calibrating the cost of the meals.

avoid consuming any food or calorific drinks between meals. For the children, the protocol included an additional morning and afternoon snack (to be consumed at regular times). The snacks were provided by us and determined by a nutritionist. The list of snacks can be found in Appendix Table A.5.

This intervention draws on a body of evidence in economics and nutrition suggesting a link between snacking and obesity. A review of changes in childhood food consumption patterns by St-Onge et al. (2003) suggests that the rising proportion of calories coming from snack foods, which are in turn associated with higher sugar and fat consumption, may be a contributor to rising overweight and obesity in children. A review paper by Bellisle (2014) suggests that snacking often contributes to the calorie count but little nutrition, especially among obese children and adults. Although snacking is often blamed for rising obesity rates (Cutler et al., 2003), research on the effects of snacking on BMI is not conclusive (Field et al., 2004; Larson and Story, 2013).

A number of studies have also shown an association between meal irregularity and poor dietary outcomes more generally (Laska et al., 2015; Leech et al., 2015), and there may be metabolic advantages to eating at more regular and structured intervals (Alhussain et al., 2016). In the case of adults, people appear more likely to choose healthier foods when they select them in advance than when they select them at the moment when they will be consumed (Read and Van Leeuwen, 1998; Sadoff et al., 2020). A related body of literature in biology hypothesizes that irregularity of food intake could have a significant impact on diet and total calories, although this hypothesis is not supported in non-animal experiments (Hume et al., 2016).

Aside from providing healthy snacks for children, families were not instructed to alter the content of other meals. The focus remained on limiting and regulating snacking for the family, to create a source of exogenous variation across groups, and study how that has an impact on diet.

We will discuss compliance with both protocols in Section 3.6. Of course, it is difficult to enforce full compliance and we will therefore only be able to estimate “intention-to-treat” estimates.

Control Group

Control group participants were instructed to carry on as usual with their daily routines. Note that we were very careful not to communicate this instruction as an endorsement of their current diet — these instructions were framed as protocol instructions and it was clear that these could not be based on any information on their current diet, since that information could not have been available to the person instructing them on the first visit.

All groups received one-on-one instructions on the protocols during their first visit to our facilities.

3.3 Randomization

When registering to take part, families were asked to indicate several dates when they would be available to come to our facilities for the first session of measurements. All initial sessions' dates had been randomly pre-assigned to a specific treatment (control or one of the two treatments), and families were randomly assigned to one of their selected dates, without knowing these corresponded to the two different treatments and/or the control group.

We deliberately assigned a higher proportion of the families to the Control and *Meal* Treatment groups: 109 families were originally assigned to the *Snack* treatment, 141 to Control, and 146 to *Meal*. We did so after realizing the challenges of recruiting such a large sample. We decided we would drop one treatment arm if we did not succeed in recruiting at least 50 families for each treatment group. To prevent having to cancel the entire study if the sample was too small for two treatment arms, we decided to allocate proportionally more families to one of the treatments (*Meal* Treatment). The show-up rate at baseline was above 71% in the *Meal* group, it was a bit higher in the *Snack* group with a show-up rate of 72%, and it was 82% in the Control group.

Table 1 presents descriptive statistics of our sample for the different treatment groups. Overall, there are no statistically significant differences between the Control and the treatment groups at baseline. As defined by the recruitment criteria, the average age of the children is about 4 years old, and the average income is below the English and Scottish median income.¹³ Every household receives at least one type of welfare payment. Our sample contains more women than men; either because they were single mothers or the father was

¹³We did not enforce the income eligibility criterion as we could not ask for a reliable proof of income.

not available to attend the session.

3.4 Chronology of Data Collection

The experiment was conducted in two waves. The first wave was conducted at the University of Edinburgh and data collection started in February 2015. The second wave was conducted at the University of Essex (located in the city of Colchester) and started in August 2015. Participating families were expected to visit the University facilities six times over four years. The chronology of the data collection is as follows:

1. Session 1: Pre-intervention visit (“baseline”, 2 weeks prior to the first week of intervention) - February 2015 in Edinburgh, August 2015 in Colchester
2. 12-week protocol starting
3. Session 2: Intermediary session (mid-intervention period)
4. End of 12-week protocol
5. Session 3: Post-intervention visit (“After”, 2 weeks after the end of intervention) - June 2015 in Edinburgh, December 2015 in Colchester
6. Session 4: 1 year after Session 1 - February 2016 in Edinburgh, August 2016 in Colchester
7. Session 5: 2 years after Session 1 - February 2017 in Edinburgh, August 2017 in Colchester
8. Session 6: 3 years after Session 1 - February 2018 in Edinburgh, August 2018 in Colchester

3.5 Compensation

Families were compensated for each visit to our facilities. We estimate that the minimum time families would have to hold free to attend a session was around 3 hours (2 hours at

the facilities and 1-hour door-to-door transportation including walking from the parking or walking from the nearest bus station). They received £50 for each session (expected to last for two hours) except a for much shorter mid-intervention session (session 2), where they received £20. The amount for the initial session was increased to £100 to motivate families to attend the first session.¹⁴

3.6 Compliance

Compliance was assessed in various ways, such as asking families to provide pictures of the meals prepared and writing feedback information (see Appendix Section A.6).

To make sure that families fully understood what was expected from them, we met with each of them (treated and control) one-to-one at baseline and provided face-to-face instructions about the study and the protocols they should follow. Families in both treatment groups were told that we were interested in learning how easy the protocols were to follow and would value feedback on the difficulties they encountered. In Appendix A.6, we present details on the compliance measures. Overall our compliance measures show relatively high compliance with the protocols. The *Meal* Treatment appears to have been easier to follow compared to the *Snack* Treatment. Of course the estimates we will report are Intention-to-Treat estimates and the compliance tasks are part of the treatments we evaluate.

4 Main Outcome Measures

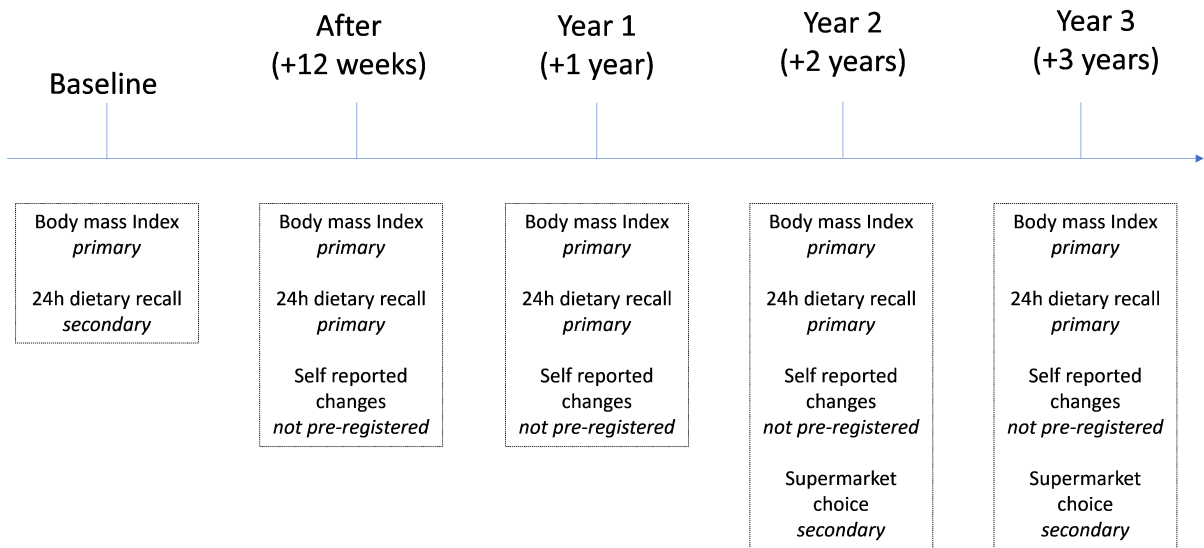
We evaluate the effects of the interventions on BMI and dietary habits. Note that these treatments were not designed as weight-loss programmes and did not impose any calorie restriction. Nevertheless, these protocols were expected to have longer-term effects on BMI through a change in dietary habits.

We collected measures at five different points in time: At baseline, after completion of

¹⁴The national living wage in the UK in 2015 was equal to £6.5 per hour. Since we invited both adults in the family (and at least one child), the equivalent pay for 2 adults spending three hours, with the compensation for a daily family bus ticket worth £8.5 would be £47.5, close to the compensation fee we offered. The compensation was higher in session 3 (£130) because participants had to bring back materials such as feedback leaflets and sim cards with pictures). The total compensation over three years added up to £350 in Edinburgh and £400 in Colchester for all visits, spread over the 6 visits.

the 12-week treatments, as well as one, two and three years after the interventions had taken place. Figure 1 shows the timeline of measurement of the different outcomes and indicate whether they were pre-registered as primary and secondary.¹⁵ All measures were collected on site at the Universities of Edinburgh and Essex.

Figure 1: Study timeline



4.1 Anthropometric Measurements and BMI (primary)

Participants were weighed and measured by a trained member of the team. Height and weight data were used to create age-adjusted BMI for children using BMI cut-offs (based on the percentiles) recommended by the Childhood Obesity Working Group of the International Obesity Taskforce (Vidmar et al. 2004).

4.2 Measures of Diet

We collected data on diet using three methods. For each method, we focus on variables that are most indicative of the healthiness or unhealthiness of a diet: total calorie intake, fruit and vegetable consumption, saturated fat, and added sugar (otherwise known as Non-Milk

¹⁵The baseline and post-experiment measurements were collected during two-week time windows before and after the 12-week treatment period.

Extrinsic Sugar (NMES) or free sugars).¹⁶ Importantly, these should not be considered as outcomes we wish to study independently. As we noted earlier, one major drawback of existing studies is that they usually focus on a single measure of diet, often based on an isolated food consumption or purchasing event. Our goal is to identify a consistent and robust pattern across these different measures and understand the mechanisms behind the changes observed.

4.2.1 24h dietary recall (primary)

The first measure of dietary intake relies on a method commonly used in nutrition research, called the “24-hour diet recall”. The method consists of asking participants to recall in detail what they have eaten in the last 24 hours. The data was recorded face-to-face with nutritionists at baseline and immediately after the intervention, and using a computer-based British software called Intake24 for the three subsequent years.

In our study, a parent (most often the mother) provided the information on behalf of the child and for themselves. The data was then entered into a nutritional analysis software (called WISP) which computes measures of dietary intake based on a large database of food items available in the UK. The software then provides direct information on calorie intake and macro-nutrient composition. More information on the method is provided in Appendix Section A.4.1.

One important challenge with our implementation of the 24-hour diet recall method is that participants could anticipate having to report their dietary intake and adjust their consumption the day before their visit. Of course, this challenge applies to all groups, and it is unclear whether the risk of bias is greater for treated or control groups. However, it is possible that those in the treated groups feel particularly inclined to under-report certain foods. The best way to address this concern is, in our view, to correlate the various measures we have and identify a pattern across the various measures we have.

¹⁶See a report from the National Diet and Nutrition survey (<https://tinyurl.com/NDNS20122014>) claiming that diets low in saturated fats and sugars and high in fruit and vegetables are typically recommended for health.

4.2.2 Self-Reported Changes (not pre-registered)

The second measure is based on a computer survey where parents were also asked to report changes in the consumption of specific food items (for their child and themselves). Specifically, participants were asked to report whether consumption increased, decreased, or did not change, relative to the previous visit for specific items. Specifically, for the first year follow-up we asked: “Could you indicate how your consumption has changed over the last 9 months for the following items?” For the second and third year follow up we asked to indicate the change over the last 12 months to reflect the timing of the visit.

We grouped these items into three categories: fruit and vegetables, foods high in fat, and foods high in added sugar. We calculate an index equal to the number of items for which an increase is reported, minus the number of items for which a decrease is reported. Therefore, these can be interpreted as net changes in the consumption of different food categories. So that the changes we observe are relative to their first visit (baseline), we sum the changes over the respective sessions. Therefore, this variable could reflect changes in just one item of the index, or many. For example, in year 1 someone may indicate that they have increased their consumption of vegetables but not fruit, so the variable would take a value of one. Then in year 2, they may indicate that over the last 12 months they have further increased their vegetable consumption, so this variable would now take a value of two. Similarly, in year 3 they have further increased their vegetable consumption but not fruit so that variable will now take a value of three. Alternatively, another participant could have a value of three in year 3 because they increased their fruit consumption and their vegetable consumption in year 1 and only their vegetable consumption in year 2.

This measure has been collected throughout the whole experiment i.e. after the treatments. Specifically, immediately after the end of the intervention, and then in the three subsequent years. Participants were also asked about other lifestyle changes (what they need to do to meet the food guidelines) and their perceived health. Note that this analysis was not initially pre-registered. We decided to include it as we believe it is helpful in the identification of a robust pattern across measures.

4.2.3 Supermarket Choice Tool (secondary)

The third measure is an incentivized measure based on a one-shot post-experimental choice of a supermarket basket worth 30 British Pounds (GBP), using the tool developed by Spiteri et al. (2019). Parents are asked to select food items among a choice set of 120 popular items from a supermarket, organized in categories that resemble typical aisles in a supermarket (fruit and vegetables, meat and fish, dairy, etc.). Screenshots of the tasks can be found in Appendix Figure A.5.

Participants are instructed to shop as they normally would. This means that the shopping basket is intended for the whole family rather than for themselves individually. The basket was delivered to one in every 10 families selected at random, two weeks after the intervention (to avoid that choices would depend on current food stocks). The tool allows us to extract information on calories and nutrients, such as sugar and saturated fat. This task was implemented in years 2 and 3 to avoid contamination with the interventions conducted in the treatment groups.

5 Analysis and Results

5.1 Baseline measurements

Table 2 shows the mean BMI (z-score), as well as the percentage of children underweight, overweight, or obese. The z-score is the number of standard deviations from the mean value. We see that the BMI z-score of the snack group is lower compared to the control and Meal groups but the differences between them are not statistically significant. The obesity rate of our sample is in line with national statistics (16.5% of the children and 33% of adults). We do not find significant differences in the distribution of weight categories between the three groups at baseline. The table also shows baseline measurements for different diet intakes (total calorie intake, added sugar, saturated fat, and fruit and vegetables). No significant differences are found for those measurements at baseline.

To get a sense of how the meals in the *Meal* treatment compare to the pre-intervention diet of participants, we compare the nutritional content holding calories constant (benchmark

of 365 calories). The average diet of participants contained twice the amount of fat (15g versus 8g) and twice the amount of sugars (20g versus 10g) found in the meals (see Appendix Table A.10). Participants' diets at baseline were lower in carbohydrates and protein relative to the *Meal* treatment.

5.2 Attendance and Attrition

Table A.7 provides descriptive statistics on attendance and attrition. The attrition rate after the 12-week intervention was very low (3.85%). The overall show-up rate remained high in the subsequent data collection years (87% in 2016, 84% in 2017, and 82% in 2018 relative to the first “baseline” session). However, the attrition rate differs slightly between treatments. There is a slightly higher rate of attrition in the *Snack* treatment at year 3 relative to baseline (20%) compared to both the *Meal* treatment (11%) and the Control group (16%). We discuss how attrition affects our results in Section 7. The main conclusion is that attrition is unlikely to drive the significant treatment effects we identify.

5.3 Intention-to-Treat (ITT) Estimates of the Impact on BMI and Diet

We now turn to the core results of the paper and present ITT estimates of the effects of the treatments. For measures with baseline information, the estimates are difference-in-differences estimates. For measures with only post-intervention information, the estimates relate to the difference between the treated and control groups. All estimates are from linear models, with standard errors clustered at the household level.

As mentioned earlier, we are looking for consistent evidence across measures. For this reason, we report all treatment effect estimates in one table for each treatment: Table 3 for the *Meal* Treatment and Table 4 for the *Snack* Treatment.

We report the treatment effects on BMI and measures of diet. For the measures of diet, we report treatment effects on measures derived from the 24h-dietary recall, measures derived from the survey (self-reported changes), and measures derived from the incentivized supermarket tool (shopping basket), the latter corresponding to parental choice.

Starting with the *Meal* treatment (Table 3), we find a precisely estimated and negative treatment effect on BMI, of the order of 0.20 standard deviations in the early years and equal to 0.34 in year 3. The size of the effects is initially very similar across both treatments. The effect is remarkably robust and persistent. Turning to the measures of diet, we present the individual coefficients as well as the tests for joint significance across measures (in a given year and across years).

We find evidence from all measures pointing at a reduction in sugar consumption in one and two years after the intervention. By year 2 we find that on average, relative to baseline, children have reduced their consumption of items high in sugar by around 3.7 items.¹⁷ In year 3, the effects across the board for sugar are less precisely estimated. We also have evidence (albeit weaker) of a reduction in fat consumption. While all except one of the point estimates is negative, only three of the 10 coefficients are precisely estimated. Jointly across all the coefficients on fat we do, however, find a significant effect. On the other hand, we do not find consistent evidence of a change in fruit and vegetable consumption. We find small but imprecise reductions in fruit and vegetable consumption in the 24 hour diet recall measures. In contrast, we find an increase in consumption when looking at the self-reported measures.

We also do not see a statistically significant drop in calories, but the reduction in added sugar consumption for the *Meal* group could be consistent with a drop in calories (corresponding to 75% of the point estimate change in calories at -109.6). After three years, we do not see a significant difference in added sugar intake, and the point estimate of the difference in calorie intake becomes positive (but remains insignificant). The self-reported measures are still significant, and the supermarket choice is also negative but is not precisely estimated.

For the *Snack* treatment (Table 4), we see less consistent effects. We find a negative treatment effect on BMI immediately after the intervention, and negative effects in each of the subsequent years albeit these are not precisely estimated ($p=0.093$ in year 1). There is also evidence of a drop in fat and sugar consumption in year 1 (only precisely estimated for sugar), but no consistent evidence emerges of significant changes in year 2 and beyond.

¹⁷As indicated in section 4.2.2 this could include different items or could be the same item over a number of years.

These results do fit well with the observed changes in BMI.¹⁸

Figure 2 presents the kernel density function for the distribution for the percentile ranks of BMI of children, at baseline, immediately after and in year 3, for the different control and treatment groups. The top left panel shows the density for the control group and the two treatments before the intervention. The top right panel shows the kernel density function for the control group and the two treatments after the intervention. The bottom left panel shows the kernel density function for the control group and the two treatments after three years. The table below the Figure also reports the p-values from two comparison statistical tests (Kolmogorov-Smirnov and Mann-Whitney tests). We can see that the distribution of the percentile rank shifts from the top of the distribution for both the *Meal* and the *Snack* treatments, albeit more pronounced in the *Meal* treatment. For both treatments, the intervention has moved children from the top of the BMI distribution more toward the bottom half of the distribution. Mann-Whitney comparison tests show that immediately after the intervention, the median percentile rank in children is significantly lower in the *Meal* and *Snack* groups compared to the control group (p-values respectively <0.05 and <0.01). The difference between the treatment groups and control group remains positive up until year 3 but is less precisely estimated (p-value 0.075 for the *Meal* and 0.095 for *Snack*).

Altogether, these results provide evidence that children reduced their consumption of foods high in sugar, and to a lesser extent fat, which in turn affected their BMI. The effects are persistent and precisely estimated for the *Meal* treatment. They are less clear for the *Snack* treatment, with estimates becoming noisier over time.¹⁹

We present additional results in the Appendix. First, we present results including individual fixed effects (see Appendix Section A.7), which are almost identical. In Appendix Section A.8, we report the results of the analysis on the adult sample, for which we find no consistent evidence of improvements in dietary habits or changes in BMI. We also collected

¹⁸As an alternative analysis (available upon request), we have also computed two healthy eating indices. The first index uses the recommendation threshold for each macronutrient (a dummy variable is created for each macronutrient, which is equal to 1 if the participant is within the recommendations' ranges, 0 otherwise. We then sum those dummies to create an index). A second index is inspired by Handbury et al. (2015). We use the share in the recommended quantities and percentage from calories allowing us to have an index informing on the diet quality. We find no significant treatment effects on any of these two indices, for children or adults.

¹⁹In Section 7 we examine the robustness of the results to alternative inference procedures, multiple hypotheses testing and attrition.

blood samples from a sub-sample of adults to identify several key biomarkers, but again did not find evidence which would indicate improvements in diet (see Appendix Section A.9).

5.4 Mediation Analysis

The evidence presented so far points to the temporary reduction in sugar intake, where we find the most consistent change in diet and this resulted in a sustained decrease in BMI. To examine this more formally we carry out the following exercise. We estimate the treatment effects for BMI and then we repeat the exercise controlling for sugar intake, as measured by the 24h-dietary recall. We then compare the change in the coefficients and test to see whether the differences we find are statistically significant. We test the difference between the effects for each treatment for each period individually, as well as jointly across all periods. We repeat this analysis for fat, and fruit and vegetable consumption.

This exercise is shown in Appendix Section A.11 Table A.26. We note that when added sugar is included as a control, the point estimates of the treatment effect decrease for each year and for both treatments. On average, the treatment effects are 11% and 11.5% lower when added sugar consumption is controlled for in the meal and snack treatments respectively. This is not the case for saturated fat, or fruit and vegetable consumption, where the pattern is not so consistent. When controlling for added sugar the difference in treatment effects are not always statistically significant. However, when we jointly test all the coefficients, we find that controlling for sugar statistically significantly reduces the point estimates in the Meal treatment. We find a similar effect for the Snack treatment albeit less precisely estimated.

As described above, while we do not find a statistically significant reduction in sugar consumption immediately after the treatment for the Meal intervention, controlling for added sugar consumption does reduce the treatment effect by 19%. Thus, even if we do not detect a significant treatment effect on added sugar immediately after the intervention, there appears to be a correlation between a change in BMI and a change in added sugar consumption in the direction we would expect: those who experience a decrease in BMI are more likely to be those who reduce their added sugar consumption. This is again evidence pointing to the reduction in sugar being a driver of the reduction in BMI. Controlling for added sugar does

not explain the entire effect on BMI though, but we would not expect this to be the case either. The measure of added sugar relates to a specific window of 24 hours. It is not a perfect measure of someone’s diet over a long period. We interpret these results as indicative that a change in diet is indeed an important driver of the change in BMI we observe.

6 Mechanisms

The results suggest that both interventions had an impact on children’s diet and body mass index in the short run, and for at least 3 years after the intervention for the *Meal* treatment. The interventions were designed to be as effective as possible and therefore exploited several potential levers simultaneously. We now discuss key potential mechanisms, as well as evidence we have for the role they may have played.

6.1 Financial constraints

In both interventions, families received ingredients for free. Thus, the first obvious mechanism driving the effects could be a relaxation of financial constraints.

Our results suggest this is unlikely to be the case. First, we see effects beyond the year of intervention, which are not plausible if the key mechanism is financial constraints. We also see no indication that children *increased* their consumption of healthy foods, rather we see an indication that they *reduced* their consumption of foods high in sugar and in fat.

6.2 Non-financial costs of implementation

Both interventions potentially alleviated other *non-financial* barriers to the adoption of healthier habits. In the *Meal* Treatment, families got everything they needed at their doorstep: The groceries are delivered directly to their home and they receive a weekly booklet of recipes using the ingredients delivered. The recipes have been chosen for their simplicity of execution. Similarly, for the *Snack* treatment, we asked families to follow a simple protocol (eat at regular times and avoid eating between these times) and healthy snacks were delivered directly at home. However, we did ask families to adhere to a specific

protocol for 12 weeks, which is quite a long period, and these protocols do obviously impose constraints.

The compliance measures discussed in Appendix Section A.6 give us an idea of how these protocols were experienced by families. For example, we know that the *Meal* protocol was perceived as relatively easy to follow. Participants from both treatments were asked whether they experienced any difficulties in implementing the protocol, and if, in general, adults and children liked and ate the food that was delivered. Appendix Table A.11 shows differences in self-reported measures regarding the ease of implementation of the protocols. We find notable differences between the two treatment groups. For instance, 42.5% of the *Snack* sample found it difficult or very difficult not to snack between the meals. In contrast, 83.7% of those in the *Meal* treatment say they found it easy or very easy to cook the recipes.

In the *Meal* treatment, one year later, families reported cooking on average 6.05 recipes, a number that remained stable over the next years (5.7 in year 2 and 6.2 recipes in year 3), and on a 5-point Likert scale (from 1 being very easy to 5 being very difficult) adults report an average difficulty of 1.7 (s.d=0.5). Those results corroborate the self-reports displayed in Appendix Table A.11, showing that this treatment has been perceived as relatively easy to follow.

In the *Snack* treatment, the feedback leaflet allows us to check the regularity in the meal times as participants were reporting the time of the three (five for the children) meals they had during the twelve weeks. For each week, we set the mode time as the regular time and we look at the frequency of a deviation of at least 30 minutes from this mode. 18.7% of the adults' breakfast deviated from their mode, and 16.5% for the children.

Thus, it is plausible that the ease of sustaining the protocol in the *Meal* treatment may explain why it had more sustained effects.

6.3 Information and Learning

Since following a healthy diet and preparing healthy meals is not trivial, and in general 'experimentation' is costly, both interventions could have had an impact by subsidizing 'experimentation' and helping families to discover new recipes to prepare healthy and affordable meals, and snack options that are to their liking.

While this is plausible, it is striking that we do not see an impact on adults. Also, adults do not indicate having difficulties accessing healthy foods at baseline, and they seem to be generally well-informed about what they would need to change to adopt a healthier diet. Section A.12 includes two tables showing the percentage of responses for different questions about information access asked during the computer surveys, at the different sessions. Hence, about 94% of the participants believe they have access to healthy food, already at baseline. 90% of them are convinced that they can have influence on their own health (see Appendix Table A.27). They also seem well aware at baseline about the different guidelines as they know which food or drink items consumption they should increase/decrease to meet the guidelines (see Appendix Table A.28). Those answers do not evolve much over time.

This said, it is plausible that parents in the *Meal* treatment added new and healthier meals to the repertoire they use with their children, whereas for the *Snack* treatment, the challenges of adhering to a regular meal pattern seemed to have been large.

6.4 Food Preferences (primary outcome)

One question is whether the treatments significantly affected children’s tastes, in line with a habit formation type of model (Becker and Murphy (1988)). To investigate this possibility, we collected direct measures of food preferences. These measures were pre-registered as primary outcomes of interest.

Due to the young age of the children at the beginning of the study, our main variable is a simple non-incentivized measure of preferences for a range of foods. We collected additional incentivized measures of food preferences based on single one-shot decisions in year 1 for adults and year 3 for children (see Appendix Section A.10).

Children and adults were asked independently to rate their liking of a set list of foods. The survey was computer-based for parents, while for children it was administered on a one-on-one basis in a separate room by a member of our team. The questionnaire included 20 food items aimed at capturing a range of different food groups and five recipes that featured in the *Meal* treatment (see Appendix Table A.22) for the full list of items). For each item, participants had to answer on a 4-point scale how much they liked the item (really dislike to really like, illustrated with smiley emoticons), with the additional possibility of an “allergic”

or “never tried” option. Items were then grouped into food categories following the “eatwell plate” food categories (fruit, vegetables, meat/fish/eggs, cheese, bread, unhealthy processed food and sweets).²⁰

The ITT estimates are presented in Table 5. We find that children in the *Meal* treatment report a weaker preference for foods high in saturated fat, immediately after the intervention. The impact remains negative although not precisely estimated in the years after the intervention. In contrast, children report liking sweets more. The point estimates for the *Meal* treatment effects are all positive, but they are neither jointly nor individually statistically significant. In the *Snack* treatment, the point estimate is initially positive and significant but then the sign changes in subsequent years. Overall, however, we find a positive and significant increase in the preference for sweets for those in the *Snack* treatment, although neither of the estimates for the effect on preferences for sweets remain statistically significant with more conservative inference procedures (Appendix Table A.30).

An increase in the liking of sweets seems a priori inconsistent with a model of habit formation, but is consistent with the most basic economic model of decreasing marginal utility of consumption, and with evidence in nutrition research, which shows that restricting access to certain foods can increase their valuation (Fisher and Birch (1999)).

In addition to the self-reported measure, we also collected a measure based on a one-shot incentivized choice in year 3 for children, when they were presumably old enough to fully understand an experimental protocol with incentives. The task was conducted one-on-one with each child and with the same investigator.

Children were offered a choice between a low-calorie food item and a high-calorie item. They were presented with four different pairs of items involving either sweet or savory food, fresh or storable. To ensure that the low-calorie item was attractive, we chose to impose a price on each item. The simplest price we thought we could implement was to attach a risk of not obtaining it. More precisely, if the child chose the low-calorie item, she would have two chances out of three of actually getting it, while if she chose the high-calorie item, she would only have one chance out of three to get it. The child had to indicate which item she chose among four different pairs of low-high calorie items (see Appendix Figure A.6)

²⁰The eatwell plate is a policy tool used to define the British government recommendations on eating healthily and achieving a balanced diet.

The outcome measure of interest is a simple count of the number of times the child picks the low-calorie (healthy) item. Great care was taken to ensure that the child understood the instructions.²¹

We find no statistically significant difference across treatments (see Table 6). That is, we have no indication that children in either treatment group developed a preference for “healthier” foods.

Altogether, these results provide little evidence that the interventions resulted in stronger preferences for healthier foods. But the results indicate that the *Meal* treatment intervention affected purchasing behavior of the parents.

6.5 Summary

Summarizing our findings, we conjecture that the most plausible mechanism for the changes observed is that parents adjusted the food they provided to their children for the longer-term. In the *Meal* treatment specifically, they continued preparing some of the healthy meals, and they purchased less sugary foods.

The fact that we do not see an effect on adults also fits with this explanation as it may be easier for parents to control their children’s diet than it is to control their own. It could also be that dietary habits are much more ingrained later on in life and are therefore much less malleable.

7 Robustness and Attrition

In this section, we document various robustness checks of the main effects. We focus on alternative inference procedures as well as the issue of attrition. We include individual fixed effects for all outcome measures that were taken pre- and post-intervention. The relevant tables are presented in Appendix Section A.13.

²¹Since the price involves risk, we also collected a measure of risk preference, inspired by the bomb task (Crosetto and Filippin (2013), see Appendix A.10.2 for further details).

7.1 Alternative Inference Procedures

We perform a randomization inference procedure set out in Young (2019). This involves a test of a sharp null (all participants of a particular treatment have a zero treatment effect rather than an average treatment effect of zero). These results are shown in Table A.29 for BMI, Table A.30 for food preferences, and Table A.31 for dietary intake. Point estimates that are statistically significant when estimated with standard errors clustered at the household level remain so under the randomization inference procedure.

7.2 Multiple Hypotheses Correction

Recall that adjustments to p-values are made in Tables 3 and 4 for multiple hypotheses testing on the joint tests for coefficients for each of the variables we examine (i.e. BMI, calories, sugar, fat and fruit and vegetables). Here we now document adjustments made within the analysis of food preferences (Table A.30) and dietary recall (Table A.31) separately. The corrected p-values which adjust for the family wise error rate (FWER) as proposed by Romano and Wolf (2005a) and Romano and Wolf (2005b) are shown in square brackets.

We find that those point estimates that are significant in Table 5 are also statistically significant when carrying out these more conservative inference procedures except for the increased preference for sweets that we found in the snacking treatment after the intervention. Table A.31 shows that those point estimates that were precisely estimated in Tables 3 and 4 remain so correcting for multiple hypothesis testing.

7.3 Attrition

As mentioned earlier, there is some attrition in our sample. To evaluate how attrition affects our main results, we carry out three exercises. First, we compare the baseline characteristics of all those who started the experiment with the baseline characteristics of those who attended the final session in year 3. The idea is to see whether those who do and do not drop out differ based on the baseline measures of interest (our outcomes). There are no statistically significant differences - the means are remarkably similar (Appendix Tables A.32, A.33 and A.34). The biggest differences occur in the *Snack* treatment, as might be expected

due to the larger attrition rate, however, the differences for both adults and children are in absolute size small.

A second way to examine the impact of attrition, and specifically to see whether differences in attrition from session to session affect our estimates, is to compare the treatment effects of the sample of participants who attended every session (or who attended the final session) with the baseline estimates. We focus on the two outcomes with significant treatment effects and for which we have repeated measures: children’s BMI and consumption of added sugar.

The estimated treatment effects on a balanced sample or on a sample of those who attended the last session are slightly smaller but are very similar to the baseline estimates (Appendix Tables A.35 and A.36, suggesting that attrition was not systematically related to improvements in outcomes.

The third exercise is to estimate bounds on the treatment effects taking attrition into account in the spirit of Lee (2009). We describe the procedure in detail in Appendix Section A.13.4. Here again we focus on children’s BMI and consumption of added sugar. For each session, we sort the difference in outcomes comparing the baseline to the particular session within each treatment group from highest to lowest. Then for that particular session, the sample is trimmed from above and below to produce groups that have equal rates of attrition. We then re-estimate the treatment effects based on the trimmed samples (Appendix Tables A.37, A.38, and A.39). We do not trim to produce equal groups because we did not begin with equal groups by design as described above. Our conclusions are robust to this exercise.

These three exercises show that attrition cannot explain the significant treatment effects we have identified for children’s BMI and added sugar consumption.

7.4 Power

As it was not standard practice when we designed the experiment, we did not perform power calculations before implementing the interventions. But we can present ex-post calculations that are based on data that existed prior to our study, and would have been natural anchors to do such calculations ex-ante.

Starting with BMI, a recent Cochrane review by Brown et al. (2019) summarizes the findings from 16 RCTs evaluating interventions targeting diet and physical activity among children between the age of 0 and 5 and find a mean effect on BMI z-score of -0.11 (95% confidence interval -0.21 to 0.01). They only report one study targeting dietary habits, by Daniels et al. (2012). The intervention tested is a universal educational intervention to promote protective feeding practices that commence in infancy. They find a reduction in BMI z-score of 0.19 six months after the intervention. Using this as a benchmark, for the *Meal* treatment we would have 0.793 power to detect such an effect. The *Snack* treatment was slightly under-powered at 0.7.

To examine the power to detect effects in nutrient intake we use data from the National Diet and Nutrition Survey (NDNS) (2012/13-2013/14) as a baseline measure. The data is grouped into various ages and we use the closest one to our sample (children aged between 4 and 10 years old). The mean daily consumption of calories from this survey are 1,462 with a standard deviation of 317. Our realized mean calorie consumption of the children in our sample (control) is 1,439, which is close to the NDNS mean, although the standard deviation is somewhat larger of 539, suggesting our measure is quite noisy. Few studies have targeted calorie reduction of children, so we test power against a 10% change and a more conservative 5% change in intake.

Under an assumption of a correlation between baseline and follow-up measures being 0.8, we have over 80% power to detect a 10% change in calories in both the *Meal* and *Snack* treatment. We are below 80% power in some circumstances however. These include: for a 5% reduction in added sugar in the snack treatment, a 5% change in saturated fat in both treatments, and we are under-powered for a 10% effect for both treatments with respect to fruit and vegetable consumption. Otherwise, we have over 80% power to detect up to 5% changes in the various nutrient intake measures.

For food preferences, we can use data collected by Wardle et al. (2003), a seminal study on food preferences. To convert their 5-point likert scale to our 4-point one we multiply their measure by 0.8. Mean preferences from Cooke and Wardle (2005) for fruit, meat, eggs, fish, dairy, and vegetables are broadly similar to our baseline means. As a benchmark for the estimates, we use results from Wardle et al. (2003) who have examined repeated exposure

to vegetables on liking. They found increases of around 1.6 to 1.3 based on adjusted likert scales. We have over 80% power for all our measures on food preferences for a change in the liking by 1, equivalent to going from on average from “good” to “awesome” which is more conservative than those changes found by Wardle et al. (2003).

Overall, while our sample size may appear at first to be small, it is relatively large in comparison to other studies, particularly in the area of diet. A recent review of studies evaluating interventions targeting the prevent of childhood obesity by Perdeu et al. (2021) only found eight studies targeting families, five of them with a sample of less than 100 families, and only two of them with a larger sample than in our study.²² These interventions were not evaluated with a RCT. On the whole, for key measures, we are not under-powered to detect reasonably and economically significant effect sizes.

8 External Validity

As with any field experiment, one can question the external validity of the results. We discuss here the transparency checklist proposed by List (2020) that should help understanding to what extent our results are generalizable.

Selection: The study focuses on low-income families in the UK. As mentioned earlier, the experimental sample was selected in various ways. First, we selected two specific locations in the UK where the study could be conducted. Second, we targeted families satisfying specific eligibility criteria. Third, families self-selected into participating in the study. For ethical and practical reasons, participants had to be informed that participation in the study could involve having to follow a protocol for 12 weeks. One may perhaps expect these families to be more receptive than the average to policies encouraging changes. It could be that the results we find for children are an upper bound if we believe the parents taking part in this study are more motivated than the average parent in encouraging healthy eating in their children. The baseline data does not provide much indication of positive selection, in particular the diet of children does not meet the UK recommendations. So we do not have

²²The study with the largest sample size had 808 families involved.

indication of positive selection on baseline characteristics.

Selection could have also be influenced by the size of the incentives provided for participation. As explained earlier, the compensation was carefully chosen to be commensurate to the time and efforts expected from participating families. However, recent work by Dutz et al. (2021) shows that incentives do not only affect participation rates but can also bias the participating pool. One interesting feature of our study is that we did increase the participation incentives between wave 1 in Edinburgh and wave 2 in Colchester. It is not as ideal as an experiment as the one discussed in Dutz et al. (2021). But interestingly these two samples compare well on observables (see Appendix Table A.7 as well). For example, the average household income is essentially identical in both samples.

Also related to selection, one may wonder how the UK compares to other countries. There are varied criteria for defining obesity in children and not all countries collect systematic surveys in nationally representative samples. Nevertheless, the available data from various countries clearly show a growing epidemic of childhood obesity in both developing and developed countries (Dalstra et al., 2005; Waters et al., 2011). That is, the UK is not an outlier in child obesity trends.²³

Attrition: The previous section addressed issues related to attrition. An important aspect of the study was that compliance was not incentivized, but participation to the measurements was. This feature of the design limits a possible correlation between compliance and attrition. The feedback leaflets made it clear to participants that we were interested in the difficulties they encountered in implementing the protocols. The anecdotal evidence we have suggests that non-participation was more likely to be caused by external factors, such as incompatible working schedules or personal circumstances unrelated to the interventions or the study.

Naturalness: One concern could be that the interventions placed participants in an artificial setting. We took great care to ensure that this was not the case. The main reason for

²³Note that for adults, the accumulated evidence also suggests an inverse association between Socioeconomic status and obesity among women in developed societies, see Bennett et al. (2008). Among all racial and ethnic groups combined, women of lower income are approximately 50% more likely to be obese than those with higher income levels. Evidence of the inverse association is less consistent among men. A few studies conducted in the US show that, in men, the association varies by race. In white men, obesity was inversely associated with income, while for black men obesity and income were positively associated Miyawaki et al. (2021).

opting to deliver food at home was to ensure that participants could eat in a familiar setting. Online supermarket deliveries are common in the UK and the ingredients were purchased from one of the largest supermarkets in the UK. The fact that the food was provided for free may appear unusual, however programs which target low-income families do exist (such as food banks).

Scaling: The primary goal of the interventions tested here was to examine the malleability of dietary habits. Providing food for free and delivering it at home is relatively costly. This said, food and meal delivery programs targeting vulnerable households do exist. In the UK, *Meals on Wheels* (MOW) programs provide home-delivered meals to over 1.5 million older adults. In the US, the Commonwealth Care Alliance (CCA), a not-for-profit community-based health plan that manages and administers care for adults, runs such a program targeting adults dually eligible for Medicare and Medicaid.

The costs of rolling out such interventions would need to include the costs of the logistics, but since online supermarket platforms are now common in most countries, organizing the delivery of the food should be feasible at low cost. The costs of the food ingredients are also relatively low. For a family of two (one parent and one child) this is approximately £12 per person per week. For a family of more than 2, this is approximately £10-11 per person per week. The costs to deliver the snacks were about £5 per child per week. Such costs are well within the range of costs of alternative interventions that have been implemented in public policy programs in the UK and in the US.

9 Discussion and Conclusion

This paper presents the evaluation of two interventions targeting dietary habits. The first changes what families eat (*Meal* treatment), while the second targets regularity of eating and snack choices and behaviors (*Snack* treatment).

We find that, relative to the control group, children in both treatments moved down the distribution in terms of weight and BMI, particularly those from the top half of the distribution. This result persists for three years for the children exposed to healthy meals (*Meal* treatment), but does not persist for those in the *Snack* treatment.

The most plausible explanation is that parents purchased less unhealthy foods (e.g. foods high in sugar) and that resulted in a lower BMI in the treated groups. The difference in consumption of added sugar disappears in year 3 though, but the effect on BMI persists. So it could be that the temporary reduction in sugar intake (for a couple of years), has led to a sustained decrease in BMI. The effects are weaker and less consistent across measures in the *Snack* intervention, and hence the lack of persistence of the BMI effect.

Is the change in BMI observed for children large? As mentioned earlier, the only other study we are aware of that also targeted young children is a study by Daniels et al. (2012), who evaluated an intervention promoting protective infant feeding practices to prevent childhood obesity and found a reduction in BMI z-score of 0.19 six months after the intervention. In comparison, the effects we find here (0.34 decrease in BMI z-score for the *Meal* treatment) are large.

The changes we observe do not seem to affect preferences for healthier foods. Thus we do not find evidence of a habit-formation mechanism in the spirit of habit-formation models (Becker and Murphy, 1988). Of course, the interventions may not be long enough to alter preferences.

Finally, even though these interventions are not meant to be policy proposals, the effects we find provide encouraging evidence that interventions aimed at changing children’s diet early on in life can have a longer-term impact. In our view, they could be realistically implemented as cost effective policy interventions.

10 Data Availability

Code replicating the tables and figures in this article can be found in Belot et al. (2024) (Replication Data for: “Changing Dietary Habits Early in Life: A Field Experiment with Low-Income Families”. 2024. doi: 10.7910/DVN/QMRM7W.)

References

- M. H. Alhussain, I. A. Macdonald, and M. A. Taylor. Irregular meal-pattern effects on energy expenditure, metabolism, and appetite regulation: a randomized controlled trial in healthy normal-weight women. *The American journal of clinical nutrition*, 104(1):21–32, 2016. [10](#)
- H. Allcott, R. Diamond, J.-P. Dubé, J. Handbury, I. Rahkovsky, and M. Schnell. Food deserts and the causes of nutritional inequality. *The Quarterly Journal of Economics*, 134(4):1793–1844, 2019. [5](#)
- S. Ambuehl, B. D. Bernheim, and A. Ockenfels. What motivates paternalism? an experimental study. *American economic review*, 111(3):787–830, 2021. [5](#)
- M. L. Anderson. Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *Journal of the American statistical Association*, 103(484):1481–1495, 2008. [42](#), [43](#)
- A. Baulos and J. Heckman. The importance of investing in early childhood development and the role of families. In *Handbook of research on innovative approaches to early childhood development and school readiness*, pages 38–54. IGI Global, 2022. [1](#), [5](#)
- G. S. Becker and K. M. Murphy. A theory of rational addiction. *Journal of political Economy*, 96(4):675–700, 1988. [24](#), [33](#)
- F. Bellisle. Meals and snacking, diet quality and energy balance. *Physiology & behavior*, 134:38–43, 2014. [1](#), [10](#)
- M. Belot and J. James. Incentivizing dietary choices among children: Review of experimental evidence. *Food Policy*, 111:102319, 2022. [1](#), [8](#)
- M. Belot, J. James, and P. Nolen. Incentives and children’s dietary choices: A field experiment in primary schools. *Journal of health economics*, 50:213–229, 2016. [4](#)
- M. Belot, N. Berlin, J. James, and V. Skafida. Replication Data for: “Changing Dietary Habits Early in Life: A Field Experiment with Low-Income Families”. 2024. doi: 10.7910/DVN/QMRM7W. URL <https://doi.org/10.7910/DVN/QMRM7W>. [33](#)
- Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1):289–300, 1995. [42](#), [43](#)

- G. G. Bennett, K. Y. Wolin, and D. T. Duncan. Social determinants of obesity. *Obesity epidemiology: methods and applications*, pages 342–376, 2008. [31](#)
- J. Beshears, H. N. Lee, K. L. Milkman, R. Mislavsky, and J. Wisdom. Creating exercise habits using incentives: The trade-off between flexibility and routinization. *Management science*, 67(7):4139–4171, 2021. [8](#)
- L. L. Birch. Development of food preferences. *Annual review of nutrition*, 19:41, 1999. [1](#)
- T. Brown, T. H. Moore, L. Hooper, Y. Gao, A. Zayegh, S. Ijaz, M. Elwenspoek, S. C. Foxen, L. Magee, C. O’Malley, et al. Interventions for preventing obesity in children. *Cochrane Database of Systematic Reviews*, (7), 2019. [29](#)
- A. Brownback, A. Imas, and M. A. Kuhn. Behavioral interventions increase the effectiveness of healthy food subsidies. 2019. [5](#)
- S. Capacci, M. Mazzocchi, B. Shankar, J. Brambila Macias, W. Verbeke, F. J. Pérez-Cueto, A. Koziol-Kozakowska, B. Piórecka, B. Niedzwiedzka, D. D’Addesa, et al. Policies to promote healthy eating in europe: a structured review of policies and their effectiveness. *Nutrition reviews*, 70(3):188–200, 2012. [1](#)
- J. Cawley, A. S. Hanks, D. R. Just, B. Wansink, et al. *Incentivizing nutritious diets: A field experiment of relative price changes and how they are framed*. National Bureau of Economic Research, 2016. [4](#), [5](#)
- G. Charness and U. Gneezy. Incentives to exercise. *Econometrica*, 77(3):909–931, 2009. [4](#)
- G. Charness, R. Cobo-Reyes, E. Eyster, G. Katz, Á. Sánchez, and M. Sutter. Improving children’s food choices: Experimental evidence from the field. *European Economic Review*, 159:104562, 2023. [3](#)
- L. J. Cooke and J. Wardle. Age and gender differences in children’s food preferences. *British Journal of Nutrition*, 93(5):741–746, 2005. [29](#)
- P. Crosetto and A. Filippin. The “bomb” risk elicitation task. *Journal of risk and uncertainty*, 47(1):31–65, 2013. [26](#)
- D. M. Cutler, E. L. Glaeser, and J. M. Shapiro. Why have americans become more obese? *Journal of Economic perspectives*, 17(3):93–118, 2003. [10](#)
- J. A. Dalstra, A. E. Kunst, C. Borrell, E. Breeze, E. Cambois, G. Costa, J. J. Geurts, E. Lahelma, H. Van Oyen, N. K. Rasmussen, et al. Socioeconomic differences in the prevalence

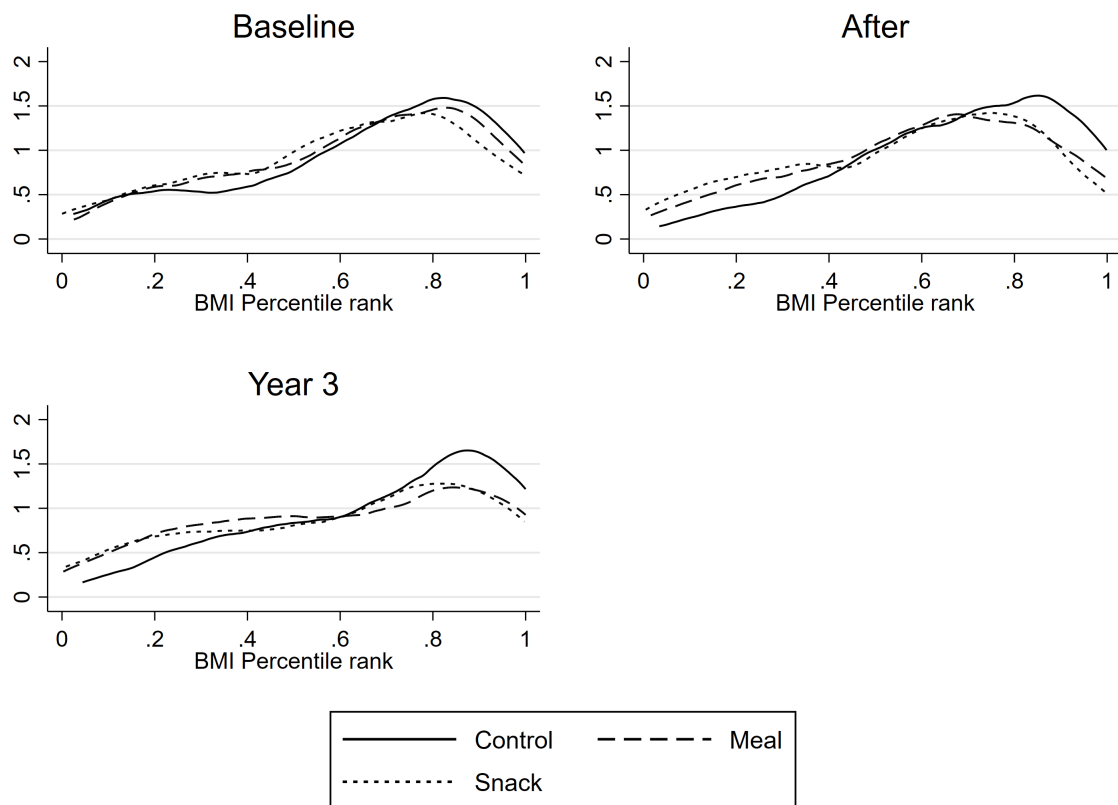
- of common chronic diseases: an overview of eight european countries. *International journal of epidemiology*, 34(2):316–326, 2005. 2, 31
- L. Daniels, K. Mallan, D. Battistutta, J. Nicholson, R. Perry, and A. Magarey. Evaluation of an intervention to promote protective infant feeding practices to prevent childhood obesity: outcomes of the nourish rct at 14 months of age and 6 months post the first of two intervention modules. *International journal of obesity*, 36(10):1292–1298, 2012. 4, 29, 33
- J. S. Downs and G. Loewenstein. Behavioral economics and obesity. In *The Oxford handbook of the social science of obesity*. 2011. 3
- D. Dutz, I. Huitfeldt, S. Lacouture, M. Mogstad, A. Torgovitsky, and W. Van Dijk. Selection in surveys: Using randomized incentives to detect and account for nonresponse bias. Technical report, National Bureau of Economic Research, 2021. 31
- A. E. Field, S. B. Austin, M. W. Gillman, B. Rosner, H. R. Rockett, and G. A. Colditz. Snack food intake does not predict weight change among children and adolescents. *International journal of obesity*, 28(10):1210–1216, 2004. 10
- J. O. Fisher and L. L. Birch. Restricting access to palatable foods affects children’s behavioral response, food selection, and intake. *The American journal of clinical nutrition*, 69(6):1264–1272, 1999. 25
- D. Frisvold and C. Lensing. *Economics of Obesity*, pages 1–21. Springer International Publishing, Cham, 2020. ISBN 978-3-319-57365-6. doi: 10.1007/978-3-319-57365-6_134-1. URL https://doi.org/10.1007/978-3-319-57365-6_134-1. 1
- R. Griffith, S. von Hinke, and S. Smith. Getting a healthy start: The effectiveness of targeted benefits for improving dietary choices. *Journal of health economics*, 58:176–187, 2018. 4
- J. Handbury, I. Rahkovsky, and M. Schnell. Is the focus on food deserts fruitless? retail access and food purchases across the socioeconomic spectrum. Technical report, National Bureau of Economic Research, 2015. 20
- C. Hume, B. Jachs, and J. Menzies. Homeostatic responses to palatable food consumption in satiated rats. *Obesity*, 24(10):2126–2132, 2016. 10
- D. R. Just and J. Price. Using incentives to encourage healthy eating in children. *Journal of Human resources*, 48(4):855–872, 2013. 4

- M. Krawczyk and L. P. Wozny. An experiment on temptation and attitude towards paternalism. *Available at SSRN 2912427*, 2017. [6](#)
- N. Larson and M. Story. A review of snacking patterns among children and adolescents: what are the implications of snacking for weight status? *Childhood obesity*, 9(2):104–115, 2013. [10](#)
- M. N. Laska, M. O. Hearst, K. Lust, L. A. Lytle, and M. Story. How we eat what we eat: identifying meal routines and practices most strongly associated with healthy and unhealthy dietary factors among young adults. *Public health nutrition*, 18(12):2135–2145, 2015. [1](#), [10](#)
- D. S. Lee. Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *The Review of Economic Studies*, 76(3):1071–1102, 2009. [28](#)
- R. M. Leech, A. Worsley, A. Timperio, and S. A. McNaughton. Understanding meal patterns: definitions, methodology and impact on nutrient intake and diet quality. *Nutrition research reviews*, 28(1):1–21, 2015. [1](#), [10](#)
- J. A. List. Non est disputandum de generalizability? a glimpse into the external validity trial. Technical report, National Bureau of Economic Research, 2020. [6](#), [30](#)
- J. A. List and A. S. Samek. The behaviorist as nutritionist: Leveraging behavioral economics to improve child food choice and consumption. *Journal of health economics*, 39:135–146, 2015. [4](#), [5](#)
- J. A. List, A. Samek, and T. Zhu. Incentives to eat healthily: Evidence from a grocery store field experiment. *Economica*, 89(354):489–509, 2022. [4](#)
- G. Loewenstein, J. Price, and K. Volpp. Habit formation in children: Evidence from incentives for healthy eating. *Journal of health economics*, 45:47–54, 2016. [4](#)
- J. Ludwig, J. R. Kling, and S. Mullainathan. Mechanism experiments and policy evaluations. *Journal of Economic Perspectives*, 25(3):17–38, 2011. [1](#)
- A. Miyawaki, C. E. L. Evans, P. J. Lucas, and Y. Kobayashi. Relationships between social spending and childhood obesity in oecd countries: an ecological study. *BMJ open*, 11(2):e044205, 2021. [31](#)
- M. Perdew, S. Liu, and P.-J. Naylor. Family-based nutrition interventions for obesity prevention among school-aged children: a systematic review. *Translational Behavioral Medicine*, 11(3):709–723, 2021. [30](#)

- D. Read and B. Van Leeuwen. Predicting hunger: The effects of appetite and delay on choice. *Organizational behavior and human decision processes*, 76(2):189–205, 1998. [10](#)
- J. P. Romano and M. Wolf. Exact and approximate stepdown methods for multiple hypothesis testing. *Journal of the American Statistical Association*, 100(469):94–108, 2005a. [27](#)
- J. P. Romano and M. Wolf. Stepwise multiple testing as formalized data snooping. *Econometrica*, 73(4):1237–1282, 2005b. [27](#)
- S. Sadoff, A. Samek, and C. Sprenger. Dynamic inconsistency in food choice: Experimental evidence from two food deserts. *The Review of Economic Studies*, 87(4):1954–1988, 2020. [10](#)
- J. Shapiro. Discounting for you, me and we. Technical report, Working Paper 39, Institute for Financial Management and Research, 2010. [6](#)
- K. L. Spalding, E. Arner, P. O. Westermark, S. Bernard, B. A. Buchholz, O. Bergmann, L. Blomqvist, J. Hoffstedt, E. Näslund, T. Britton, et al. Dynamics of fat cell turnover in humans. *Nature*, 453(7196):783–787, 2008. [1](#)
- J. Spiteri, J. James, and M. Belot. A computer-based incentivized food basket choice tool: Presentation and evaluation. *PloS one*, 14(1):e0210061, 2019. [17](#)
- M.-P. St-Onge, K. L. Keller, and S. B. Heymsfield. Changes in childhood food consumption patterns: a cause for concern in light of increasing body weights. *The American journal of clinical nutrition*, 78(6):1068–1073, 2003. [1](#), [10](#)
- J. Wardle, M. L. Herrera, L. Cooke, and E. L. Gibson. Modifying children’s food preferences: the effects of exposure and reward on acceptance of an unfamiliar vegetable. *European journal of clinical nutrition*, 57(2):341–348, 2003. [29](#), [30](#)
- E. Waters, A. de Silva-Sanigorski, B. J. Burford, T. Brown, K. J. Campbell, Y. Gao, R. Armstrong, L. Prosser, and C. D. Summerbell. Interventions for preventing obesity in children. *Cochrane database of systematic reviews*, (12), 2011. [31](#)
- A. Young. Channeling fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results. *The Quarterly Journal of Economics*, 134(2):557–598, 2019. [27](#)
- M. Zumbuehl, T. Dohmen, and G. Pfann. Parental involvement and the intergenerational transmission of economic preferences, attitudes and personality traits. *The Economic Journal*, 131(638):2642–2670, 2021. [5](#)

Tables and Figures

Figure 2: Changes in the distribution of the BMI percentile rank in children



Note: Each line shows the kernel density function for the distribution for the percentile ranks of BMI of children. The top left panel shows the kernel density function for the control group and the two treatments before the intervention. The top right panel shows the the kernel density function for the control group and the two treatments After the intervention. The bottom left panel shows the kernel density function for the control group and the two treatments after three years.

	Control vs. Meal		Control vs. Snack	
	Comparison tests - reported <i>p-values</i>			
	Kolmogorov-Smirnov	Mann-Whitney test	Kolmogorov-Smirnov	Mann-Whitney test
Baseline	0.928	0.591	0.368	0.252
After	0.144	0.046	0.102	0.013
Year 3	0.215	0.075	0.390	0.095

Table 1: Demographic characteristics at baseline and across groups

	Control	Meal	Snack	P-value	P-value
	(1)	Mean (std)	(3)	(1)=(2)	(1)=(3)
	(1)	(2)	(3)	(1)=(2)	(1)=(3)
Sample size (families) (Present at the start of study)	111	103	71	-	-
% Female adults among all sample	72.2	79.6	75.3	0.15	0.59
% Female pregnant among female sample	8.1	6.0	1.9	0.62	0.13
# Adults in household	1.7 (0.9)	1.61 (0.6)	1.7 (0.7)	0.43	0.85
# Children in household	1.9 (0.9)	1.8 (1.0)	1.9 (1.0)	0.76	0.75
Age (adults)	35.1 (7.5)	34.7 (6.6)	34.0 (6.9)	0.67	0.23
Age (study child)	4.1 (1.4)	4.0 (1.6)	3.9 (1.4)	0.71	0.22
Mean annual household income	20,855 (10,056)	21,167 (19,227)	23,928 (21,844)	0.15	0.42
% Receiving child benefit	85.4	85.6	84.9	0.97	0.92
% Receiving tax credit	74.8	70.5	75.3	0.41	0.94
% Receiving job allowance	2.6	5.3	2.2	0.25	0.81
% Receiving housing benefits	35.1	40.9	34.4	0.31	0.91
% Receiving income support	21.9	15.2	18.3	0.15	0.50
% Receiving other benefits	9.3	9.1	6.5	0.96	0.44
% Higher degree	19.6	16.9	18.3	0.57	0.80
% No qualifications	2.7	3.1	3.2	0.85	0.81

Note: Means with standard deviations in parentheses. Col. (4) and (5) report the P-value of a t-test of equality of estimated parameters in Col. (1) and (2) and in Col. (1) and (3) respectively. “Higher Degree” includes having a first degree, higher degree, SVQ level 5, professional qualifications and other higher education qualification. “No Qualifications” corresponds to respondents who ticked the “No Qualifications” option. A descriptive statistics table for panel A Edinburgh and panel B Colchester can be found in Appendix Table A.7. Pregnant women at baseline: 6 in the control group, 4 in the *Meal* treatment, 1 in the *Snack* treatment.

Table 2: Baseline measurements (children)

	Control	Meal	Snack	P-value	P-value
	Mean (std)				
	(1)	(2)	(3)	(1)=(2)	(1)=(3)
Body Size					
BMI z-score	0.5	0.5	0.3	0.59	0.25
% Underweight	0.0	0.0	1.5	1.00	1.00
% Normal weight	71.0	76.0	75.0		
% Overweight	12.2	6.8	8.8		
% Obese	16.8	17.3	14.7		
N	107	104	68		
Diet Intake					
Total calorie intake (Kcal)	1407.8 (464.5)	1460.2 (476.0)	1386.8 (379.4)	0.29	0.85
Added sugar (g)	18.0 (22.7)	25.9 (34.7)	18.1 (20.2)	0.25	0.45
Saturated Fat (g)	23.1 (9.6)	26.1 (13.0)	23.9 (11.6)	0.14	0.87
Fruit and veg (g)	101.8 (125.8)	122.8 (126.4)	119.6 (138.2)	0.10	0.41
N	110	103	72		

Note: Means with standard deviations in parentheses. 4th and 5th columns report the *p-value* of a Wilcoxon signrank test of estimated parameters in Col. (1) and (2) and in Col. (1) and (3) respectively. *p-values* from Kolmogorv-Smirnov test of distribution are reported to compare the BMI categories distribution between groups, rank-sum tests were performed to compare BMI levels. A descriptive statistics table for Edinburgh and Colchester sample separately can be found in Appendix Table A.7. Self-reported preferences: An item that has never been tried or for which the participants declares to be allergic to is considered missing. 1 corresponds to not liking at all, 4 to liking very much.

Table 3: Impact of Meal Treatment on Body Mass Index and Diet of Children

	Estimate	Post	Year 1	Year 2	Year 3
Body Mass Index (z-score for children)	DD	-0.20 (0.06)**	-0.22 (0.08)**	-0.21 (0.09)**	-0.34 (0.12)**
<i>Joint significance of all coefficients</i>			0.003 [0.004]		
Total calories (kcal)					
24h-dietary recall	DD	-54.0 (71.7)	-109.6 (107.7)	-170.3 (115.5)	68.2 (116.8)
Supermarket parental choice	Treat-Control			-19.8 (21.4)	10.2 (22.3)
<i>Joint significance coefficients by year</i>				0.2158	0.7730
<i>Joint significance of all coefficients</i>			0.3657 [0.366]		
Sugar					
24h-dietary recall (added sugar, g)	DD	-5.9 (5.2)	-21.4 (8.1)**	-28.8 (10.3)**	1.2 (9.0)
Self-reported changes (foods high in sugar) : # changes (-8 to 8)	Treat-Control	0.19 (0.61)	-2.0 (0.76)***	-3.69 (0.95)***	-5.20 (1.10)***
Supermarket parental choice (g)	Treat-Control			-4.23 (1.61)**	-1.92 (1.58)
<i>Joint significance coefficients by year</i>		0.49360	0.00110	0.00000	0.00010
<i>Joint significance of all coefficients</i>			0.000 [0.001]		
Fat					
24h-dietary recall (saturated fat, g)	DD	-2.82 (1.84)	-3.23 (2.47)	-3.42 (2.48)	-1.07 (2.59)
Self-reported changes (foods high in fat) : # changes (-8 to 8)	Treat-Control	0.40 (0.57)	-1.89 (0.70)**	-3.48 (0.85)**	-4.64 (1.01)**
Supermarket parental choice (saturated fat in g)	Treat-Control			-0.98 (0.88)	-0.41 (0.92)
<i>Joint significance coefficients by year</i>		0.2334	0.0100	0.0004	0.0001
<i>Joint significance of all coefficients</i>			0.0000 [0.001]		
Fruit and vegetables					
24h-dietary recall (g)	DD	-14.2 (21.2)	-6.90 (40.0)	-28.3 (27.0)	-38.9 (26.5)
Self-report changes (# of changes out of 2 categories)	Treat-Control	-0.28 (0.19)	0.66 (0.24)**	1.46 (0.30)**	2.26 (0.37)**
Supermarket parental choice (% spent on F&V)	Treat-Control			0.24 (0.56)	-0.25 (0.54)
<i>Joint significance coefficients by year</i>		0.2600	0.0234	0.0000	0.0000
<i>Joint significance of all coefficients</i>			0.0000 [0.001]		

Note: Standard errors in parenthesis clustered at the household level. $p < 0.01$ **, $p < 0.05$ *. Joint significance of all coefficients is the p-value of the test as to whether all coefficients are jointly significant for that particular outcome, in square brackets are the FDR q-value calculated using the method from Benjamini and Hochberg (1995) and Anderson (2008). Estimate indicates the estimation strategy with "DD" being difference-in-differences and "Treat-Control" tests the difference between the treatment and control group at for that particular time period. For self-reported changes for foods high in sugar and fat these range from -8 to +8, and for for fruit and vegetables, these are reported as the number of changes out of 2 categories.

Table 4: Impact of Snack Treatment on Body Mass Index and Diet of Children

	Estimate	Post	Year 1	Year 2	Year 3
Body Mass Index	DD	-0.22 (0.09)*	-0.18 (0.11)	-0.12 (0.14)	-0.15 (0.19)
<i>Joint significance of all coefficients</i>			0.078 [0.098]		
Total calories (kcal)					
24h-dietary recall	DD	-48.4 (74.0)	-149.8 (101.2)	-188.9 (111.3)	-103.0 (107.0)
Supermarket parental choice	Treat-Control			8.85 (26.1)	32.3 (29.1)
<i>Joint significance coefficients by year</i>				0.2282	0.2898
<i>Joint significance of all coefficients</i>			0.4586 [0.459]		
Sugar					
24h-dietary recall (added sugar, g)	DD	-0.53 (4.7)	-18.9 (8.6)*	-20.8 (10.8)	-8.7 (9.2)
Self-reported changes (foods high in sugar): : # changes (-8 to 8)	Treat-Control	-0.45 (0.68)	-2.21 (0.87)*	-2.90 (1.04)**	-3.08 (1.21)*
Supermarket parental choice (g)	Treat-Control			0.83 (1.90)	2.21 (2.23)
<i>Joint significance coefficients by year</i>		0.7962	0.0054	0.0142	0.0413
<i>Joint significance of all coefficients</i>			0.009 [0.012]		
Fat					
24h-dietary recall (saturated fat, g)	DD	-1.59 (2.01)	-1.76 (2.42)	-3.19 (2.45)	-2.13 (2.54)
Self-reported changes (foods high in fat) : : # changes (-8 to 8)	Treat-Control	0.43 (0.58)**	-1.11 (0.76)	-2.16 (0.90)*	-2.35 (1.07)*
Supermarket parental choice (saturated fat in g)	Treat-Control			0.11 (1.19)	0.42 (1.11)
<i>Joint significance coefficients by year</i>		0.5488	0.2627	0.0713	0.1272
<i>Joint significance of all coefficients</i>			0.0054 [0.011]		
Fruit and vegetable consumption					
24h-dietary recall (g)	DD	-23.9 (25.7)	-33.4 (32.9)	-34.4 (31.0)	-27.1 (30.0)
Self-report changes (# of changes out of 2 categories)	Treat-Control	-0.18 (0.20)	0.86 (0.26)**	1.49 (0.32)**	2.30 (0.39)**
Supermarket parental choice (% spent on F&V)	Treat-Control			0.48 (0.76)	-0.44 (0.84)
<i>Joint significance coefficients by year</i>		0.4318	0.0046	0.0000	0.0000
<i>Joint significance of all coefficients</i>			0.000 [0.001]		

Note: Standard errors in parenthesis clustered at the household level. $p < 0.01$ **, $p < 0.05$ *. Joint significance of all coefficients is the p-value of the test as to whether all coefficients are jointly significant for that particular outcome, in square brackets are the FDR q-value calculated using the method from Benjamini and Hochberg (1995) and Anderson (2008). Estimate indicates the estimation strategy with "DD" being difference-in-differences and "Treat-Control" tests the difference between the treatment and control group at for that particular time period. For self-reported changes for foods high in sugar and fat these range from -8 to +8, and for for fruit and vegetables these are reported as the number of changes out of 2 categories.

Table 5: Meal and Snack treatment effects on food preferences (children)

	Fruits	Vegetables	Meat Fish Eggs	Processed Food	Sweets	Bread	Cheese
After	0.13*	-0.06	-0.00	0.07	-0.14*	0.05	0.20*
	(0.06)	(0.07)	(0.10)	(0.05)	(0.07)	(0.10)	(0.10)
1-year	0.05	-0.00	0.16	0.05	-0.09	0.01	0.01
	(0.07)	(0.08)	(0.10)	(0.07)	(0.08)	(0.10)	(0.11)
2-year	0.07	-0.08	0.00	0.08	-0.17*	0.02	0.02
	(0.09)	(0.09)	(0.12)	(0.07)	(0.08)	(0.11)	(0.13)
3-year	0.02	-0.12	0.09	-0.07	-0.30**	-0.15	-0.24
	(0.08)	(0.09)	(0.11)	(0.07)	(0.09)	(0.10)	(0.14)
Meal x After	-0.11	0.02	-0.08	-0.21**	0.17	-0.16	-0.32*
	(0.09)	(0.10)	(0.14)	(0.08)	(0.09)	(0.13)	(0.13)
Meal x 1-year	-0.01	0.04	-0.09	-0.09	0.11	-0.11	-0.07
	(0.10)	(0.11)	(0.15)	(0.09)	(0.13)	(0.13)	(0.16)
Meal x 2-year	-0.05	0.04	-0.07	-0.13	0.24	-0.19	-0.18
	(0.12)	(0.12)	(0.16)	(0.09)	(0.13)	(0.14)	(0.18)
Meal x 3-year	-0.02	0.10	-0.18	-0.03	0.21	-0.12	0.03
	(0.11)	(0.12)	(0.16)	(0.09)	(0.13)	(0.14)	(0.18)
Snack x After	0.03	0.14	-0.14	0.02	0.21*	0.24	-0.18
	(0.11)	(0.11)	(0.14)	(0.09)	(0.10)	(0.15)	(0.16)
Snack x 1-year	0.15	0.13	-0.26	0.05	-0.04	0.13	-0.11
	(0.12)	(0.12)	(0.19)	(0.11)	(0.15)	(0.14)	(0.17)
Snack x 2-year	0.17	0.24*	-0.18	0.14	0.19	0.13	-0.20
	(0.13)	(0.12)	(0.18)	(0.12)	(0.11)	(0.17)	(0.20)
Snack x 3-year	0.11	0.07	-0.15	0.07	-0.09	-0.03	-0.23
	(0.15)	(0.12)	(0.17)	(0.12)	(0.14)	(0.17)	(0.21)
Constant	3.20**	2.58**	2.67**	3.35**	3.53**	3.47**	3.29**
	(0.03)	(0.03)	(0.04)	(0.02)	(0.03)	(0.04)	(0.04)
# Obs	1,239	1,250	1,250	1,253	1,247	1,244	1,240
R-squared	0.01	0.01	0.01	0.02	0.04	0.03	0.03
# individuals	287	288	288	288	288	288	287

Note: Each column is from a separate regression. All regressions include individual fixed effects. Standard errors in parentheses are clustered at the household level, ** p<0.01, * p<0.05.

Table 6: Meal and Snack treatment effects on children’s incentivized food preferences

	Number of unhealthy items chosen		
Meal	0.05 (0.16)	0.06 (0.17)	0.08 (0.17)
Snack	-0.18 (0.19)	-0.15 (0.20)	-0.14 (0.20)
Risk preference		0.01 (0.03)	0.004 (0.03)
Controls for gender and age	N	N	Y
Observations	224	212	212

Note: Each column in each panel represents a separate regression. Data was collected on children’s incentivized food preferences in year 3 only.